Accurately Predicting the Reflectance of Rough Metal Surfaces From One-Dimensional Surface Profile Measurements

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

Citation
Zhu, Yuanchen. 2018. Accurately Predicting the Reflectance of Rough Metal Surfaces From One-Dimensional Surface Profile Measurements. Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences.

Citable link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:41121229

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA
Accurately Predicting the Reflectance of Rough Metal Surfaces from One-Dimensional Surface Profile Measurements

A dissertation presented
by
Yuanchen Zhu
to
The School of Engineering and Applied Sciences
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Computer Science

Harvard University
Cambridge, Massachusetts
August 2018
Abstract

This thesis investigates the problem of using surface microgeometry measurements to predict the reflectance of rough metal surfaces. Because light cannot penetrate much into metals, the observable reflectance of a metal surface is due almost entirely to interfacial reflection, i.e., light reflection at the interface between the metal and air. According to the laws of optics, interfacial reflection is determined completely by surface topography and material refractive index. Hence this thesis seeks to accurately predict the reflectance of metals, as observed and measured by a gonioreflectometer, from topographic measurements and material refractive index alone.

In order to achieve accurate prediction, it is vital to acquire accurate topographic measurements as input. This thesis identifies why modern profilometers based on white light interferometry (WLI) and atomic force microscopy (AFM) are unlikely to measure the two-dimensional topography of real-world rough surfaces with sufficient fidelity to allow accurate reflectance prediction. A new experimental procedure for accurately measuring the one-dimensional surface profiles using AFM is consequently proposed.

This thesis then derives how reflectance of a two dimensional surface is related to one-dimensional surface profiles, resulting in an ill-posed linear inverse problem. Two algorithms are subsequently designed for solving this inverse problem: The
first algorithm introduces and employs the *shaped microfacet model*, an extension to the classic microfacet model under the Kirchhoff approximation, to provide the necessary priors for tackling the inverse problem using optimization; The second algorithm employs the *Fourier slice theorem* to construct a solution to the inverse problem in its Fourier domain using simple interpolation. Both algorithms require no back-fitting of free parameters to reflectance measurements, thereby allowing reflectance computation from profilometer measurements to be truly predictive.

This thesis also experimentally evaluated the proposed measurement procedures and computational algorithms on surface samples manufactured via *metallic coating* using off-the-shelf *leafing* aluminum flakes and polyurethane binder. The proposed methods, including both algorithms, achieved much higher prediction accuracy than demonstrated before for non-precision-fabricated real-world two-dimensionally rough surfaces.
# Contents

Abstract ................................................................. iii
Acknowledgments ......................................................... xii

1 Introduction ........................................................... 1
  1.1 Contributions .................................................. 8
  1.2 Limitations .................................................... 10
  1.3 Organization .................................................. 11

2 Related Work ........................................................ 13
  2.1 Bidirectional Reflectance Distribution Function ............... 14
    2.1.1 Definition ............................................... 14
    2.1.2 BRDF Acquisition ....................................... 16
    2.1.3 BRDF Modeling .......................................... 17
  2.2 Microfacet Theory ............................................. 19
    2.2.1 Torrance and Sparrow’s Formulation .................... 19
    2.2.2 Introduction of Microfacet Theory to Computer Graphics . 21
    2.2.3 Zoo of Microfacet-Based BRDF Models .................. 21
  2.3 Kirchhoff Approximation ...................................... 22
    2.3.1 Kirchhoff Approximation and Beckmann-Kirchhoff Theory .. 23
    2.3.2 Physical Optics in Early Computer Graphics .............. 23
    2.3.3 Zoo of Kirchhoff-Approximation-Based BRDF Models ....... 24
  2.4 Experimental Validation of Reflectance Models ............... 25
    2.4.1 Fitting Models to Measured Reflectance ................ 26
    2.4.2 Reflectance Fabrication ................................ 28
  2.5 Reflectance Prediction from Profilometer Measurements ....... 29
    2.5.1 Strictly Gaussian Surfaces ............................. 31
    2.5.2 Modeling Arbitrary Surfaces as Parametric Random Surfaces 32
    2.5.3 Computing Directly from Topography of Arbitrary Surfaces . 34
5.3 Profile Measurement via Atomic Force Microscopy .................................. 86
  5.3.1 Basics of AFM Operation ................................................. 87
  5.3.2 Inter-Scanline Drift .................................................. 89
  5.3.3 Head Tilt ..................................................................... 90
  5.3.4 Comparison with White Light Interferometry ......................... 93
  5.3.5 Results ....................................................................... 95
5.4 Summary ........................................................................... 95

