2006

Analytic Framework for Calculating BRDFs of Randomly Rough Surfaces

Yinlong Sun

Report Number:
06-010

http://docs.lib.purdue.edu/cstech/1653

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.
ANALYTIC FRAMEWORK FOR CALCULATING BRDFS OF RANDOMLY ROUGH SURFACES

Yinlong Sun

Department of Computer Sciences
Purdue University
W. Lafayette, IN 47907, USA

CS TR #06-010
May 2006
Analytic Framework for Calculating BRDFs of Randomly Rough Surfaces

Yinlong Sun
Department of Computer Sciences, Purdue University, W. Lafayette, IN 47907, USA

Abstract

This report presents an analytic framework for calculating BRDFs (bi-directional reflectance distribution function) of randomly rough surfaces. The fundamental assumptions are the tangent plane approximation and statistical approach sufficiency. By treating a local surface area as its tangent plane, the tangent plane approximation validates the description of waves using geometric rays, and the statistical approach sufficiency validates the calculation of reflection using statistical concepts such as surface height probability and correlation function. Based on these assumptions, we derive a generic equation for reflection due to single scattering at a rough surface. As an example, the generic equation is applied to an isotropic surface with Gaussian statistics to obtain a specific analytic form of BRDF. An interesting result we have obtained is that as the surface smoothness varies, the derived BRDF shows four distinct reflection regimes, namely, mirror reflection, grazing reflection, retro-reflection, and normal reflection. In this report, we have also derived an explicit form of self-shadowing for a surface with Gaussian statistics, and the result agrees well with computer simulation. Our analytic solution for single scattering is useful to describe the entire reflection of an opaque surface that is highly or moderately smooth. It also offers a basis for calculating reflection that involves multiple scattering. While this study focuses on optical reflection, the method and solution also apply to reflections of other forms of waves.

Keywords:
Surface reflection, rough surface, single scattering, analytic models, statistical method.
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>Related Researches</td>
<td>6</td>
</tr>
<tr>
<td>2.1</td>
<td>Empirical Approach</td>
<td>6</td>
</tr>
<tr>
<td>2.2</td>
<td>Fraunhofer Diffraction</td>
<td>6</td>
</tr>
<tr>
<td>2.3</td>
<td>Kirchhoff-Beckmann Approach</td>
<td>7</td>
</tr>
<tr>
<td>2.4</td>
<td>Perturbation Approach</td>
<td>7</td>
</tr>
<tr>
<td>2.5</td>
<td>Microfacet Method</td>
<td>8</td>
</tr>
<tr>
<td>2.6</td>
<td>Summary</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>Statistical Ray Method</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>Visibility Functions</td>
<td>15</td>
</tr>
<tr>
<td>4.1</td>
<td>Individual Visibility Function</td>
<td>16</td>
</tr>
<tr>
<td>4.2</td>
<td>Combined Visibility Functions</td>
<td>17</td>
</tr>
<tr>
<td>5</td>
<td>Isotropic Surfaces with Gaussian Statistics</td>
<td>17</td>
</tr>
<tr>
<td>5.1</td>
<td>Additional Conditions</td>
<td>17</td>
</tr>
<tr>
<td>5.2</td>
<td>Gaussian Height Field</td>
<td>18</td>
</tr>
<tr>
<td>5.3</td>
<td>Visibility Function</td>
<td>18</td>
</tr>
<tr>
<td>5.4</td>
<td>Summary</td>
<td>21</td>
</tr>
<tr>
<td>6</td>
<td>Discussion</td>
<td>22</td>
</tr>
<tr>
<td>6.1</td>
<td>Extreme Cases of Surface Smoothness</td>
<td>22</td>
</tr>
<tr>
<td>6.2</td>
<td>Reciprocity and Energy Conservation</td>
<td>22</td>
</tr>
<tr>
<td>6.3</td>
<td>Visibility Function</td>
<td>23</td>
</tr>
<tr>
<td>6.4</td>
<td>Probability Density of Surface Orientation</td>
<td>25</td>
</tr>
<tr>
<td>6.5</td>
<td>Behavior of the Derived BRDF</td>
<td>26</td>
</tr>
<tr>
<td>7</td>
<td>Concluding Remarks</td>
<td>29</td>
</tr>
<tr>
<td>8</td>
<td>Appendices</td>
<td>33</td>
</tr>
<tr>
<td>8.1</td>
<td>Radiometric Concepts</td>
<td>33</td>
</tr>
<tr>
<td>8.2</td>
<td>Statistical Description of Rough Surfaces</td>
<td>33</td>
</tr>
<tr>
<td>8.3</td>
<td>Fresnel’s Coefficients and Optical Constants</td>
<td>36</td>
</tr>
<tr>
<td>8.4</td>
<td>Relationship between Differential Solid Angles</td>
<td>36</td>
</tr>
<tr>
<td>8.5</td>
<td>Separable Combined Probability</td>
<td>37</td>
</tr>
<tr>
<td>8.6</td>
<td>Scaling Invariance Principle</td>
<td>38</td>
</tr>
<tr>
<td>9</td>
<td>References</td>
<td>39</td>
</tr>
</tbody>
</table>
1 Introduction

Reflections of electromagnetic waves at randomly rough surfaces are an essential problem in many scientific and engineering fields. Traditional applications include optical engineering\(^1\)-\(^{12}\), radiophysics\(^{13, 14}\), remote sensing\(^{15-19}\), radar imaging\(^{20-23}\), and metrology\(^{24-26}\). In recent decades, great interest in surface reflection has arisen in areas of computer science, such as computer graphics to synthesize images\(^{27-30}\) and machine vision to recognize objects\(^{31-38}\). In all applications mentioned above, a central task is to determine surface reflection as a function of relevant parameters. In principle Maxwell’s equations\(^{39-41}\) provide a sufficient basis for determining surface reflection and based on that many sophisticated calculations\(^{15, 42-46}\) have been made on wave scattering at rough surfaces, but a number of key problems remain to be solved.

We start with the definition of a bi-directional reflectance distribution function (BRDF), a way commonly used to quantify surface reflection.\(^1, 28, 47-50\) A BRDF is the ratio between a differential radiance \(L_v(\theta, \varphi, \lambda)\) reflected into the viewing (outgoing) direction and the incident irradiance \(L_c(\theta, \varphi, \lambda)\cos \theta \cos \omega\) in the lighting (incident) direction (Figure 1):

\[
\rho(\theta, \varphi, \theta, \varphi') = \frac{dL_v(\theta, \varphi, \lambda)}{dE_c(\theta, \varphi, \lambda)} = \frac{dL_v(\theta, \varphi', \lambda)}{L_c(\theta, \varphi, \lambda)\cos \theta \cos \omega},
\]

where \(\theta\) and \(\varphi\) are the polar and azimuthal angles of the lighting direction, \(\theta\) and \(\varphi\) the polar and azimuthal angles of the viewing direction, and \(\lambda\) is wavelength. Appendix A provides the definitions of radiance and irradiance. In practice, it is more convenient to regard BRDF applying to a small area \(\Delta A\) (which we call a local illumination area) than to a mathematical point on a surface.

![Figure 1: Geometry and notations for defining BRDF.](image)

Analytically deriving BRDFs for rough surfaces is very difficult. The problem involves multiple factors, including the incident and outgoing directions, surface roughness, surface anisotropy, surface composition, material properties, wavelength, and polarization. Surface reflection also involves single and multiple scattering. At a rough surface, a local illumination area \(\Delta A\) may have a complicated profile (Figure 2), so an incident ray may be reflected once, twice, or more times. If the surface material is transparent or translucent, a scattering path (such as path 3-3’ in Fig. 2) may contain both reflection and transmission events. Moreover, natural surfaces not only are often rough, but also differ statistically\(^{24-26, 43, 46}\) (see Appendix B). This means that rough surfaces may have different forms of probabilistic height distributions. For example, the height distributions of many natural surfaces are Gaussian, but non-Gaussian surfaces also exist commonly (such as those generated by turning or engraved by chemicals or...
Finally, surface reflection may involve complex scattering mechanisms and considerable volume absorption. For example, light reflection at a surface of a translucent medium involves surface scattering at the medium’s boundary as well as subsurface scattering by small particles within the medium.

![Figure 2: Microscopic view of a rough surface and ray scattering. The path from 3 to 3' occurs when the medium is transparent. The path from 4 to 4' involves subsurface scattering, which occurs if the material is translucent.](image)

A perfectly smooth opaque surface can be sufficiently described by the mirror-reflection principle and Fresnel’s coefficients (see Appendix C). Imagine that the surface varies from smooth to rough; an effect called surface self-shadowing will play an increasing role. This effect describes the probability for a surface point being blocked by another part of the surface. When a reflected ray is blocked by the surface, a secondary reflection happens. Similarly, tertiary and even higher-order reflections may occur. We call a process with two or more reflections as multiple scattering. Thus, the total reflected energy power has contributions from single and multiple scattering:

\[ \Phi_{\text{surf}} = \Phi_{\text{single}} + \Phi_{\text{multiple}}. \]  

The importance of \( \Phi_{\text{multiple}} \) increases with surface roughness. To determine \( \Phi_{\text{multiple}} \), we need first to determine \( \Phi_{\text{single}} \) because a multiple-scattering process is formed by single-scattering events.

Now suppose that the surface material is translucent instead of opaque; the total reflected power may be expressed as

\[ \Phi_{\text{total}} = \Phi_{\text{surf}} + \Phi_{\text{subsurf}}, \]  

where \( \Phi_{\text{surf}} \) is from surface scattering including single and multiple scattering as shown in Eq. (1.2). However, \( \Phi_{\text{subsurf}} \) involves not only surface scattering but also subsurface scattering by particles inside the translucent material volume. To determine \( \Phi_{\text{subsurf}} \), we must first determine the surface scattering because it is involved as a part for calculating \( \Phi_{\text{subsurf}} \).

Previous approaches on deriving reflection models have a number of limitations. First, the approaches are restricted to particular surface types instead of offering a potential for modeling generic rough surfaces. For example, the Kirchhoff-Beckmann’s approach is difficult to handle multiple scattering and the analytic solutions derived by Beckmann and Spizzichino for rough surfaces require the condition of perfectly conducting material, or the extended version that requires a number of parameters being constants (see pages 17-33 in the book by Beckmann and Spizzichino). But such required conditions are unlikely to hold for generic surfaces and it is often infeasible and inconvenient to verify the condition validity.
Moreover, existing reflection modeling approaches tend to oversimplify the problem or involve empirical treatments that lack adequate justification. For example, the decomposition of the total reflection into diffuse and specular terms is often used, but this treatment has not been seriously justified. In fact, a difficulty occurs when one attempts to interpret the physical meanings of the diffuse and specular terms (i.e., what are the physical processes that are responsible for the diffuse and specular terms). If one interprets the diffuse term as subsurface scattering processes (i.e. corresponding to $\Phi_{\text{subsurf}}$ in Eq. (1.3)) at a surface of translucent material or with multiple layers, then the diffuse term should not be present for opaque surfaces such as metallic. On the other hand, if one interprets the diffuse term as all multiple-scattering processes as a part of surface scattering (i.e. corresponding to $\Phi_{\text{multiple}}$ in Eq. (1.2)), then it is difficult to interpret the meaning of the subsurface-scattering term $\Phi_{\text{subsurf}}$ in Eq. (1.3). One may attempt to use the diffuse term to account for both $\Phi_{\text{multiple}}$ and $\Phi_{\text{subsurf}}$, but then there is an immediate question why the diffuse term should have the same behavior (i.e. Lambertian) for distinctly different scattering processes. The fact that those difficult components (such as for $\Phi_{\text{multiple}}$ or $\Phi_{\text{subsurf}}$) are regarded as the diffuse term in practice is perhaps only for the sake of convenience to achieve quick solutions. However, this point is rarely stated in most existing researches, and as a result the critical problems are hidden and quantifying the diffuse term as Lambertian is commonly taken for granted. Finally, the decomposition of the total reflection into the diffuse and specular terms leads to a question of assigning the coefficients for the two terms, and in practice the coefficients are often chosen with a certain degree of arbitrariness. One may hope to choose the coefficients as functions of the surface roughness such that the diffuse coefficient vanishes as the surface approaches perfectly smooth, but correctly quantifying the coefficients is in fact the same problem as calculating $\Phi_{\text{single}}$ in Eq. (1.2) precisely.

In this report, we attempt to solve an essential component of the reflection problem: single scattering at an opaque rough surface. The current report is a rigorous formulation that significantly extends the basic idea and preliminary analytic calculations published earlier. In the current report, while our derived result for single scattering applies to the entire reflection at an opaque surface that is highly or moderately smooth, the solution also reveals insights of single scattering and offers a useful basis for calculating multiple scattering. Our study will focus on the case of light, but the calculation is generically applicable to other waves as well. This is because when a surface is locally smooth sufficiently, a wave can be described using a ray, just as light waves are treated in geometric optics.

As a contribution, we propose an approach called statistical ray method. This method is established upon two assumptions: the tangent plane approximation and the statistical approach sufficiency. The tangent plane approximation, which is same as in Kirchhoff-Beckmann’s approach, validates the description of waves using geometric rays. The statistical approach sufficiency allows us to calculate the reflected light power using statistical concepts such as surface height probability and correlation function. Using the statistical ray method, we have obtained a generic equation of surface reflection due to single scattering. By applying the equation to an isotropic surface with Gaussian statistics, we found that the derived BRDF demonstrates four distinct reflection regimes, namely, mirror reflection, grazing reflection, retro-reflection, and normal reflection. These regimes are results of competitions of different terms in the derived BRDF.

Another contribution of this report is on calculating surface self-shadowing. This is accomplished in terms of the visibility function, the probability of a surface point visible with respect to a given direction. Note that the visibility function defined in this report is related to the shadowing function in the past research, and we believe that the name “visibility function” is more appropriate because it refers to the surface points that are visible instead of being shadowed. We have derived a generic expression of the visibility function, and obtained an explicit analytic form for a surface with Gaussian height field and Gaussian correlation function. A scaling invariance principle has been introduced to complete the determination of the visibility function. Our derived visibility function for a Gaussian surface agrees well with the computer simulation obtained by Brockelman and Hagfors.

The following sections are organized as follows. Section 2 reviews the previous approaches. Since the primary goal of this paper is on the method of deriving reflection models, our reviews will focus on the existing approaches for deriving reflection models. Section 3 presents the statistical ray method along with the assumptions and calculation. Section 4 focuses on the derivation of visibility functions. Section 5 specifically studies isotropic surfaces with Gaussian statistics. Section 6 discusses the derived BRDF through numerical analysis. Section 7 concludes the report and discusses future research directions. Relevant background materials are provided in the Appendices.
2 Related Researches

2.1 Empirical Approach

An empirical approach specifies the entire BRDF expression or a part of it without giving the detailed analytic steps that lead to the expression. The best approaches for describing scattering over a $2\pi$ sr viewing hemisphere for an actual surface require an interplay of theory and experiment. For instance, the BRDF for a known surface geometry could be obtained, and from the data, a few parameters could be extracted for a model which then describes scattering over a wide range of surface geometries. As an alternate approach, one may start with 2D roughness measurements of the surface height field to determine the profile statistics\textsuperscript{24,26, 43, 46} and then infer the form of BRDFs. But the latter approach might be less accurate as the former approach.

Typically, empirical approaches use simple, intuitive functions to characterize surface scattering behavior, but the conditions or assumptions are usually not specified sufficiently. One example using the empirical approach is the Phong reflection model\textsuperscript{68}, which describes the specular reflection lobe through the dot product between the viewing direction and the mirror-reflection direction. The shape of the specular lobe is controlled by the power of the dot product. This power parameter does not have a physical counterpart. (The power parameter is related to the surface roughness, but the relationship has to be obtained through a comparison with a physically-based model at the asymptotic condition that the surface is very smooth. In other words, the Phong model itself cannot reveal the physical meaning of the power parameter.)

Another well-known example is the Lambertian model where the BRDF is independent of the viewing direction. This model was proposed based on the observation that rough surfaces tend to appear with the same brightness in all viewing directions. The condition of the model is vaguely stated as applying to diffuse surfaces, but the meaning of “diffuse” is not quantitatively specified with respect to the physical and geometrical aspects of the surface.

Many reflection models decompose the entire reflection into the diffuse and specular terms. The diffuse term is commonly assumed Lambertian, while the specular term varies. Unfortunately, there has been little discussion that validates the decomposition. When using the decomposition, one question is how to assign the weights to the diffuse and specular terms. In practice, choosing the weights often involves some arbitrariness. Another question is how to interpret the physical meanings of the diffuse and specular terms. We have discussed the difficulty for interpreting the diffuse and specular terms in Sec. 1, in the paragraph following Eq. (1.3). In addition, when attempting to interpret the diffuse and specular terms, a conceptual conflict may occur. Suppose that the surface reflection can be modeled as a summation of the diffuse and specular terms, and that the specular term contains ray contributions from single scattering. Also, suppose that when a surface is rough enough, the Lambertian term alone sufficiently describes the entire reflection. Thus the Lambertian term contains contributions from single scattering as well. However, single-scattering events have already been included in the specular term. This implies that each one-bounce reflection is counted twice, one in the diffuse term and one in the specular term. This conflict is caused by the decomposition of the entire reflection into the diffuse and specular terms that do not have clear physical interpretations and justifications.

Empirical models are simple and easy to use, but they are not sufficiently accurate. The empirical approach is not able to describe surface self-shadowing and multiple scattering effects. Since the physical basis and steps for analytic expression are not given and the conditions are not specified quantitatively, the empirical approach does not offer a systematic way for improving the model’s accuracy and for generalizing the application conditions.

2.2 Fraunhofer Diffraction

The \textit{Fraunhofer diffraction} theory is based on the Huygens-Fresnel principle\textsuperscript{40, 41} that every unobstructed point of a wavefront at a given instance serves as a source of spherical secondary wavelets with the same wavelength as the primary wave. The amplitude of the field strength at any point beyond is the superposition of the secondary wavelets with consideration of their amplitudes and relative phases. Thus, we can calculate the field through an integral of all contributing secondary wavelets over a wavefront surface, and calculate the reflected energy power through the
product of the field with its complex conjugate. The Fraunhofer diffraction handles the far-field case where the considered point is located far away from the wavefront (the near-field case is called the Fresnel diffraction).

The Fraunhofer theory is rigorous and works for both reflection and transmission. The theory has been widely used to obtain the reflection behavior of surfaces with periodic structures (such as gratings) or simple structures. However, the theory requires that the surface profile is known deterministically. This is a major limitation for deriving reflection models of rough surfaces because a rough surface is often known statistically instead of deterministically. Besides, the Fraunhofer theory is not effective to handle multiple scattering.

2.3 Kirchhoff-Beckmann Approach

It is important to distinguish the two parts involved in the so-called Kirchhoff-Beckmann approach. The first part is the Kirchhoff’s diffraction theory that calculates a field through an integral (i.e. Kirchhoff integral) over the space boundary using a Green’s function. This part is a direct result of the Helmholtz theorem and is rigorous. The second part of Kirchhoff-Beckmann approach focuses on solving the Kirchhoff integral. Given a rough surface, the Kirchhoff integral usually cannot be solved exactly, because the surface profile is complicated. Beckmann and Spizzichino proposed the tangent plane approximation that treats a local surface area as its tangent plane. This approximation allows the use of Fresnel’s law at a local surface region. Another assumption of Beckmann and Spizzichino is the infinite conductivity or its extension version that a number of quantities in the integrand of the Kirchhoff integral must satisfy specific conditions (refer to pages 17-33 in the book of Beckmann and Spizzichino).

The Kirchhoff-Beckmann approach is attractive in the sense that it has led to explicit analytic expressions for some types of rough surfaces. The approach is also able to calculate diffraction effects. However, inherently, the approach is not effective to handle multiple scattering. It does not work for a surface of transparent material where a light scattering path may contain reflection and transmission events, such as path 3-3’ in Figure 2. The approach is also not effective to handle a complex surface that has multiple layers or a translucent material where subsurface scattering is important. This is because the Kirchhoff integral needs to be conducted over the boundary of the interested space and the presence of subsurface scattering will make the boundary conditions too complicated to solve the Kirchhoff integral. Moreover, when using the analytic solutions of Beckmann and Spizzichino, it is important to verify for all required conditions. While the tangent plane approximation is valid in many cases, the assumption of the infinite conductivity or its extension version may not hold. Finally, note that the obtained analytic solutions of Beckmann and Spizzichino’s require that the surface be a Gaussian height field, that is, the height probability of the surface profile is a Gaussian function.

2.4 Perturbation Approach

One perturbation approach applies to surfaces that are slightly rough. A common method is to express the entire solution as a sum of the solution for a smooth surface (the 0th-order solution) and terms arising from the slight surface roughness. The solution for a smooth surface is usually easy to obtain. The 1st-order approximation can be calculated using the Kirchhoff integral (refer to Section 2.3) in which the field at the surface boundary in the integrand takes the values of the 0th-order solution. Similarly, the 2nd-order approximation can be obtained by using the field values of the 1st-order solution in the integrand. The method is restricted to slightly rough surfaces.

Another perturbation method is based on Rayleigh theory. The theory was originally developed by Rayleigh to study reflection of a wave normally incident onto a sinusoidally corrugated rough surface, and was later extended to arbitrary incidence. The basic idea is to express the unknown scattered field as a sum of outgoing plane waves and to determine the unknown coefficients in the sum by satisfying the boundary conditions at the surface. As major representative work, Church introduced the vector perturbation technique and Elson and Bennett developed a similar approach to optical scattering theory. The technique is currently known as the Rayleigh-Rice vector perturbation theory. The Rayleigh method requires that the surface only be slightly rough such that the sum convergence for the outgoing scattered waves is easily achievable.
2.5 Microfacet Method

The foundation of the microfacet method is the microfacet surface model that was proposed by Torrance and Sparrow in 1967 to quantify light reflection at rough surfaces. The microfacet model assumes that a surface is comprised of many planar, perfectly smooth, and isotropic facets. Thus light reflection at each microfacet is perfectly specular. Furthermore, within the microfacet model, Torrance and Sparrow assumed that a microfacet surface is formed with V-shaped grooves, which assumption allows the derivation of a shadowing factor based on intuitive geometric consideration.

A well-known reflection model obtained using this approach is the Cook-Torrance model, where specular reflection is expressed as a product of the Fresnel coefficient, the surface self-shadowing factor, and the surface orientation probability. The microfacet method has a potential to handle multiple-scattering and wave reflections from surfaces of transparent or translucent materials. A major limitation is that the derivation of the shadowing factor is based on a surface composition of V-shaped grooves, which is often not the case for practical rough surfaces. Recently, Ashikhmin et al. have made a more rigorous approach for deriving BRDFs based on the microfacet model without the limitation of a surface composition of V-shaped grooves.

While the microfacet model has a major impact in modeling reflection at a rough surface in general, the key idea of the model is also useful to describe scattering of flat particles. For example, Germer and collaborators have applied the facet model and ray description to quantify reflection from pigment flakes within a transparent layer of coating. They have derived analytic BRDFs of surfaces with coating based on the probability distribution of surface normals of pigment flakes. The self-shadowing effect has not been considered because the density of particles is assumed low. Also, polarized single scattering using the Mueller matrix for a micro-facet has been investigated by Priest and Meier. They obtained closed-form result for the polarimetric BRDF from the microfacet model.

2.6 Summary

Table 1 summarizes the existing approaches along with the statistical ray method developed in this study. The first column lists the criteria that we believe important for modeling surface reflection. The Fraunhofer diffraction, Kirchhoff-Beckmann, as well as perturbation approaches can model diffraction effects. All approaches have a potential to support polarization, transmission, and surface anisotropy. The effect of polarization could be handled by decomposing the incident wave into the parallel and perpendicular polarizations, and by calculating the two components separately. This calculation could be very complicated, but in principle possible in all approaches. Regarding the case of light transmission, the Kirchhoff integral in the Kirchhoff-Beckmann and perturbation approaches should be applied to the space on the other side of the surface with respect to the incident wave. Surface anisotropy could be described using statistical parameters such as the surface correlation lengths (see Appendix B). On surface self-shadowing, the microfacet and statistical ray methods contain a shadowing term to calculate reflection, and the empirical, Fraunhofer, and Kirchhoff-Beckmann approaches may include a multiplicative shadowing factor in the final expression or in the integrand of the Kirchhoff integral. However, the perturbation approach does not need to handle self-shadowing because its applied surfaces are only slightly rough. On multiple scattering, only the microfacet and statistical ray methods are capable because these methods may handle a scattering process by tracing the ray path. Note that when using the Kirchhoff integral, calculating multiple scattering is the same problem as solving the integral over a rough surface, for which a generic solution is impossible. Also, because the microfacet and statistical ray methods allow to trace rays, they have the potentials of handling reflection at a transparent surface or a complex surface with multiple layers. Finally, on surface roughness and statistics, the Kirchhoff-Beckmann approach and the microfacet method require a Gaussian statistics to achieve explicit solutions, but the statistical ray method works for both Gaussian and non-Gaussian surfaces (this will be shown in the later sections).
Table 1: Comparison of the approaches for deriving reflection models. The first column lists the criteria.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Example model</td>
<td>Phong⁶⁸</td>
<td>Iridescence of CD ROM⁵⁴</td>
<td>Beckmann and Spizzichino’s solution¹⁵</td>
<td>Rayleigh-Rice model¹⁵</td>
<td>Cook-Torrance model¹⁵</td>
<td>Derived BRDF in the current report</td>
</tr>
<tr>
<td>Required conditions</td>
<td>Not specified</td>
<td>Generally valid</td>
<td>- Tangent plane approximation - Infinite conductivity (or its extension)</td>
<td>Slightly rough surfaces</td>
<td>Surface consists of V-shape, perfectly smooth facets</td>
<td>- Tangent plane approximation - Statistical approach sufficiency</td>
</tr>
<tr>
<td>Support diffraction</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Support polarization</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Support transmission</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Support anisotropy</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Handle self-shadowing</td>
<td>Possible</td>
<td>Possible</td>
<td>Possible</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Handle multiple scattering</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Possible</td>
</tr>
<tr>
<td>Reflection of transparent material</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Possible</td>
<td>Possible</td>
</tr>
<tr>
<td>Reflection of complex surfaces</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Possible</td>
<td>Possible</td>
</tr>
<tr>
<td>Surface roughness and statistics</td>
<td>Not specified</td>
<td>Surface profiles are known deterministically</td>
<td>Solutions are possible for Gaussian rough surfaces</td>
<td>Slightly rough surfaces</td>
<td>Surfaces of V-shape, perfectly smooth facets - Gaussian</td>
<td>- Various roughness - Gaussian and non-Gaussian</td>
</tr>
<tr>
<td>Others</td>
<td>- Inaccurate</td>
<td>- Unfeasible for most surfaces because profiles are not known deterministically</td>
<td>- With analytic solutions for some surface types</td>
<td>- Work only for slightly rough surfaces</td>
<td>- Need empirical treatment (such as decomposition of total reflection)</td>
<td>- Provide a generic equation for deriving reflection models</td>
</tr>
</tbody>
</table>

The importance of surface self-shadowing for reflection was recognized long ago. A simple analytic form of the shadowing factor was derived based on the V-shaped geometry of surface by Torrance and Sparrow⁷³, and was later used to model surface reflection⁷³, ⁸¹. The first attempt of analytic derivation of the shadowing factor was made by Beckmann⁶¹, but his result was found in marked disagreement with the computer simulation of Brockelman and Hagfors⁶². Wagner⁶³ presented a rigorous calculation of the shadowing function based on the consideration of the visibility probability of a surface with Gaussian height field. However, Wagner’s analytic results, including his approximate version, are rather complicated. Smith⁶⁴ simplified the calculation, but his final expression involves an error function, which is not desirable. Bistatic shadowing factors were also calculated by Wagner⁶³ and Smith⁸².