6 Shaped Microfacet Model .......................................................... 98
  6.1 Surface Formation by Shaped Microfacets ............................... 99
  6.2 Expected Modulated Power Spectral Density ............................ 100
  6.3 Nonparametric Modeling of the Expected MPSD ....................... 103
  6.4 Discussions .................................................................... 105
    6.4.1 Probabilistic Interpretation of the Basis Coefficients .......... 105
    6.4.2 Properties of a Shape Basis ............................................ 106
    6.4.3 Azimuthal Basis for Modeling Anisotropy .......................... 109
  6.5 Summary ........................................................................ 109

7 Estimation of Model Parameters .................................................. 111
  7.1 Bases Choice for Shaped Microfacet Model ............................. 112
    7.1.1 Shape Basis: Scaled Sombreros ...................................... 112
    7.1.2 Slant Basis: Shifted Dirac Deltas .................................... 114
    7.1.3 Azimuthal Basis: Piecewise-Linear “Hats” ......................... 114
    7.1.4 Dictionary for the Expected MPSD .................................... 115
  7.2 Least-Square Optimization .................................................... 115
    7.2.1 Discretization .............................................................. 116
    7.2.2 Solving at Multiple $h_z$ ............................................... 116
    7.2.3 Smoothness Regularization .............................................. 117
    7.2.4 Anisotropy Regularization .............................................. 117
    7.2.5 Total Energy Functional ............................................... 118
  7.3 Results and Discussion ......................................................... 118
    7.3.1 Overall Prediction ....................................................... 119
    7.3.2 Effect of Smoothness Regularization Weight ($\kappa_s$) .......... 119
    7.3.3 Effect of Estimated Upper Bound on Facet Radius ($r_{\text{max}}$) . 122
    7.3.4 Effect of Azimuthal Basis Size ($N_\alpha$) .......................... 123
    7.3.5 Effect of Anisotropy Regularization Weight ($\kappa_a$) .......... 124
7.3.6 Effect of Using a Single Shape Basis Function \((N_\Omega = 1)\) .... 124
7.4 Summary ......................................................... 126

8 Modulated Autocorrelation: A Dual Perspective 128
8.1 Modulated Autocorrelation Function ................................. 129
8.2 Interpretation Using Stochastic Processes ......................... 131
8.3 Estimating Expected MACF Using Interpolation ................. 134
  8.3.1 Circumferential Linear Interpolation .......................... 136
  8.3.2 Interpolating Property of MPSD Corresponding to Interpolated MACF ................................................. 138
  8.3.3 Computational Considerations .................................. 139
8.4 Denoising MPSD .................................................. 140
  8.4.1 Truncating MACF .............................................. 141
  8.4.2 Choosing Bandlimits By Assuming Nonincreasing MACF .. 141
8.5 Results and Discussion ........................................... 144
  8.5.1 Prediction using Circumferential Linear Interpolation .... 144
  8.5.2 Effects of Bandlimits on Optimization-Based Prediction .. 145
8.6 Summary ......................................................... 146

9 Conclusion 149

References 152

Appendix A Notations and Glossary 159
  A.1 Operators ...................................................... 159
  A.2 Notation ....................................................... 160
  A.3 Glossary ....................................................... 163
List of Tables

2.1 Literature on reflectance prediction from surface profilometer measurements ......................................................... 39
# List of Figures