Many reflection models have been developed in the areas of computer graphics and machine vision. Since the current report focuses on model-derivation methodology, we will not review each existing model in detail but will just briefly discuss their commonalities. The existing reflection models may be classified according to the derivation approaches (see Table 1), namely, using the empirical approach⁶⁸, ⁸³-⁸⁸, Fraunhofer diffraction⁸⁰, Kirchhoff-Beckmann approach⁸⁹-⁹⁵, and microfacet method⁷⁴, ⁷⁵, ⁸¹, ⁹⁶-⁹⁸. Using an approach similar to the microfacet method, Oren and Nayar⁹⁹ proposed a non-Lambertian model assuming each facet Lambertian instead of perfectly specular. Most models adopt the assumption that the total reflection is decomposed into the diffuse and specular terms. One exception is the model proposed by He et al.⁹⁰, where the total reflection consists of three terms corresponding to the uniform diffuse, directional diffuse and specular contributions, respectively.

As discussed in Section 2.1, the decomposition of the total reflection into the diffuse and specular terms lacks justification. In the strict sense, models that adopt the rigid decomposition are not entirely physically based. Also, in
many modeling processes, physically-based arguments or calculations are mixed with empirical treatments, and the
distinctions and boundaries between the physically-based and empirical components are not specified clearly. For
example, the shadowing factor derived by Torrance and Sparrow\textsuperscript{33} was used in a number of reflection models, but
this shadowing factor only applies to a surface composed of V-shaped facets, not to generic rough surfaces. Finally,
existing models tend to be limited to particular surface types or scattering effects instead of emphasizing the
methodology for the systematic derivability of reflection model and the generalization of required conditions.

3 Statistical Ray Method

In this report, we present a statistical ray method for deriving analytic BRDFs of rough surfaces. The method is
established on two assumptions: the tangent plane approximation and the statistical approach sufficiency. The first
assumption validates the description of waves using geometric rays. The second assumption validates the use of
statistical concepts and techniques. A number of other assumptions are involved to characterize surface types, but
they are not fundamental to the ray statistical method.

For convenience of discussion, we define a new concept called surface microarea. A surface microarea $\delta A$ is a
surface region of size with dimension of $\tau$, where $\tau$ is the surface correlation length (see Appendix A). A surface
microarea is much smaller than the local illumination area $\Delta A$ on which a BRDF is defined. Figure 3 shows the
relationships between an entire object, a local illumination area $\Delta A$, and a surface microarea $\delta A$. A surface
microarea does not necessarily mean that its size must be on microns. The key point is that $\delta A$ is much smaller than
$\Delta A$. For example, if the dimension of $\Delta A$ is one the scale of $10^{-3}$ meters, then the dimension of $\delta A$ could be $10^{-5}$
meters (this means that $\Delta A$ is $10^4$ times of $\delta A$).

![Figure 3: (a) An object with a marked local illumination area $\Delta A$. (b) The surface profile within $\Delta A$ and a
marked microarea $\delta A$. (c) The enlarged view of $\delta A$. The size of the object is much larger than $\Delta A$, which is
much larger than $\delta A$.](image)

**Assumption 1 (tangent plane approximation):** Any surface microarea $\delta A$ has size much larger than wavelength $\lambda$
and is sufficiently smooth such that $\delta A$ can be replaced with the local tangent plane.

Consider an incident plane wave on a surface microarea $\delta A$. Because $\delta A$ can be replaced with the tangent plane
and the size of $\delta A$ is much larger than $\lambda$, according to Huygens’s principle\textsuperscript{41}, the reflected wave remains a plane
wave. This validates the description of the reflected waves using geometric rays. Since a plane wave corresponds to
a geometric ray traveling in the direction of wave propagation, the reflected wave corresponds to a geometric ray
traveling in the mirror-reflection direction with respect to the local tangent plane. In addition, Fresnel’s law (see
Appendix C) applies to the local reflection at $\delta A$. Therefore, in our discussion below, we will describe waves using geometric rays.

From now on, we will focus on the case of light reflection. Note, however, that the following calculation and discussion equally apply to other waves as long as Assumption 1 holds.

In essence, the tangent plane approximation stated in Assumption 1 is equivalent to the tangent plane approximation used by Beckmann and Spizzichino in their analytic calculation.\textsuperscript{15} Note that in Beckmann and Spizzichino’s calculation, the purpose of the tangent plane approximation was to validate Fresnel’s law at the surface boundary. However, in this report, the tangent plane approximation validates not only Fresnel’s law, but also the ray description. Moreover, the tangent plane approximation in Assumption 1 is more generic than the assumption used in Torrance-Sparrow’s microfacet model\textsuperscript{73} where the surface is assumed comprised of microfacets. In the real world, a naturally rough surface with smooth varying surface profile (such as shown in Figure 3) is not a collection of flat facets, but can still satisfy Assumption 1.

**Assumption 2 (statistical approach sufficiency):** Any local illumination area $\Delta A$ contains a large number of surface microareas. As a result, it is valid to use the concept of probability of microareas within $\Delta A$.

Assumption 2 validates the use of statistical concepts and techniques to calculate surface reflection. This is very important because natural surfaces are typically rough and their profiles are often only known statistically, instead of deterministically\textsuperscript{24-26, 43}. When a local illumination area $\Delta A$ contains a large number of microareas $\delta A$, to evaluate the entire reflection from $\Delta A$, the statistical approach is not only convenient but also sufficient.\textsuperscript{100-103}

Determining the sizes of $\Delta A$ and $\delta A$ often requires considering the spatial resolution of the detector as well as the viewing distance. As an example, consider the case of light reflection with a human eye as the detector. If $A_m$ is the minimal resolvable area of the eye\textsuperscript{41}, then $\Delta A$ should be chosen not smaller than $A_m$ because the surface details below the size of $A_m$ cannot be recognized by the eye. Let $D$ be the viewing distance (Figure 4). The subtended angle of $A_m$ with respect to the viewpoint is the minimal resolvable angle $\alpha_m$, which is about 1 minute of arc.\textsuperscript{41} Thus the diameter of $A_m$ is $\alpha_m D$. Since the eye has the best acuity when the object is located at about $D = 25$ cm in front of the eye, the minimal resolvable length $l_m$ is estimated to be 0.084 mm and $A_m$ is about $l_m^2$. If we choose the local illumination area $\Delta A$ with the same size as $A_m$, a microarea $\delta A$ may have the size of $10 \lambda$, which is roughly 0.7 microns. Similar analysis applies to other cases of surface reflection and detector.
Figure 4: The minimal resolvable angle $\alpha_m$, length $l_m$, and area $\alpha_m D$. Angle $\beta$ is small if the object is located in front of the eye.

Many natural surfaces (such as metal objects), although appearing rough macroscopically, are sufficiently smooth on the scale of $10\lambda$. Therefore, Assumptions 1 and 2 are valid. However, there are exceptions. When a surface contains structures of size comparable to $\lambda$, the tangent plane approximation is no longer valid. One example is a compact disk’s surface which has tracks and pits of size on microns. In this case, Assumption 1 fails and surface reflection must be handled using waves instead of geometric rays.

Now we start to present the basic equations. Let $l$ and $v$ be the unit vectors for the incident (lighting) and outgoing (viewing) directions (see Figure 1). Given $l$ and $v$, only a portion of the surface in the local illumination area $\Delta A$ can reflect directly from the incident solid angle $ld\omega$ into the outgoing solid angle $vd\omega$. If a microarea $\delta A$ reflects from $d\omega_i$ into $d\omega_r$, the surface normal of $\delta A$ must equally subdivide $l$ and $v$ (Figure 5) and it is given by

$$h = \frac{l+v}{2\cos \alpha} \tag{3.1}$$

where $2\alpha$ is the angle between $l$ and $v$. We call $h$ the halfway vector of $l$ and $v$.

![Diagram of angle and vectors](image)

Figure 5: The surface normal $h$ of a microarea $\delta A$ equally subdivides the lighting direction $l$ and the viewing direction $v$.

Applying Assumption 1 to $\delta A$, the incident and reflected radiances (see Appendix A) for $\delta A$ are related as

$$L_v(\theta, \phi) = \bar{F}(\alpha, \lambda)L_i(\theta, \phi)\delta(h, v, l) \tag{3.2}$$

$\bar{F}(\alpha, \lambda)$ is the Fresnel coefficient at incident angle $\alpha$ and is averaged for unpolarized light or over all polarizations of the incident light (see Appendix C). $\delta(h, v, l)$ is a Dirac delta function defined as

$$\delta(h, v, l) = \begin{cases} 1, & \text{when } h = \frac{l+v}{2\cos \alpha} \\ 0, & \text{otherwise} \end{cases} \tag{3.3}$$

Using Eq. (A.5), the outgoing radiant power from $\delta A$ can be written as

$$\delta\Phi(\theta, \phi) = L_i(\theta, \phi) \cos \alpha \delta A d\omega_i = \bar{F}(\alpha, \lambda)L_i(\theta, \phi)\delta(h, v, l) \cos \alpha \delta A d\omega_i. \tag{3.4}$$
To obtain the total reflected power for the local illumination area $\Delta A$, we should include contributions from all microareas within $\Delta A$ (note that $\Delta A$ contains a large number of microareas). However, not every microarea with normal $\mathbf{h}$ will actually contribute, because some might be blocked by the surface. We can describe this self-shadowing effect using a visibility function $V(\Delta A)$ that represents the probability of a microarea $\delta A$ visible in both incident and viewing directions. Note that it is valid to use the probability concept because the illumination area $\Delta A$ contains a large number of microareas $\delta A$ (Assumption 2). Thus, the total reflected power can be written as

$$\Phi_{\Delta A}(\theta, \phi) = L(\theta, \phi) \cos \alpha d\omega \sum_{\Delta A} \bar{F}(\alpha, \lambda) V(\Delta A) \delta(h, v, l) \delta A .$$

(3.5)

**Assumption 3** (surface homogeneity): The surface properties remain the same across a local illumination area $\Delta A$. These properties include the physical aspect such as the optical constants of the material (see Appendix C) as well as the geometric aspect such as the statistics as well as statistical parameters of the surface profile (see Appendix B).

From Eq. (A.5) in Appendix A and using Eq. (3.5), the radiance for $\Delta A$ is

$$L_{\Delta A}(\theta, \phi) = \Phi_{\Delta A}(\theta, \phi) = \frac{L(\theta, \phi) \cos \alpha F(\alpha, \lambda) d\omega}{\Delta A \cos \theta, d\omega} \sum_{\Delta A} V(\Delta A) \delta(h, v, l) \delta A .$$

(3.6)

In Eq. (3.6), $F(\alpha, \lambda)$ appears outside the summation because $\alpha$ is independent of the surface location and the surface material remains the same within $\Delta A$ (Assumption 3). Applying Eq. (3.6) to the BRDF definition in Eq. (1.1), we obtain

$$\rho_{\text{single}}(\theta, \phi, \theta, \phi, \lambda) = \frac{\cos \alpha F(\alpha, \lambda)}{\Delta A \cos \theta, d\omega} \sum_{\Delta A} V(\Delta A) \delta(h, v, l) \delta A .$$

(3.7)

Note that $\rho_{\text{single}}(\theta, \phi, \theta, \phi, \lambda)$ is the BRDF due to single scattering at a rough surface.

**Assumption 4** (height-field surface): The surface profile is a height field. That is, for any line parallel with the z-axis, the line will intersect with the surface profile exactly one time.

This assumption is true for most natural surfaces except those with cavities or pores. Now let $(\Delta A)_{\perp}$ be the projected area of $\delta A$ on the xy-plane. Then,

$$(\Delta A)_{\perp} = \delta A \cos \theta,$$

(3.8)

where $\theta$ is the polar angle of the half vector $\mathbf{h}$, which is the surface normal of $\delta A$ (see Figure 5). Because the surface is a height field (Assumption 4), $(\Delta A)_{\perp} / \Delta A$ is the probability of a surface point with normal $\mathbf{h}$, and this probability depends on partial derivatives $\zeta_x'$ and $\zeta_y'$ of $\delta A$, that is,

$$\frac{1}{\Delta A} \sum_{\Delta A} \delta(h, v, l)(\Delta A)_{\perp} = d\zeta_x' d\zeta_y' \int p(\zeta, \zeta_x', \zeta_y') d\zeta ,$$

(3.9)

where $p(\zeta, \zeta_x', \zeta_y') d\zeta d\zeta_x' d\zeta_y'$ is the probability of a surface point with height in $[\zeta, \zeta + d\zeta]$ and partial derivatives in $[\zeta_x', \zeta_x' + d\zeta_x']$ and $[\zeta_y', \zeta_y' + d\zeta_y']$, and we call $p(\zeta, \zeta_x', \zeta_y')$ as the combined probability density function (for more information refer to Ogilvy’s book). From $p(\zeta, \zeta_x', \zeta_y')$ we can obtain the individual probability density functions of variables $\zeta$, $\zeta_x'$, and $\zeta_y'$ as

$$p(\zeta) = \int p(\zeta, \zeta_x', \zeta_y') d\zeta_x' d\zeta_y' ,$$

(3.10)

$$p(\zeta_x') = \int p(\zeta, \zeta_x', \zeta_y') d\zeta d\zeta_y' ,$$

(3.11)

$$p(\zeta_y') = \int p(\zeta, \zeta_x', \zeta_y') d\zeta d\zeta_x' ,$$

(3.12)

where $p(\zeta)$ is the probability density function of surface height $\zeta$, $p(\zeta_x')$ the probability density function of $\zeta_x'$, and $p(\zeta_y')$ the probability density function of $\zeta_y'$.

Combining Eqs. (3.7)-(3.9), we obtain...
\[ \rho_{\text{single}}(\theta, \phi, \theta_h, \phi_h, \lambda) = \frac{\cos \alpha F(\alpha, \lambda)}{\cos \theta_h \cos \theta \cos \theta} d\zeta d\zeta' \int d\zeta p(\zeta, \zeta', \zeta') V(\zeta, I, V). \]  

(3.13)

The above integral is over \( \zeta \) because the contributing microareas may have different heights. Since \( \zeta' \) and \( \zeta'' \) are variables independent of \( \zeta \), \( d\zeta' d\zeta'' \) is put outside of the integral. In addition, visibility function \( V(\delta A) \) is replaced with \( V(\zeta, I, V) \) to show explicitly the dependent parameters. \( V(\zeta, I, V) \) is a bistatic probability and it is interpreted as a surface point at height \( \zeta \) being visible in directions \( I \) and \( V \) simultaneously.

Because both \( h \) and \( (\zeta', \zeta'') \) specify the orientation of a microarea \( \delta A \), there must be a relationship between \( h \) and \( (\zeta', \zeta'') \). Let the surface profile be described by

\[ \zeta = h(x, y), \]  

(3.14)

and then the partial derivatives \( \zeta_x' \) and \( \zeta_y' \) are

\[ \zeta_x' = \frac{\partial h(x, y)}{\partial x}, \quad \zeta_y' = \frac{\partial h(x, y)}{\partial y}. \]  

(3.15)

For any coordinate combination \( (x, y) \), i.e. a point on the xy-plane, the corresponding point on the surface profile is

\[ r = (x, y, h(x, y)). \]  

(3.16)

If we define a new function

\[ f(x, y, z) = z - h(x, y), \]  

(3.17)

then the surface normal \( h \) (toward the positive z-direction) is given by

\[ h = \frac{\nabla f}{|\nabla f|}, \]  

(3.18)

where

\[ \nabla f(x, y, z) = \left( \frac{\partial f(x, y, z)}{\partial x}, \frac{\partial f(x, y, z)}{\partial y}, \frac{\partial f(x, y, z)}{\partial z} \right) = (-\zeta_x', -\zeta_y', 1). \]  

(3.19)

Let us express \( d\zeta' d\zeta'' \) in terms of solid angle \( d\omega \). Consider \( \delta A \) as shown in Figure 6. Let \( \theta_h \) and \( \phi_h \) be the polar and azimuthal angles of normal \( h \), and let \( \theta_s \) and \( \phi_s \) be the polar and azimuthal angles for the gradient vector \( s \) (which has the maximum slope of \( \delta A \)). Considering that \( s \) and \( h \) are perpendicular and that \( s \), \( h \), and \( z \) are coplanar, the angle between \( s \) and the xy-plane is \( \theta_s = \theta_h \), and the azimuthal angle of \( s \) is \( \phi_s = \phi_h + \pi \). Therefore, the surface gradient value is

\[ \zeta' = |\nabla h(x, y)| = \left| \frac{\partial h(x, y)}{\partial R_s} \right| = \tan \theta_s, \]  

(3.20)

where \( R_s = [x^2 + y^2]^{1/2} \) is the distance along OB (OB is the projection of \( s \) on the xy-plane). Because

\[ \zeta_x' = \frac{\partial h(x, y)}{\partial x} = \frac{\partial R_s \partial h(x, y)}{\partial x} = \cos \theta_s \zeta', \]  

(3.21)

and

\[ \zeta_y' = \frac{\partial h(x, y)}{\partial y} = \frac{\partial R_s \partial h(x, y)}{\partial y} = \sin \phi_s \zeta', \]  

(3.22)

Eqs (3.21) and (3.22) are equivalent to

\[ \begin{cases} \zeta_x'^2 + \zeta_y'^2 = \zeta'^2 \\ \zeta_y'/\zeta_x' = \tan \phi_s \end{cases} \]  

(3.23)

Therefore,

\[ d\zeta' d\zeta'' = \zeta' d\zeta'' d\phi_s = \tan \theta_s d(\tan \theta_s) d\phi_s = \frac{\sin \theta_s d\phi_h d\phi_s}{\cos^3 \theta_h} = \frac{d\omega_h}{\cos^3 \theta_h}, \]  

(3.24)

where \( d\omega_h = \sin \theta_h d\theta_h d\phi_h \) is the differential solid angle along direction \( h \). Moreover, as derived from Appendix D,
Substituting Eqs. (3.24) and (3.25) into Eq. (3.13),

\[
\rho_{\text{single}}(\theta, \varphi, \theta', \varphi', \lambda) = \frac{\mathcal{F}(\alpha, \lambda)}{4 \cos \theta \cos \theta'} \int d\zeta p(\zeta, \zeta', \zeta') V(\zeta, 1, v).
\] (3.26)

**Assumption 5** (separable combined probability): The combined probability function can be approximated as a product of the individual probability functions of all involved variables.

Specifically, Assumption 5 implies that

\[
p(\zeta, \zeta', \zeta'') = p(\zeta) p(\zeta') p(\zeta''),
\] (3.27)

where the individual probability density functions have been defined in Eqs. (3.10)-(3.12). Substituting Eq. (3.27) into (3.26), we obtain

\[
\rho_{\text{single}}(\theta, \varphi, \theta', \varphi', \lambda) = \frac{\mathcal{F}(\alpha, \lambda) p(\zeta') p(\zeta'')}{4 \cos \theta \cos \theta'} V(1, v),
\] (3.28)

where

\[
V(1, v) = \langle V(\zeta, 1, v) \rangle = \int d\zeta p(\zeta) V(\zeta, 1, v)
\] (3.29)

is the bistatic visibility function averaged over height (see Eq. (C.5) in Appendix C). Eq. (3.28) is a **generic equation** when Assumptions 1-4 hold. From Eq. (3.28), we can derive BRDFs when specific surface properties are given, as will be shown in Sec. 5. Note that \( \theta_a \) in Eq. (3.28) can be determined by a dot product of unit vector \( z \) to Eq. (3.1)

\[
\cos \theta_a = \frac{\cos \theta + \cos \theta'}{2 \cos \alpha} = \frac{\cos \theta + \cos \theta'}{\sqrt{2 + \cos(2\alpha)}},
\] (3.30)

where \( \alpha \) is given by

\[
\cos 2\alpha = 1 \cdot v = \sin \theta \sin \theta' \cos(\varphi - \varphi') + \cos \theta \cos \theta'.
\] (3.31)

### 4 Visibility Functions
4.1 Individual Visibility Function

Our strategy is to calculate the individual visibility function of the one-dimensional case and from it we will obtain the result for the two-dimensional case. Consider a one-dimensional surface profile as shown in Figure 7. Because the surface is a height field (Assumption 4), the surface profile can be described by

\[ \zeta = h(x). \]  (4.1)

Let a ray start from a surface point of height \( \zeta_1 \) and propagate with polar angle \( \theta \). Imagine that we move the start point of the ray along the surface profile while the ray direction is fixed; then sometimes the ray is blocked by the surface and sometimes not, depending on the location of the ray start point. We define the individual visibility function \( V(\zeta, d) \) as the probability of a ray that starts from height \( \zeta \) and travels in direction \( d \) without blocking by the surface.

![Figure 7: A ray starts from \( x = 0 \) at height \( \zeta_1 \) at polar angle \( \theta \). The height at distance \( x \) is \( \zeta_2 \).](image)

Let \( T(x) \) be the probability for a ray to travel over distance \( x \) without blocking (see Figure 7). The change of \( T(x) \) in interval \([x, x + dx]\) can be written as

\[ dT(x) = T(x + dx) - T(x) = -kT(x)Q(x)dx. \]  (4.2)

The minus sign means a decrease in \( T(x) \), parameter \( k \) is positive and independent of \( x \) and \( \theta \), and

\[ Q(x) = \text{Prob}(h(x) > \zeta_1 + x\cot\theta) \]  (4.3)

is the probability that the surface height \( \zeta_2 = h(x) \) at \( x \) is higher than \( \zeta_1 + x\cot\theta \), which is the z value of ray at \( x \).

Note that in Eq. (4.2) the change of \( T(x) \) is proportional to \( T(x)Q(x) \) instead of \( Q(x) \) only because the ray must be first visible before reaching the interval \([x, x + dx]\). \( Q(x) \) can be written as

\[ Q(x) = \int_{\zeta_1 + x\cot\theta}^{\infty} p(\zeta_2 | \zeta_1) d\zeta_2, \]  (4.4)

where \( p(\zeta_2 | \zeta_1) \) is a conditional probability for a surface point with height \( \zeta_2 = h(x) \) at \( x \) given that the surface height is \( \zeta_1 = h(0) \) at \( x = 0 \). Because \( p(\zeta_2 | \zeta_1) \) is a conditional probability,

\[ p(\zeta_2 | \zeta_1) = \frac{p(\zeta_1, \zeta_2)}{p(\zeta_1)}, \]  (4.5)

where \( p(\zeta_1) \) is one-point probability density and \( p(\zeta_1, \zeta_2) \) is two-point probability density. From Eq. (4.2),

\[ d\ln T(x) = \frac{dT(x)}{T(x)} = -kQ(x)dx. \]  (4.6)

Since \( T(0) = 1 \) (the ray is not blocked at the ray start point),
Now we define
\[ V(\zeta, d, l) = \exp \left[ -k \int_0^l Q(x) dx \right] \] (4.8)
as the probability of a ray starting at surface height \( \zeta \) and traveling in direction \( d \) over distance \( l \) without blocking. Thus, the individual visibility function \( V(\zeta, d) \) is
\[ V(\zeta, d) = \lim_{l \to \infty} V(\zeta, d, l) = \exp \left[ -k \int_0^\zeta Q(x) dx \right]. \] (4.9)

Eqs. (4.1)-(4.9) are valid for all surfaces as long as they are height fields (Assumption 4).

Note that the formulations of Eqs. (4.1)-(4.9) appear similar to those occurring in the calculations of Wagner\(^{63}\) and Smith\(^{64}\). However, there are two essential differences. The first difference is that our fundamental equation, Eq. (4.2), contains a constant \( k \). This is not only mathematically correct, but also important for satisfying the scaling invariance principle (see Sec. 5.3 and Appendix F). Since \( dT(x)/dx \) is proportional to \( T(x) \) and \( Q(x) \), to make the equality in Eq. (4.2), the introduction of constant \( k \) is necessary. As the second difference, our calculation has considered the correlation between surface heights at coordinates \( x \) and \( 0 \), but this correlation was ignored in the calculations of Wagner and Smith.

4.2 Combined Visibility Functions

To obtain the analytic expression of \( V(l, v) \) that appears in Eqs. (3.28) and (3.29), one follow a similar idea of the calculation in Sec. 4.1 for a two-dimensional surface. However, the calculation is very complicated. In this report, we take a simple approach by obtaining \( V(l, v) \) using the individual visibility function obtained in Sec. 4.1. If we neglect the correlation between \( l \) and \( v \), then
\[ V(l, v) = V(\theta_l)V(\theta_v), \] (4.10)
where \( V(\theta_l) \) and \( V(\theta_v) \) are the averaged individual visibility functions for \( l \) and \( v \) given by
\[ V(\theta) = V(d) = \int d\zeta \rho(\zeta) V(\zeta, d). \] (4.11)
However, it is important to note that \( l \) and \( v \) may have a considerable correlation. For example, in the case when \( l \) and \( v \) coincide, if a surface point is visible for direction \( l \), the point is also visible in direction \( v \). Therefore, Eq. (4.10) is an approximation, which is stated in Assumption 6 below. Note, however, that this assumption can be easily removed when a thorough calculation for the bistatic visibility function is available.

Assumption 6 (negligible directional correlation): The correlation between the incident and outgoing visibility functions is ignored.

5 Isotropic Surfaces with Gaussian Statistics

5.1 Additional Conditions

As an example of applying the statistical ray method to a specific surface type, now we focus on an isotropic surface with Gaussian statistics. Besides Assumptions 1-6, we assume that the surface also satisfies the following conditions:

A. The surface is isotropic.
B. The height probability density is Gaussian.
C. The height correlation function is Gaussian.

Condition A implies that the surface parameters are independent of the horizontal direction (or the azimuthal angle). Imagine that we cut \( \Delta A \) along two different horizontal directions; although the cross-sectional profiles might be
different, the values of the statistical parameters (such as the height deviation and correlation length) will be the
same. On Condition B, when a surface height field is generated purely by a single random process, the height
probability density is Gaussian. Many natural surfaces belong to this type. However, there are considerable
exceptions\textsuperscript{24, 25} because the factor of generating a rough surface may not be purely random. Such examples include
the surfaces generated by engineering methods such as turning, or by engraving of liquids where the surface height
is decreased preferentially toward the negative to create cracks and valleys. On Condition C, for a randomly rough
surface, the correlation function for a surface profile can be described with different analytic expressions and a
Gaussian function is often sufficient (see Appendix B).\textsuperscript{43} Also, Condition C is related to the validity of Assumption
5 (see Appendix E).