1.1 Example of reflectance prediction by our methods .......................... 5

2.1 Geometry for defining the BRDF ........................................... 14

3.1 Setup for deriving the integral theorem of Helmholtz and Kirchhoff ... 46
3.2 Geometric configuration used in Fresnel’s equations ...................... 51
3.3 Setup for evaluating the Kirchhoff integral under the Kirchhoff approximati

4.1 Estimated MPSD projections ................................................. 69

5.1 Common morphological varieties of metallic pigments .................... 78
5.2 Schematic illustration of leafing and nonleafing metallic flakes .......... 78
5.3 Photographs of physical surface samples ................................. 82
5.4 Our gonioreflectometer setup ............................................. 83
5.5 Examples of raw images captured and processed during gonioreflectometer operation .......................................................... 85
5.6 Optically measured reflectance ............................................. 86
5.7 AFM setup ................................................................. 88
5.8 Typical AFM measurements of our surface samples .................... 96
5.9 Height autocorrelation of measure surface profiles ...................... 97

6.1 Slant and azimuthal angles of a shaped microfacet ...................... 99
6.2 Illustration of the formation of a MPSD dictionary element .......... 104
6.3 PSD of a disk and a triangle shape ....................................... 107

7.1 Estimated MPSD projections and predicted reflectance .................. 120
7.2 Effect of smoothness regularization ....................................... 121
7.3 Effect of estimated upper bound on facet radius ........................ 122
7.4 Effect of using a single azimuthal basis function ........................ 123
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.5</td>
<td>Effect of using a single azimuthal basis function and no smoothness regularization</td>
<td>124</td>
</tr>
<tr>
<td>7.6</td>
<td>Effect of anisotropy regularization</td>
<td>124</td>
</tr>
<tr>
<td>7.7</td>
<td>Effect of using a single shape basis element</td>
<td>125</td>
</tr>
<tr>
<td>8.1</td>
<td>Reflectance computed using circumferential linear interpolation</td>
<td>140</td>
</tr>
<tr>
<td>8.2</td>
<td>One-dimensional expected MPSD and MACF</td>
<td>142</td>
</tr>
<tr>
<td>8.3</td>
<td>Effect of truncating the MACF</td>
<td>144</td>
</tr>
<tr>
<td>8.4</td>
<td>Reflectance prediction via circumferential linear interpolation</td>
<td>145</td>
</tr>
<tr>
<td>8.5</td>
<td>Using optimization-based algorithm with denoised MPSD projections</td>
<td>146</td>
</tr>
<tr>
<td>8.6</td>
<td>Interaction between smoothness regularization and denoising MPSD projections</td>
<td>147</td>
</tr>
</tbody>
</table>
Acknowledgments

I would like to thank my advisors Professor Steven Gortler and Professor Todd Zickler for their guidance through out my thesis work. I learned from them to seek and appreciate principled solutions to problems and to be extra meticulousness when dealing with experimental studies. Their insight, expertise, encouragement, and inspiration are indispensable to the completion of this study.

The work discussed in this thesis started out as collaboration work with Anat Levin and Daniel Glasner. I am grateful to them for introducing me to the topic of physical-optics-based reflectance models. They also contributed directly to this study. In particular, the ideas that the reflectance of two-dimensional surfaces can be specified by one-dimensional surface profiles, and that leafing flakes can be used to manufacture surface samples with similar optical property as rough metal surfaces, emerged from discussions with them.

I am grateful to Karen Schneider at Schlenk AG for answering all my questions concerning metallic coating, and for helping me obtain the materials needed to manufacture the surface samples. I benefited greatly from technical discussions with my office mate Ioannis Gkioulekas, who also answered many of my questions concerning building a gonioreflectometer on the optical bench. Finally I am indebted to Wenshou Wang at MIT for teaching me how to use the fun and sometimes dangerous equipments in a chemistry lab.

I would like to thank my parents for their loving and persistent support.
To mom and dad
Chapter 1

Introduction

Reflection of light from surfaces is studied in computer graphics, computer vision, and applied optics. Various models that characterize the directional distribution of light reflected from real-world surfaces have appeared in the literature. In computer graphics, these models determine the photo-realistic appearances of rendered surfaces. In computer vision, these models allow recovery of scene information such as surface normals and lighting conditions from images. In applied optics, how well these models explain and predict experimental measurements along with their applicable regimes and limitations is of interest.