5.2 Gaussian Height Field

Using Condition B, the surface height field is described by
\begin{equation}
p(\zeta) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{\zeta^2}{2\sigma^2}\right) = \frac{1}{\sigma} g\left(\frac{\zeta}{\sigma}\right),
\end{equation}
where $\sigma$ is the standard deviation and $g(t)$ is the standard Gaussian function (see Appendix B). From statistics, if a
random variable $\zeta$ is Gaussian, it is known that the derivative $\zeta'$ is also Gaussian with deviation\textsuperscript{15}
\begin{equation}
\sigma' = \frac{\sqrt{2\sigma}}{\tau},
\end{equation}
where $\tau$ is the surface correlation length. In other words, the probability densities for $\zeta'$ and $\zeta''$ are
\begin{equation}
p(\zeta') = \frac{1}{\sqrt{2\pi\sigma'}} \exp\left(-\frac{\zeta'^2}{2\sigma'^2}\right) = \frac{1}{\sigma'} g\left(\frac{\zeta'}{\sigma'}\right)
\end{equation}
and
\begin{equation}
p(\zeta'') = \frac{1}{\sqrt{2\pi\sigma''}} \exp\left(-\frac{\zeta''^2}{2\sigma''^2}\right) = \frac{1}{\sigma''} g\left(\frac{\zeta''}{\sigma''}\right).
\end{equation}
Therefore,
\begin{equation}
p(\zeta') p(\zeta'') = \frac{1}{2\pi\sigma''} \exp\left(-\frac{\zeta'^2}{2\sigma'^2}\right) = \frac{\tau^2}{4\pi\sigma^2} \exp\left(-\pi^2 \tan^2 \theta / 4\sigma^2\right)
\end{equation}
where $\tan \theta = \zeta' = \left(\zeta'^2 + \zeta''^2\right)^{1/2}$ (refer to Eq. (3.20)). Applying Eq. (5.5) to Eq. (3.28), we obtain
\begin{equation}
\rho_{\text{angle}}(\theta_1, \phi_1, \theta_2, \phi_2, \lambda) = \frac{\tau^2 F(\alpha, \lambda) \exp\left(-\pi^2 \tan^2 \theta / 4\sigma^2\right)}{16\pi^2 \cos \theta_1 \cos \theta_2 \cos^2 \theta_2} V(1, v).
\end{equation}

5.3 Visibility Function

For a Gaussian height field with deviation $\sigma$ and correlation function $C(x)$, we know that the two-point probability
density has the following form\textsuperscript{15, 43, 104}
\begin{equation}
p(\zeta_1, \zeta_2) = \frac{1}{\sqrt{2\pi\sigma^2 \sqrt{1-C^2}}} \exp\left[-\frac{\zeta_1^2 - 2\zeta_1 \zeta_2 C + \zeta_2^2}{2\sigma^2(1-C^2)}\right].
\end{equation}
Substituting Eq. (5.7) into Eq. (4.5), we obtain
\begin{equation}
p(\zeta_1 \mid \zeta_2) = \frac{1}{\sqrt{2\pi\sigma \sqrt{1-C^2}}} \exp\left[-\frac{(\zeta_2 - \zeta_2 C)^2}{2\sigma^2(1-C^2)}\right].
\end{equation}
Substituting into Eq. (4.4),
\begin{equation}
Q(x) = \frac{1}{\sqrt{2\pi\sigma \sqrt{1-C^2}}} \int_{\zeta_2 + \cos \theta}^{\phi} \exp\left[-\frac{(\zeta_2 - \zeta_2 C)^2}{2\sigma^2(1-C^2)}\right] d\zeta_2 = 1 - G(w(x)),
\end{equation}
where function $G$ is given in Eq. (B.11) in Appendix B and
\[
\begin{align*}
\text{w}(x) & \equiv \frac{\zeta_1(1-C) + x\cot\theta}{\sigma\sqrt{1-C^2}}. \\
\text{(5.10)}
\end{align*}
\]

From Eq. (4.4), integrating \(Q(x)\) in parts,
\[
\int_0^x Q(x)dx = xQ(x) - \int_0^x xdQ(x),
\]
and for a Gaussian height field,
\[
\int_0^x Q(x)dx = x[1-G(w(x))] - \int_0^x x(1-G(w(x))) = x[1-G(w(x))] + \int_{w_0}^w x(w)g(w)dw,
\]
where we have used \(dG(t)/dt = g(t)\) (see Appendix B) and we define
\[
w_i \equiv w(0) = \tau\cot\theta/\sqrt{2\sigma},
\]
which is the value of \(w(x)\) at \(x = 0\) from Eq. (5.10).

From Sec. 4.1, the individual visibility function is \(T(x)\) when \(x \to \infty\)
\[
V(\zeta_1, \theta) = T(x)\bigg|_{x \to \infty}.
\]

According to Eq. (5.10), when \(x \to \infty\), \(w(x) \to \infty\). Also, when \(x \to \infty\), the first term on the right hand side of Eq. (5.12) vanishes, because \(1-G(w)\) approaches 0 faster than \(x\) approaches the infinity. Thus Eq. (5.12) becomes
\[
\int_0^x Q(x)dx = \int_{w_0}^w x(w)g(w)dw,
\]
where the integral is performed over variable \(w\), and \(x\) is regarded as a function of \(w\). Thus,
\[
V(\zeta_1, \theta) = \exp\left[-k\int_{w_0}^w x(w)g(w)dw\right].
\]

The expression in Eq. (5.16) depends on \(\zeta_1\) via \(x(w)\) and is rigorous as long as Condition B holds. The averaged individual visibility function is
\[
\tilde{V}(\theta) = \left\langle V(\zeta_1, \theta) \right\rangle = \int V(\zeta_1, \theta)p(\zeta_1)d\zeta_1.
\]

To obtain an explicit form of \(\tilde{V}(\theta)\), we adopt approximations. The first approximation is
\[
\tilde{V}(\theta) \approx V(\zeta_1 = 0, \theta).
\]

which means that the averaged individual visibility is approximately equal to the individual visibility function for \(\zeta_1 = 0\). The consideration is that \(V(\zeta_1, \theta)\) is small for negative height and large for positive height, because a ray starting from a valley is more likely to be blocked and from a hill more likely to be unblocked. This consideration appears rather crude, but the final result it leads to turns out agreeing with numerical simulation very well (see Sec. 6.3). In our discussion below, we adopt approximation of Eq. (5.18) simply because it easily leads to a simple explicit expression that allows straightforward analytic analysis to reveal helpful insights of the reflection behavior of rough surfaces.

When \(\zeta_1 = 0\), Eq. (5.10) becomes
\[
\text{w}(x) = \frac{x\cot\theta}{\sigma\sqrt{1-C^2}}.
\]

For large \(x\), \(C \to 0\) and
\[
w \to \frac{x\cot\theta}{\sigma}.
\]

Therefore, for large \(x\), we can write
\[
x = \sigma w \tan \theta.
\]

If we use Eq. (5.21) to finish the integral in Eq. (5.16), the result is
\[
\int_0^\infty Q(x)dx = \sigma \tan \theta \int_{w_0}^\infty w g(w)dw = \frac{\sigma \tan \theta}{\sqrt{2\pi}} \exp(-w_0^2/2).
\]

(5.22)

The replacement of Eq. (5.10) with (5.21) corresponds to approximating the curve of \( w(x) \) using its asymptotic line (Figure 8). Combining Eqs. (5.16), (5.18), and (5.22),

\[
V(\theta) = \exp \left[ -\frac{k \sigma \tan \theta}{\sqrt{2\pi}} \exp(-\tau^2/4\sigma^2 \tan^2 \theta) \right].
\]

(5.23)

Figure 8: Relations between \( w \) and \( x \). The curve of \( w(x) \) is approximated with the line \( x = \sigma w \tan \theta \) for \( w \geq w_0 \).

From the scaling invariance principle (Appendix F), \( k \) must be inversely proportional to \( \tau \) and can be written as

\[
k = \frac{2\pi k_0}{\tau},
\]

(5.24)

where \( k_0 \) is a positive constant. The value of \( k_0 \) can be determined by comparing the expression of Eq. (5.23) with the result from experiment or numerical simulation of a Gaussian rough surface. We leave this issue to be handled in Sec.6. Thus, \( V(\theta) \) can be written as

\[
V(\theta) = \exp \left[ -\frac{k_0 \sigma \tan \theta}{\tau} \exp(-\tau^2/4\sigma^2 \tan^2 \theta) \right].
\]

(5.25)

or

\[
V(\theta) = \exp \left[ -\frac{k_0 \tan \theta}{s} \exp(-s^2/4 \tan^2 \theta) \right].
\]

(5.26)

where \( s \) is defined as the surface smoothness parameter

\[
s = \frac{\tau}{\sigma}.
\]

(5.27)

Correspondingly, the surface roughness parameter is

\[
r = \frac{s}{\sigma} = \frac{1}{\tau}.
\]

(5.28)

Parameters \( s \) and \( r \) are non-negative. Large \( s \) or small \( r \) (i.e. large \( \tau \) and small \( \sigma \)) implies a smooth surface, and small \( s \) or large \( r \) (i.e. small \( \tau \) and large \( \sigma \)) implies a rough surface. For a perfectly smooth surface, \( s \to \infty \) and \( r = 0 \).
It is important to note that the concept of *surface roughness* has different meanings in literature of wave scattering at rough surfaces, where a surface is regarded as rough when its characteristic length is comparable with the wavelength with respect to a given incident or outgoing direction. In the current report, however, the entire framework is established upon the tangent plane approximation, which requires that the surface is sufficiently smooth locally with respect to the wavelength. Among the real surfaces satisfying the tangent plane approximation, many surfaces appear very rough in the macroscopic scales. Therefore, in this report, it is more appropriate to define the surface smoothness parameter as given in Eq. (5.28). The definition of Eq. (5.28) has also an advantage that the surface smoothness or roughness is an inherent property of the surface, independent of the incident wavelength or relevant directions. As matter of fact, the BRDF we have derived is a function of the surface smoothness parameter $s$ or roughness parameter $r$. As shown in the discussion below, an important result we obtain is that it is parameter $s$ or $r$ (instead of $\sigma$ and $\tau$ independently) that directly affects the surface reflection behavior.

### 5.4 Summary

Using the surface smoothness parameter $s$ defined in Eq. (5.27) and substituting Eq. (4.10) into Eq. (5.6),

$$
\rho_{\text{angle}}(\theta_l, \varphi_l, \theta_v, \varphi_v, \lambda) = \frac{s^2 \tilde{F}(\alpha, \lambda) \exp(-s^2 \tan^2 \theta_l / 4)}{16\pi \cos \theta_l \cos \theta_v \cos^2 \theta_h} V(\theta) V(\theta_t).
$$

(5.29)

Eq. (5.29) is the derived BRDF for single scattering at an isotropic rough surface that has a Gaussian height field and a Gaussian correlation function of the height field. Recall that $\theta_l$ and $\varphi_l$ are the polar and azimuthal angles of the lighting direction, $\theta_v$ and $\varphi_v$ the polar and azimuthal angles of the viewing direction, and $\lambda$ is wavelength. $\theta_h$ is the polar angle of the halfway vector $h$ and can be determined by

$$
\cos \theta_h = \frac{\cos \theta_l + \cos \theta_v}{2 \cos \alpha} = \frac{\cos \theta_l + \cos \theta_v}{\sqrt{2 + \cos(2\alpha)}},
$$

(5.30)

where $\alpha$ is given by

$$
\cos 2\alpha = 1 \cdot v = \sin \theta_l \sin \theta_v \cos(\varphi_l - \varphi_v) + \cos \theta_l \cos \theta_v.
$$

(5.31)

Also, from Eq. (5.26), the individual visibility function is

$$
V(\theta) = \exp \left[ -\frac{k_0 \tan \theta}{s} \exp(-s^2 / 4 \tan^2 \theta_v) \right],
$$

(5.32)

where $k_0$ is a positive constant to be determined in Sec. 6.3.

It is helpful to rewrite Eq. (5.29) in the following form:

$$
\rho_{\text{angle}}(\theta_l, \varphi_l, \theta_v, \varphi_v, \lambda) = \tilde{F}(\alpha, \lambda) \chi(\theta_l, \varphi_l, \theta_v, \varphi_v) D(\theta_h) V(\theta_l) V(\theta_t).
$$

(5.33)

$D(\theta_h)$ is the probability density function of a surface microarea whose normal has polar angle $\theta_h$

$$
D(\theta_h) = \frac{s^2 \exp(-s^2 \tan^2 \theta_h / 4)}{4\pi \cos^3 \theta_h}.
$$

(5.34)

Note that the surface is isotropic and thus $D(\theta_h)$ is independent of the azimuthal angle $\varphi_h$, and

$$
\int D(\theta_h) d\omega_h = \frac{s^2 \exp(-s^2 \tan^2 \theta_h / 4)}{4\pi \cos^3 \theta_h} \sin \theta_h d\theta_h d\varphi_h = \frac{s^2}{4} \int \exp(-s^2 \tan^2 \theta_h / 4) d\tan^2 \theta_h = 1.
$$

(5.35)

Moreover, at the limit of a perfectly smooth surface, i.e. $s \to \infty$,

$$
D(\theta_h) \to \frac{1}{2\pi} \delta(1 - \cos \theta_h).
$$

(5.36)

In Eq. (5.33), function $\chi(\theta_l, \varphi_l, \theta_v, \varphi_v)$ is called the normalization factor defined as

$$
\chi(\theta_l, \varphi_l, \theta_v, \varphi_v) = \frac{1}{4\cos \theta_l \cos \theta_v \cos \theta_h},
$$

(5.37)

which is necessary to normalize the BRDF to satisfy the energy conservation law (see Sec. 6.2).
6 Discussion

6.1 Extreme Cases of Surface Smoothness

For a very smooth surface, $s \to \infty$, $V(\theta) \to 1$, the derived BRDF in Eq. (5.29) reduces to the following form

$$
\rho_{\text{single}}(\theta, \varphi, \varphi_i, \lambda) = \frac{s^2 F(\alpha, \lambda)}{16\pi \cos \theta_i \cos \theta \cos^4 \theta_i} \exp(-s^2 \tan^2 \theta_i / 4),
$$

which is dominated by the exponential term. Recall that $\theta_i$ is the polar angle of the halfway vector $h$ (see Figure 5).

Since $s \to \infty$, the exponential term drops very quickly when $\theta_i$ increases from 0, meaning that the BRDF has significant values only at very small $\theta_i$. In other words, there is a strong highlight occurring in the condition of $\theta_i = 0$. This is the condition when $h$ is parallel with $z$, i.e. $\theta_i = \theta_i = \varphi_i = \pi \pm \varphi_i$, which is the mirror-reflection condition. Also, when $s$ is large, from Eq. (6.1), the reflection intensity at the highlight is proportional to $s^2$.

On the other hand, for a very rough surface, $s \to 0$, $V(\theta) \to 1$, the derived BRDF is dominated by the product of the visibility functions

$$
\rho_{\text{single}}(\theta, \varphi, \varphi_i, \lambda) = \exp\left(-\frac{k_s \tan \theta_i}{s}\right) \exp\left(-\frac{k_s \tan \theta_i}{s}\right).
$$

Thus the BRDF tends to have a maximum at $\theta_i = 0$ or $\theta_i = 0$. But remember that the current BRDF only includes contribution from single scattering. To describe the full reflection behavior, multiple scattering must be included for a very rough surface.

6.2 Reciprocity and Energy Conservation

Surface reflections should satisfy the principle of Helmholtz reciprocity\(^{40,41}\). Consider path 2-2' in Figure 2. If a light ray impinges on the surface from the direction of 2', then the corresponding outgoing ray should be along the direction of 2. If we change the incident direction from $l$ to $v$, the light path reverses exactly the original path. This is the principle of Helmholtz reciprocity. Applying this principle to BRDFs, the impact is that the value of a BRDF should remain the same when directions $l$ to $v$ are switched. Obviously this is true for Eq. (5.29) because the expression is symmetric about $l$ and $v$.

On energy conservation, consider that the incident light has a uniform radiance $L_i$ in a small solid angle $\Delta\omega_i$ in lighting direction $l$. From Eq. (A.1) in Appendix A, the incident radiant power is

$$
\Phi_i = L_i \cos \theta_i \Delta\omega_i.
$$

By using Eq. (1.1), the outgoing radiance in direction $(\theta_i, \varphi_i)$ is

$$
L_i(\theta_i, \varphi_i) = \int L_i = \int \rho_{\text{single}}(\theta, \varphi, \varphi_i, \varphi_i, \lambda) L_i \cos \theta_i d\omega_i = \rho_{\text{single}}(\theta, \varphi_i, \varphi_i, \varphi_i, \lambda) L_{\text{e}} \cos \theta_i \Delta\omega_i.
$$

Thus the total outgoing radiant power from $\Delta A$ is

$$
\Phi_v = \Delta A \int L_i(\theta_i, \varphi_i, \lambda) \cos \theta_i d\omega_i = \Delta A L_{\text{e}} \cos \theta_i \Delta\omega_i \int \rho_{\text{single}}(\theta, \varphi_i, \varphi_i, \varphi_i, \lambda) \cos \theta_i d\omega_i.
$$

The principle of energy conservation requires that

$$
\frac{\Phi_v}{\Phi_i} \leq 1.
$$

Taking the ratio of Eq. (6.5) to (6.3), the energy conservation principle implies that

$$
\int \rho_{\text{single}}(\theta, \varphi_i, \varphi_i, \varphi_i, \lambda) \cos \theta_i d\omega_i \leq 1.
$$

If we substitute Eq. (5.29) into (6.7), the integral cannot be completed in general cases. However, it is possible to verify Eq. (6.7) in the two extreme cases. When $s \to \infty$ (a perfectly smooth surface), $D(\theta_i) \to \delta(1 - \cos \theta_i)/2\pi$, 

22
Suppose that \( F(\alpha, \lambda) = 1 \) (completely reflecting surface and no absorption). Using Eqs. (5.32), (5.35), (5.36), and (3.25),

\[
\int \rho_{\text{angle}}(\theta, \varphi, \iota, \varphi, \lambda) \cos \theta d\omega_\theta \leq \frac{1}{4 \cos \theta_1} \int \delta(1 - \cos \theta_1) / 2\pi \cos \theta_1 \cos \theta_\lambda d\omega_\theta = \frac{1}{2\pi \cos \theta_1} \int \frac{\cos \alpha \delta(1 - \cos \theta_\lambda)}{\cos \theta_\lambda} d\omega_\theta = \frac{1}{\cos \theta_1} \cos \theta_\lambda = 1.
\]

(6.8)

Note that \( \cos \theta_\lambda = 1 \) for non-zero of delta function \( \delta(1 - \cos \theta_\lambda) \) corresponds to the condition \( \cos \alpha = \cos \theta_1 \). In Eq. (6.8), the equality holds for a surface that is perfectly smooth and perfectly reflecting.

On the other hand, when \( s \to 0 \), from Eq. (6.2),

\[
\int \rho_{\text{angle}}(\theta, \varphi, \iota, \varphi, \lambda) \cos \theta d\omega_\theta \propto \int \exp\left(-\frac{k_0 \tan \theta}{s}\right) \sin \theta d\theta \to 0,
\]

(6.9) because \( \exp[-k_0 \tan \theta / s] \) behaves like a Dirac delta function \( \delta(\theta) \) as \( s \to 0 \). Therefore the energy conservation condition is satisfied in both extreme cases.

### 6.3 Visibility Function

As mentioned earlier, the individual visibility function \( V(\theta) \) we have derived in Eq. (5.26) involves a constant \( k_0 \) whose value needs to be determined. Now we will use the simulation result of Brockelman and Hagfors\(^{62}\) to determine the value of \( k_0 \). In Brockelman and Hagfors’s approach, they first generated one-dimensional rough surfaces with desired Gaussian height deviations and Gaussian correlation lengths, and then computed the shadowing functions (which corresponds to the average individual visibility function in this report) for surface smoothness with values of 1, 2, 4 and 10. Our approach is to plot \( V(\theta) \) given in Eq. (5.26) for various values of \( k_0 \) and select the value that gives the best match to the data of Brockelman and Hagfors. Figure 9 shows \( V(\theta) \) given in Eq. (5.26) for surface smoothness \( s \) with values of 1, 2, 4 and 10. For each of these values of \( s \), we plot the curves of \( V(\theta) \) for \( k_0 \) equal to 0.5, 0.6, 0.7, 0.8 and 1. In all cases, \( k_0 = 0.7 \) seems to have the best fit with Brockelman and Hagfors’s data, which are displayed with points in diamond shape in Figure 9. Note that \( k_0 = 0.6 \), or any value between 0.6 and 0.7, would also have a good match with Brockelman and Hagfors’s data. But our focus here is on showing how to determine \( k_0 \) instead of on deciding the optimal value of \( k_0 \). Overall, \( V(\theta) \) given in Eq. (5.26) with \( k_0 = 0.7 \) matches the data of Brockelman and Hagfors very well. When \( s \) is 1 or 2 (rather rough surfaces), \( V(\theta) \) has a significant difference from Brockelman and Hagfors’s data when \( \theta \) is large, but the agreement is very good when \( \theta \) is below 60 degrees. Therefore we will use \( k_0 = 0.7 \) in the discussion below.
Figure 9: Determining the value of $k_0$ in visibility function $V(\theta)$ based on cases (a) $s = 1$, (b) $s = 2$, (c) $s = 4$, and (d) $s = 10$. In each case the solid curves are $V(\theta)$ for $k_0$ equal to 0.5, 0.6, 0.7, 0.8 and 1, and the points in the diamond shape are from Brockelman and Hagfors's computer simulation. In all cases, $V(\theta)$ with $k_0 = 0.7$ matches Brockelman and Hagfors's data very well.

Given the surface smoothness parameter $s$, when $\theta$ increases, the visibility function decreases monotonically. Figure 10 plots the curves of $V(\theta)$ given in Eq. (5.26) with $k_0 = 0.7$ for different values of $s$. Every curve decreases monotonically from 1 to 0 as $\theta$ increases. When $s$ is small (rough surfaces), $V(\theta)$ drops quickly. On the other hand, when $s$ is large (smooth surfaces), $V(\theta)$ is almost 1 except when $\theta$ is close to 90 degrees. As the extreme cases, $V(0) = 1$ and $V(90^\circ) = 0$. These agree with the intuition that a ray upward vertically will never be blocked and a horizontal ray starting at $\zeta = 0$ will always be blocked.

Note that the rigorous solution of Wagner agrees very well with the computer simulation of Brockelman and Hagfors. However, Wagner’s analytic expression is rather complicated, which reduces its value for intuitive analysis or practical application. The approximate expressions of Wagner and Smith are simplified in the forms, but still quite complex, involving undesired error functions. In contrast, our derived visibility function given in Eq. (5.26) is entirely explicit and has a much simpler form. From our deriving process, our approach involves an integration by parts and an approximation of function $w(x)$, and uses the scaling invariance principle and numerical data to determine constant $k_0$. Although the calculations of Wagner and Smith generate the visibility functions directly (i.e. without the need of constant fitting), their approaches seem to be effective only to Gaussian surfaces. If the surface has a non-Gaussian statistics, it might be impossible to complete the integral directly. In this regard, we believe that our approach (using integration by parts and approximating the integrand) has a better potential to handle surfaces with generic statistics.
Figure 10: Curves of the visibility function $V(\theta)$ with $k_o = 0.7$ against the polar angle for different values of surface smoothness $s$, which are shown near the corresponding curves.

6.4 Probability Density of Surface Orientation

Eq. (5.34) defines $D(\theta_d) d\omega_d$, which is the probability of surface microareas with normal in differential solid angle $d\omega_d$ in direction $(\theta_d, \phi_d)$ (see Figure 5 and Figure 6). As shown in Eq. (5.34), the probability density $D(\theta_d)$ is normalized over the hemisphere. Since the surface we consider is isotropic, the probability density is independent of the azimuthal angle $\phi_d$ and depends only on the polar angle $\theta_d$.

The probability density $D(\theta_d)$ contains $\exp(-s^2 \tan^2 \theta_d / 4)$ in the numerator and $\cos^3 \theta_d$ in the denominator. The exponential term has large values at small $\theta_d$, while the cosine term contributes large values at large $\theta_d$. In $D(\theta_d)$, the exponential term and the cosine term compete, and the dominating term depends on the surface smoothness parameter $s$. Figure 11 plots the curves of $D(\theta_d)$ for different values of $s$. When $s \leq 1$ (rough surfaces), the cosine term dominates and $D(\theta_d)$ shows a peak at a large $\theta_d$. This may be interpreted as a rough surface has a high probability of surface microareas of normals with large $\theta_d$. On the other hand, when $s > 4$ (smooth surfaces), the exponential term dominates and $D(\theta_d)$ shows a peak when $\theta_d = 0$. This means that a smooth surface has a high probability of surface microareas of normals with small $\theta_d$. In Figure 11, the curves for intermediate values of $s$, i.e. $1 \leq s \leq 4$, show how $D(\theta_d)$ varies gradually from the case of rough to smooth surfaces. In particular, it is interesting to note that $D(\theta_d)$ for $s = 2$ has a plateau over a significant range of $\theta_d$ (from 0 to 45 degrees).
6.5 Behavior of the Derived BRDF

Now we analyze the derived BRDF given in Eq. (5.29) for various values of surface smoothness. We first focus on the behavior of reflection within the incident plane. Let the incident direction be $(\theta, \pi)$. Thus the incident plane is formed by the x and z-axes. For convenience of discussion, we define the slope angle $\gamma$ as the angle between the x-axis and the viewing direction $v(\theta, \phi)$ (Figure 12). The range of $\gamma$ is from 0 to $\pi$, and the relationship with the polar angle is

$$
\theta_v = \begin{cases} 
\pi / 2 - \gamma, & \text{when } \phi = 0 \\
\gamma - \pi / 2, & \text{when } \phi = \pi.
\end{cases}
$$

(6.10)
Figure 12: The relationship between the slope angle and the corresponding polar angle within the incident plane. Note that the range of the slope angles is from 0 to $\pi$.

Figure 13 plots the derived BRDF in Eq. (5.29) against the slope angle $\gamma$, for various values of the surface smoothness parameter $s$. The incident direction is specified by $\theta_i = \pi/4$ and $\phi_i = \pi$, corresponding to $\gamma_i = 3\pi/4$. The effect of Fresnel coefficient $F(\alpha, \lambda)$ is excluded, that is, the material properties are not involved in our current discussion. From the curves in Figure 13, we observe four reflection regimes of the derived BRDF. When $s > 10$ (very smooth surfaces), the BRDFs show a clear peak near $\gamma = \pi/4$; this is the mirror-reflection regime. When $3 < s < 10$ (moderately smooth surfaces), the BRDFs have a peak at an outgoing polar angle larger than that for the mirror-reflection; we call this case grazing reflection regime. (This seems to be related to the off-peak specular reflection discovered in early research.) Note that the curve for $s = 2$ is flat over a large range of the outgoing slope angle, and it is interesting to recall that the BRDF for a Lambertian surface is constant. When $0.5 < s < 3$ (moderately rough surfaces), the BRDFs tend to have strong reflection when the viewing direction is close to the lighting direction; thus we call this regime the retro-reflection regime. When $s < 0.5$ (very rough surfaces), the BRDFs have a peak near the normal direction of the surface mean plane (i.e. $z = 0$), and therefore we call the regime normal reflection regime. The key finding here is the presence of the four distinct reflection regimes. The specific values of $s$ that divide the regimes are not important, as the transitional values of $s$ depend on the incident angle. Also, note that the BRDFs shown in Figure 13 account only for single scattering, and they may differ significantly from the entire reflection behavior if the surface is considerably rough (under which condition multiple scattering is important).
Figure 13: The derived BRDF excluding $\bar{F}(\alpha, \lambda)$ against the slope angle $\gamma$, for various values of the surface smoothness parameter $s$. The incident direction is $\gamma_i = 3\pi / 4$. The curves show how the BRDF changes as $s$ increases.

The four reflection regimes are caused by the competition among the terms in the BRDF expression in Eq. (5.29). When parameter $s$ is very large, $\exp(-s^2 \tan^2 \theta_s / 4)$ dominates such that the BRDF has a strong peak at $\theta_s = 0$, which corresponds to the mirror-reflection condition. When $s$ is decreased but is still moderately large, $\cos \theta_s$ in the denominator has a considerable impact, favoring forward reflection at an angle larger than the mirror-reflection angle. When $s$ is decreased further, $4 \cos \theta_s$ will have a dominating effect favoring reflection at large $\theta_s$, which is the case of retro-reflection. In addition, $\cos \theta_i$ in the denominator favors that the maximum occurs at a large outgoing angle. Finally, when $s$ is very small, the visibility function $V(\theta_i)$ dominates such that the BRDF has significant values only when $\theta_i$ is very small (refer to Figure 10).

Figure 14 plots the BRDF expression in Eq. (5.29), with exclusion of $\bar{F}(\alpha, \lambda)$, as a function of the outgoing direction $(\theta_i, \varphi_i)$. The incident direction is given by $\theta_i = \pi / 4$ and $\varphi_i = \pi$. The polar angle $\theta_i$ varies from 0 to 90 degrees and the azimuthal angle $\varphi_i$ from -90 to 270 degrees (this choice of the interval for $\varphi_i$ helps to present the entire BRDF peak occurring at $\varphi_i = 0$). In Figure 14a, $s = 0.2$, which is in the normal reflection regime, and the BRDF values are significant only at very small outgoing polar angles. In Figure 14b, $s = 1$, which is the case of retro-reflection regime. Note that the BRDF has a peak roughly at a polar angle of 70 degrees and azimuthal angle of 180 degrees (refer to the curve for $s = 1$ in Figure 13b). This means that strong back-scattering occurs. It is also interesting to note that the BRDF has relatively smaller values when the azimuthal angle is 0, implying that the forward reflection is minimized. In Figure 14c, $s = 4$, the case of grazing reflection regime. We can see that the BRDF maximum occurs at 0 azimuthal angle and a polar angle greater than 45 degrees (refer to Figure 13c). Finally, in Figure 14d, $s = 10$, which is in the mirror-reflection regime, where the BRDF has a peak roughly at 0 degrees of the azimuthal angle and 45 degrees of the polar angle.
Figure 14: The derived BRDF excluding $F(\alpha, \lambda)$ as a function of the outgoing direction $(\theta, \phi)$. The incident direction is given by $\theta_i = \pi/4$ and $\phi_i = \pi$. (a) $s = 0.2$ (normal reflection regime). (b) $s = 1$ (back-reflection regime). (c) $s = 4$ (grazing reflection regime). (d) $s = 10$ (mirror-reflection regime).

7 Concluding Remarks

This report has developed a statistical ray method to derive analytic reflection models of rough surfaces. The fundamental assumptions of the method are the tangent plane approximation and the statistical approach sufficiency. The tangent plane approximation validates the description of waves using geometric rays. The statistical approach sufficiency validates the calculation of the reflected energy power using statistical concepts such as the surface height probability and correlation function. The calculation in this report has focused on optical reflection, but the method and result apply equally to other waves as long as the required assumptions are satisfied.

Using the statistical ray method, we have obtained generic equations for single scattering at an opaque rough surface under various conditions. The very initial equation is Eq. (3.5), which gives the reflected energy based on the
tangent plane approximation and the statistical approach sufficiency. Furthermore, with the assumptions of surface homogeneity and height field, the BRDF for single scattering is expressed in terms of an integral over the surface profile height, as given in Eq. (3.26). Finally, the assumption of separable combined probability leads to Eqs. (3.28) and (3.29). These equations can be used as the starting equations to calculate BRDFs for various specific types of surfaces.

As an example of applying the statistical ray method, starting from Eqs. (3.28) and (3.29), we have calculated the BRDF due to single scattering for an isotropic surface with Gaussian statistics. The analytic results are summarized in Eqs. (5.29)-(5.32). Our numerical study has found that the derived BRDF demonstrates four distinct reflection regimes, namely, the regimes of mirror reflection, grazing reflection, retro-reflection, and normal reflection, as the surface varies from being perfectly smooth to very rough. These reflection regimes are results of the competition among the terms in the BRDF expression. Eq (5.29), in spite of accounting only for single scattering, may be used to describe the entire reflection at an opaque surface that is highly or moderately smooth. This equation also offers a useful basis for calculating reflection due to multiple scattering as potential future work.

A key component of this report is the analytic study of visibility function, the probability of a surface point visible with respect to a given direction. We have presented a generic expression of the individual visibility function for surfaces of various statistics, and obtained an explicit analytic form for a surface with Gaussian height field and Gaussian correlation function. A scaling invariance principle has been introduced to complete the determination of the visibility function. With a similar accuracy as verified with the simulation of Brockelman and Hagfors, our calculation has several advantages over previous studies of Wagner and Smith. The expressions of Wagner and Smith are quite complex, involving undesired error functions, but our derived visibility function, given in Eq. (5.26), has a much simpler form and does not involve the error functions. Our result is therefore more useful and convenient for reflection analysis or other practical applications. Regarding the deriving process, our approach involves an integration by parts and an approximation of function $\psi(x)$, and uses the scaling invariance principle and numerical simulation to determine a dimensionless constant. Although the calculations of Wagner and Smith generate the visibility functions directly, their approaches seem effective only for Gaussian surfaces, because completing the integral might be impossible for non-Gaussian surfaces. In contrast, our approach (using integration by parts and approximation of the integrand) has a potential to handle surfaces with generic statistics. Finally, our approach has considered the correlation between the surface height at different locations on the surface, while this correlation has been ignored in the previous studies.

The method and calculation presented in this report open a number of future research directions. First, several further studies may be conducted on visibility functions. In this report, we determined constant $k_0$ by comparing the analytic expression with numerical simulation, but it is still highly desirable to determine $k_0$ analytically, independent of numerical data. Besides, our calculation has used an approximation of function $\psi(x)$ that is defined in Eq. (5.10), and it is worth studying to improve the approximation. Moreover, although it is helpful to verify with the data of Brockelman and Hagfors, other independent data from experiment or computer simulation would be valuable to verify the analytic results independently. Also, it is necessary to improve the calculation for the bistatic visibility function. The studies of Wagner and Smith provide a useful basis for obtaining an expression that is analytically manageable but still sufficiently accurate. Finally, it is interesting to explore for solution of the visibility functions for non-Gaussian surfaces.

Another major direction for the future work is associated with the potential extensions of the assumptions and conditions used in this report. For example, the application example in Sec. 5 in this report requires that the surface is isotropic and the surface height probability density and correlation function are Gaussian. If one removes the condition of surface isotropy, the BRDF for anisotropic surfaces may be derived starting from Eqs. (3.28) and (3.29). This extension would need two correlation lengths in two orthogonal directions along the surface mean plane (i.e. $z = 0$), and the bistatic visibility function needs to be calculated corresponding to the anisotropic surface. Starting from Eqs. (3.28) and (3.29), the extension from Gaussian to non-Gaussian surfaces is also possible, but it could be challenging to obtain the explicit forms for the orientation probability and visibility function. Besides, this report focuses on the case of an opaque surface, but the BRDF expressions for single scattering equally apply to a surface of transparent material and one only needs to use the Fresnel coefficient corresponding to the transparent material. Finally, the derivation process in this report can be easily extended to obtain illumination models of transmission.
The only key difference one needs to make in the model-deriving process is to replace the mirror-reflection principle for local reflection with the Snell’s law for local transmission.

In this report, we have not calculated the BRDF due to multiple scattering. As indicated by the visibility functions shown in Figure 10, when the surface smoothness is below 2 (moderately or highly rough surfaces), the surface self-shadowing effect is strong. Under the same condition (the surface smoothness is below 2), multiple scattering is not negligible. However, a serious calculation of the BRDF due to multiple scattering could be very difficult. One possible approach is to calculate the BRDF due to double-scattering processes first. This may lead to a useful idea to handle scattering processes of higher orders. Another study might be doing numerical simulation to reveal the behavior of multiple scattering processes, and hopefully the simulation can offer a clue to handle multiple scattering. Approximations are unavoidable to calculate the BRDF for multiple scattering. The key issue is to find the appropriate and effective approximations.

Although it could be very difficult to obtain the BRDF for multiple scattering, it is not difficult to estimate the overall reflected flux (in the sense of the numbers of rays) due to the multiple-scattering processes. This is because we are able to calculate the total flux for single scattering, which is complementary to the flux for multiple scattering. The total flux for single scattering can be obtained by integrating overall the hemisphere the BRDF for single scattering with exclusion of the Fresnel coefficient. With the knowledge of the total flux for multiple scattering, some simple forms of the BRDF for multiple scattering may be assumed. For example, if we assume that the flux for multiple scattering is uniformly distributed over all outgoing directions, the BRDF for multiple scattering would only involve the Fresnel coefficient in power because of multiple bounces of the involved processes. Certainly, this simple assumption could be verified by numerical simulation.

With the BRDF for multiple scattering in place, a full reflection model is obtained. Such a model could be used in application such as computer graphics to improve the realism of image synthesis. However, such a model might not be sufficient (especially when the surface is rough) for machine vision because the application demands high accuracy to achieve reliable determination. When applying the results derived in this report, it is important to verify the conditions carefully. On the tangent plane approximation, metallic surfaces usually satisfy this assumption because a metallic surface is typically smooth enough at the microscopic scale (although appearing rough at macroscopic scales). However, the tangent plane approximation may not hold for many natural surfaces (such as surfaces of clay or wood). Also, it is also important to verify whether or not the surface has Gaussian statistics.

Our study has been focusing on optical reflection, but the method and calculation apply equally to other waves as long as the required assumptions are satisfied. Therefore, it is interesting to test the method in scattering problems where the involved waves are not optical, for example, electromagnetic or acoustic waves. In such applications, it is important to note that the concepts of surface smoothness or roughness in this report have different meanings from those used in literature of wave scattering, where a surface is regarded as rough when its characteristic length is comparable with the wavelength. In our report, the entire framework is established upon the tangent plane approximation, which requires that the surface is sufficiently smooth locally with respect to the wavelength. Therefore, in this report, the surface smoothness is defined as the ratio between the correlation length and the surface height deviation, as given in Eq. (5.27). From our definition, smoothness or roughness is regarded as an inherent property of the surface, independent of the incident wavelength or relevant directions.

It is important to evaluate the method and calculation results presented in this report. Experimental measurement is certainly the most objective way for the evaluation. However, conducting the experiments is very challenging. This is because one should measure not only the optical reflection as function of the incident and outgoing directions, but also the surface physical and statistical parameters including the Fresnel coefficient, and the surface height deviation and correlation length. Also, one needs to verify that the tested surface satisfies the required assumptions including the tangent plane approximation, statistical approach sufficiency, surface homogeneity, height field, surface isotropy, Gaussian surface statistics, and so on. In addition, we have only derived the BRDF of single scattering, and in practical measurement it is difficult to separate single scattering from multiple scattering. Considering these factors, numerical simulation would be valuable for verifying the method because one can specifically compute the result for single scattering. In such effort, for the same reason as pointed in the beginning of Section 2.1, it would be better to generate the required micro-area probability distributions or associated parameters from deliberate BRDF data than starting from surface statistics and pristine ideal height fields that produce the BRDFs. Finally, a method or model can also be evaluated by examining the involved analytic process for the model derivation. This analytic derivation
process offers the clues for identifying the sources of errors, as well as when and why the errors are significant or negligible.

For the problem of reflection at rough surfaces, a long-term goal is to develop a generic theoretical framework to solve the problem systematically. The theory should be built upon physical principles, and the involved assumptions should be justified carefully. The theory should be generic, capable of handling different surface cases. Assumptions and approximations should be clearly stated so that the validities of the conditions can be traced and verified. Thus the sources of errors can be tracked and analyzed, and better reflection models can be achieved by generalizing the conditions and improving the approximations. The current report is an endeavor toward this goal.
# Appendices

## A. Radiometric Concepts

Here is a brief review of radiometric concepts related to this report. Refer to literature\(^1\),\(^2\),\(^28\) for further information.

- **Radiant energy**, denoted by \( Q \), is the emitted, transferred or received energy of radiation.
- **Radiant power** (or radiant flux), denoted by \( \Phi \), is the radiant energy per unit time:
  \[
  \Phi = \frac{dQ}{dt}.
  \]  
  (A.1)
- **Irradiance**, denoted by \( E \), is the radiant power incident on a unit surface area (not beam cross-sectional area):
  \[
  E = \frac{d\Phi}{dA}.
  \]  
  (A.2)
- **Exitance**, denoted by \( M \), is the radiant power leaving a unit surface area (not beam cross-sectional area):
  \[
  M = \frac{d\Phi}{dA}.
  \]  
  (A.3)
- **Radiant intensity**, denoted by \( I \), is the radiant power per solid angle:
  \[
  I = \frac{d\Phi}{d\omega}.
  \]  
  (A.4)
- **Radiance**, denoted by \( L \), is the radiant power per solid angle per projected surface area (or beam cross-sectional area):
  \[
  L = \frac{d^2\Phi}{dA^* d\omega} = \frac{d^2\Phi}{\cos\theta dA d\omega} = \frac{d^2\Phi}{dA d\omega^*}.
  \]  
  (A.5)

  where \( \theta \) is the outgoing angle, \( dA^* = \cos\theta dA \) is the project surface area of \( dA \) and \( d\omega^* = \cos\theta d\omega \) is the project solid angle of \( d\omega \). Combining Eqs. (A.5) and (A.2),

  \[
  L = \frac{dE}{\cos\theta d\omega}.
  \]  
  (A.6)

  Similarly, combining Eqs. (A.5) and (A.3),

  \[
  L = \frac{dM}{\cos\theta d\omega}.
  \]  
  (A.7)
- **Bi-directional reflectance distribution function** (BRDF) is the ratio between a differential reflected radiance \( dL_v(\theta, \varphi, \lambda) \) in the viewing (outgoing) direction and the incident irradiance \( L_i(\theta, \varphi, \lambda) \cos\theta d\omega_i \) in the lighting (incident) direction:

  \[
  \rho(\theta, \varphi, \theta_i, \varphi_i, \lambda) = \frac{\text{differential radiance}}{\text{differential irradiance}} = \frac{dL_v(\theta, \varphi, \lambda)}{dE_v(\theta, \varphi, \lambda)} = \frac{dL_v(\theta, \varphi, \lambda)}{L_i(\theta, \varphi, \lambda) \cos\theta_i d\omega_i}.
  \]  
  (A.8)

  where \( \theta, \varphi \) are the polar and azimuthal angles of the lighting direction, \( \theta_i \) and \( \varphi_i \) the polar and azimuthal angles of the viewing direction, and \( \lambda \) is wavelength. The BRDF definition was first introduced by Nicodemus and collaborators\(^1\),\(^47\) to describe surface reflections that are neither truly diffuse nor specular.

## B. Statistical Description of Rough Surfaces

Given a rough surface, the plane for the average surface height in a long range is called the **mean plane** or **ground plane**. There are two essential statistical aspects of a random rough surface: the spread of the surface height in the vertical direction (the \( z \)-direction) and the speed of height variation along a horizontal direction (in the \( xy \) plane). Various statistical distributions and parameters may be used to describe the surface profile, but they are basically...
equivalent. We will focus on the **height probability density** and **surface correlation function**, whose relevant parameters are illustrated in Figure 15.

**Figure 15**: Microscopic view of a rough surface and ray scattering. Parameters $\sigma$ and $\tau$ are the surface height deviation and correlation length.

Let $\zeta$ denote the height of any point on a rough surface relative to the ground plane. Mathematically, $\zeta$ may vary from $-\infty$ to $+\infty$. If we set up the coordinates such that the average surface height coincides with the $z = 0$ plane, the surface profile can be described by

$$\zeta = h(R) = h(x, y), \quad (B.1)$$

where $R = (x, y)$ is a point on the ground plane and $(x, y, \zeta)$ is a point on the surface profile.

In the statistical approach, the surface height is described by a **height probability density** $p(\zeta)$, which is positive and normalized, i.e.,

$$\int_{-\infty}^{+\infty} p(\zeta) d\zeta = 1, \quad (B.2)$$

and $p(\zeta) d\zeta$ is the probability of a surface point in interval $[\zeta, \zeta + d\zeta]$. The **distribution function** of $p(\zeta)$ is

$$P(z) = \int_{-\infty}^{z} p(\zeta) d\zeta, \quad (B.3)$$

which gives the probability of the height below $z$. Note that $p(\zeta)$ is the derivative of $P(z)$

$$p(\zeta) = \frac{dP(\zeta)}{d\zeta}. \quad (B.4)$$

Note that in some literature $p(\zeta)$ is called height probability distribution. But this name is easy to be confused with the distribution function $P(z)$. In this report, we call $p(\zeta)$ as **height probability density** and this convention is in agreement with most literature in statistics.

Given any function $f(\zeta)$, its average can be calculated from

$$\langle f(\zeta) \rangle = \int_{-\infty}^{+\infty} f(\zeta) p(\zeta) d\zeta, \quad (B.5)$$

where $\langle \cdots \rangle$ is the **spatial averaging** across the surface, equivalent to the average over an **ensemble** of surface profiles according to the **ergodic theorem**. Thus, the average height is given by

$$\zeta_0 = \langle \zeta \rangle = \int_{-\infty}^{+\infty} \zeta p(\zeta) d\zeta, \quad (B.6)$$

and for the coordinates we have set up, the average height is zero

$$\zeta_0 = 0. \quad (B.7)$$
The height probability density $p(\zeta)$ may have different analytic form, depending on a particular rough surface. If the height field is Gaussian, $p(\zeta)$ can be written as

$$p(\zeta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\zeta^2}{2\sigma^2}\right), \quad (B.8)$$

where $\sigma$ is the standard deviation or root mean square (RMS) of the surface height

$$\sigma = \sqrt{\left\langle \zeta^2 \right\rangle} = \left[ \int_{-\infty}^{\infty} \zeta^2 p(\zeta) d\zeta \right]^{1/2}. \quad (B.9)$$

For convenience of calculation, we use

$$g(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) \quad (B.10)$$

to denote the dimensionless standard Gaussian function, and let

$$G(t) = \int_{-\infty}^{t} g(t')dt' = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \exp\left(-\frac{t'^2}{2}\right)dt' \quad (B.11)$$

be the distribution function for $g(t)$. Thus,

$$p(\zeta) = \frac{1}{\sigma} g\left(\zeta / \sigma\right) \quad (B.12)$$

and

$$P(z) = \int_{-\infty}^{z} p(\zeta)d\zeta = G\left(z / \sigma\right). \quad (B.13)$$

It is easy to verify that

$$G(+\infty) = 1, \quad P(+\infty) = 1. \quad (B.14)$$

Two surfaces with the same standard deviation may have different degrees of roughness. To distinguish such cases, we need the surface correlation function

$$C(\mathbf{R}_1, \mathbf{R}_2) = \frac{\left\langle h(\mathbf{R}_1)h(\mathbf{R}_2)\right\rangle}{\sigma^2}, \quad (B.15)$$

where $\mathbf{R}_1$ and $\mathbf{R}_2$ are points on the ground plane (the xy-plane). When a surface is homogeneous, the correlation function depends only on the relative distance vector $\mathbf{R} = \mathbf{R}_2 - \mathbf{R}_1$. Thus the correlation function can be written as

$$C(\mathbf{R}) = \frac{\left\langle h(\mathbf{R})h(\mathbf{R} + \mathbf{R})\right\rangle}{\sigma^2}. \quad (B.16)$$

Furthermore, if a surface is isotropic, the correlation function is simplified to

$$C(\mathbf{R}) = \frac{\left\langle h(\mathbf{R})h(\mathbf{R})\right\rangle}{\sigma^2}, \quad (B.17)$$

where $R = |\mathbf{R}_2 - \mathbf{R}_1|$. As a special case, when $\mathbf{R}_1$ and $\mathbf{R}_2$ coincide

$$C(0) = 1. \quad (B.18)$$

If the surface is purely random, as $R$ increases, $C(R)$ will eventually decay to zero

$$C(\infty) = 0. \quad (B.19)$$

But $C(\infty) = 0$ does not apply to surfaces that are not completely random. For example, the correlation function of sinusoidal surface will have the form of a cosine function, reflecting the periodic nature of the surface.

Surface correlation functions can be described well with a Gaussian function

$$C(R) = \exp(-R^2 / \tau^2), \quad (B.20)$$

where $\tau$ is the correlation length. An alternative form is exponential

$$C(R) = \exp(-R / \tau). \quad (B.21)$$

Because of its continuity feature at the origin, the exponential form causes difficulties when considering higher-order properties such as surface derivatives. Surface roughness needs to be characterized using both standard
deviation $\sigma$ and correlation length $\tau$. Increasing $\sigma$ or decreasing $\tau$ will result in an increase in surface roughness. On the other hand, decreasing $\sigma$ or increasing $\tau$ will result in a decrease in the roughness.

C. Fresnel’s Coefficients and Optical Constants

At a perfectly smooth interface of two media, the incident and outgoing radiant intensities are related in terms of Fresnel formulas.\textsuperscript{40,41} If $R_\parallel$ and $R_\perp$ denote the ratios of the reflected intensity to the incident intensity for the polarizations that are parallel and perpendicular to the incident plane, then

$$
\begin{align*}
R_\parallel &= \frac{(n_2 \cos \theta_i - n_1 \cos \theta_t)^2}{(n_1 \cos \theta_i + n_2 \cos \theta_t)^2} \\
R_\perp &= \frac{(n_2 \cos \theta_i - n_1 \cos \theta_t)^2}{(n_1 \cos \theta_i + n_2 \cos \theta_t)^2}
\end{align*}
$$

where $n$ is the index of refraction of medium 2 relative to medium 1, and $\theta_i$ and $\theta_t$ are the incident and refractive angles. For an unpolarized wave, the Fresnel coefficient is averaged over the parallel and perpendicular polarizations $\mathcal{F} = (R_\parallel + R_\perp)/2$. (C.2)

$R_\parallel$, $R_\perp$ and $\mathcal{F}$ are called Fresnel’s coefficients of reflection. These coefficients depend on the incident angle, the wavelength, and the media on both sides of the surface.

For a generic material, the index of refraction is a complex number and depends on wavelength:

$$n(\lambda) = \eta(\lambda) + i\kappa(\lambda),$$

where $\eta(\lambda)$ is the simple index of refraction and $\kappa(\lambda)$ is the extinction coefficient. $\eta(\lambda)$ and $\kappa(\lambda)$ are real numbers and they are called optical constants. Fresnel coefficients can be calculated from\textsuperscript{28}

$$
\begin{align*}
R_\parallel &= \frac{a^2 + b^2 - 2a \cos \theta_i + \cos^2 \theta_t}{a^2 + b^2 + 2a \cos \theta_i + \cos^2 \theta_t} \\
R_\perp &= \frac{a^2 + b^2 - 2a \sin \theta_i \tan \theta_t + \sin^2 \theta_i \tan^2 \theta_t}{a^2 + b^2 + 2a \sin \theta_i \tan \theta_t + \sin^2 \theta_i \tan^2 \theta_t}
\end{align*}
$$

where

$$
\begin{align*}
2a^2 &= \sqrt{(\eta^2 - \kappa^2 - \sin^2 \theta_i)^2 + 4\eta^2 \kappa^2} + (\eta^2 - \kappa^2 - \sin^2 \theta_i)^2 \\
2b^2 &= \sqrt{(\eta^2 - \kappa^2 - \sin^2 \theta_i)^2 + 4\eta^2 \kappa^2} - (\eta^2 - \kappa^2 - \sin^2 \theta_i)^2
\end{align*}
$$

The spectra of $\eta(\lambda)$ and $\kappa(\lambda)$ are inherent properties of the material, and can be obtained from measurement or numerical calculation.\textsuperscript{28}

D. Relationship between Differential Solid Angles

The relationship between differential solid angles $d\omega_i$ and $d\omega_h$ can be found as follows. On the unit hemisphere, let spherical area $A_1B_1C_1D_1$ correspond to $d\omega_i$ and let $A_2B_2C_2D_2$ correspond to $d\omega_h$ (see Figure 16). Note that curve segments $A_1B_1$ and $A_2B_2$ are coplanar with line $OL$. Similarly, $C_1D_1$ and $C_2D_2$ are coplanar with line $OL$. In addition, line $OA_1$ equally subdivides angle $\angle A_1OL$, and line $OB_2$ equally subdivides angle $\angle B_2OL$. Therefore, we immediately obtain

$$\|A_1B_1\| = 2\|A_2B_2\|.$$ (D.1)
Let $\delta$ be the angle formed by the plane of triangle $\Delta A_1OL$ and the plane of triangle $\Delta D_1OL$. Since $\delta$ is small, $|B_2C_2| = \delta |B_2Q| = \delta \sin \alpha$ because $\angle B_2OL = \alpha$. Similarly, $|B_1C_1| = \delta \sin 2\alpha$ because $\angle B_1OL = 2\alpha$. Thus,

$$|B_1C_1| = 2 |B_2C_2| \cos \alpha$$  \hspace{1cm} (D.2)

Therefore, the area of $A_1B_1C_1D_1$ is

$$|A_1B_1| \times |B_1C_1| = 4 \cos \alpha |A_2B_2| \times |B_2C_2|,$$  \hspace{1cm} (D.3)

which is $4 \cos \alpha$ times of area of $A_2B_2C_2D_2$. This implies that

$$d\omega_v = 4 \cos \alpha d\omega_h.$$  \hspace{1cm} (D.4)

This relationship has been obtained in previous research\textsuperscript{73}.

![Figure 16: The relationship between differential solid angles $d\omega_v$ and $d\omega_h$.](image)

**E. Separable Combined Probability**

For a homogeneous and isotropic surface, it can be shown that that $\zeta$, $\zeta'$, and $\zeta''$ are mutually independent in $p(\zeta, \zeta', \zeta'')$. Consider the correlation between the height at $R_1$ and the height partial derivative in $x$ at $R_2$ ($R_1$ and $R_2$ are points in the xy-plane)

$$\langle \zeta(R_1)\zeta'(R_2) \rangle = \left\langle h(R_1) \frac{\partial h(R_2)}{\partial x} \right\rangle = \frac{\partial}{\partial x} \left\langle h(R_1)h(R_2) \right\rangle.$$  \hspace{1cm} (E.1)

Let

$$R = R_2 - R_1$$  \hspace{1cm} (E.2)

and

$$R = |R| = \sqrt{x^2 + y^2}.$$  \hspace{1cm} (E.3)

If the surface is homogeneous and isotropic, using Eq. (C.17), we have

$$\langle \zeta(R_1)\zeta'(R_2) \rangle = \frac{\partial}{\partial x} \sigma^2 C(R) = \sigma^2 \frac{\partial}{\partial x} C(R).$$  \hspace{1cm} (E.4)

If the correlation function is Gaussian,

$$\frac{\partial}{\partial x} C(R) = \frac{\partial R}{\partial x} \frac{\partial}{\partial R} C(R) = -\frac{2x}{\tau^2} \exp\left(-\frac{R^2}{\tau^2}\right),$$  \hspace{1cm} (E.5)

which approaches zero when $R \to 0$, i.e.,

$$\frac{\partial}{\partial x} C(R) \to 0 \text{ when } R \to 0.$$  \hspace{1cm} (E.6)
Combining Eqs. (E.1)-(E.6), we have
\[ \langle \zeta(R_1)\zeta'(R_2) \rangle \to 0 \text{ when } R_2 \to R_1, \tag{E.7} \]
This means that \( \zeta \) and \( \zeta' \) are not correlated. The same result applies to the correlation between \( \zeta \) and \( \zeta' \). That is,
\[ \langle \zeta(R_1)\zeta'(R_2) \rangle = \sigma^2 \frac{\partial}{\partial y} C(R) = -\frac{2y\sigma^2}{\tau^2} \exp\left(-\frac{R^2}{\tau^2}\right), \tag{E.8} \]
which approaches zero when \( R \to 0 \). So \( \zeta \) and \( \zeta' \) are not correlated. For the correlation between \( \zeta' \) and \( \zeta' \),
\[ \langle \zeta'(R_1)\zeta'(R_2) \rangle = \frac{\partial^2}{\partial x \partial y} \langle h(R_1)h(R_2) \rangle = \sigma^2 \frac{\partial^2}{\partial x \partial y} C(R) = \frac{4xy\sigma^2}{\tau^4} \exp\left(-\frac{R^2}{\tau^2}\right), \tag{E.9} \]
which vanishes when \( R \to 0 \). Therefore \( \zeta' \) and \( \zeta' \) are not correlated. Because \( \zeta' \), \( \zeta' \), and \( \zeta' \) are uncorrelated each other, their combined probability can be factored into the individual probabilities, that is,
\[ p(\zeta', \zeta', \zeta') = p(\zeta')p(\zeta')p(\zeta'). \tag{E.10} \]

F. Scaling Invariance Principle

To help deriving and verifying individual visibility functions, we propose the scaling invariance principle. This principle states that an individual visibility function should be a function of a dimensionless parameter
\[ \xi = \sigma \tan \theta / \tau. \tag{F.1} \]
where \( \sigma \) is the surface height deviation, \( \tau \) is the surface height correlation length, and \( \theta \) is the polar angle of the direction for the visibility function (see Fig. 8). From this principle, the same value of \( \xi \) should result in the same individual visibility function. Note that the visibility functions derived by Wagner\(^63\) and Smith\(^64\) satisfy this principle.

This principle is proved as follows. Imagine that we scale the z-axis by any positive constant \( a \). After this vertical scaling, \( \sigma \) becomes \( a\sigma \), \( \tan \theta \) becomes \( \tan \theta / a \), \( \tau \) is not changed, and \( \xi \) remains unchanged. Similarly, if we only scale the space horizontally, \( \sigma \) is not changed, \( \tan \theta \) becomes \( a\tan \theta \), \( \tau \) becomes \( a\tau \), and \( \xi \) remains unchanged. On the other hand, although the vertical or horizontal scaling changes the surface profile and ray direction, it will not change the state whether or not a ray intersects with the surface. That is, if a ray intersects with the surface before the scaling, the ray will still intersect the surface after the scaling. The same is true for a ray not intersecting with the surface. Overall the individual visibility function remains unchanged after the scaling. This is possible if and only if the individual visibility function is a function of \( \xi \).
9 References