A reflectance model establishes a deterministic mapping from its parameters to some description of surface reflectance, such as the bidirectional reflectance distribution function (BRDF). When a reflectance model is developed based on the laws of physics, its parameters are usually physically-defined characterizations of the surface. Hence in theory it is possible to experimentally measure these model parameters directly from a physical surface, and then evaluate the model to predict the reflectance distribution of the surface.

The formulation of a reflectance model and the experimental procedures for acquiring its parameters depend on many factors: The material under consideration
can be metal, dielectric, layered translucent dielectric such as skin, surfaces rich in microstructures such as fur and fabric, and so on; The type of light transport modeled can include surface reflection, subsurface scattering, florescence, and phosphorescence; The scale of the surface area for which reflectance is measured can range from sub-millimeter, corresponding to taking a close-up photo with a camera, to planet-scale in planetary photometry. These different scenarios all require different reflectance models and experimental procedures tailored to the scenario.

The scenario considered in this thesis is light reflection by rough metal surfaces. We seek to experimentally predict the directional light reflection from rough metal surfaces. Because incident light cannot penetrate much beneath the surface of metal, the appearance of metal depends almost entirely on light reflection at the media interface, that is, the boundary between the metal and air. Interfacial reflection is characterized by certain laws of optics which are direct consequences of Maxwell’s equations. It is thus remarkable that the macroscopic appearance of metals, as studied in the realms of computer graphics and computer vision, is almost completely specified by these physical first principles acting at a microscopic level. There is little need for arbitrary, fitted parameters or constants to account for quantities not directly measurable except those already assumed by these first principles.

According to physics, then, the reflectance of a metal surface is determined by its (complex) refractive index and surface topography. Refractive indices of many materials have been measured and published. Surface topography are routinely measured with sub-wavelength resolution (though not necessarily accuracy) using modern profilometers based on white light interferometry (WLI) or atomic force microscopy (AFM). Thus, it is indeed possible and practical to compute surface reflectance based on these intrinsic surface characterizations. The important question of course is: How accurate is the prediction?

Accurate prediction of this sort has many implications, such as follows:
• The reflectance model, which invariably takes some approximation when deriving from physical first principles, is validated for the type of surfaces that allow accurate prediction.

• Measuring surface topography can be a cheaper and more accurate alternative to optically measuring the BRDF of a surface [Dong et al., 2015], which is a 4D function without even considering wavelength or polarization. Moreover, a hybrid strategy is also possible where optical measurement is made only for a sparse subset of the BRDF domain for calibration and validation, or for regions within the BRDF domain that are hard to simulate exactly, such as when grazing incident and reflection angles are small.

• The measured topography can be digitally altered, for example, to simulate scratching [Yan et al., 2014], roughening, and other mechanical processes. The BRDF corresponding to the altered topography can be recomputed accurately, without the need to alter the real-world surfaces and remeasuring its BRDF.

• Topographical statistics can be inferred from optical measurements in a principled manner [e.g. Bennett and Porteus, 1961; Marx and Vorburger, 1990; Chakrabarti et al., 2014].

In order to be able to discuss previous prediction studies, here we digress to introduce some terminologies. Represent the surface height as a two-dimensional function \( z(x), x \in \mathbb{R}^2 \). One way to model rough surfaces is to view them as samples of a stationary Gaussian process, meaning that \( \{z(x)\} \) is seen as a collection of random variables indexed by \( x \) such that i) each \( z(x) \) follows the same Gaussian distribution, known as the height distribution, and ii) the covariance between the height of any two points is defined by the autocorrelation function \( R \) as \( \mathbb{E}[z(x_1)z(x_2)] = R(x_1 - x_2) \). A Gaussian surface is a stationary Gaussian process, as defined above, whose
autocorrelation function is also Gaussian. For describing the “roughness” of a surface, the terms *vertical* and *lateral roughness scales* are sometimes used to denote the spread of the height distribution and the autocorrelation function, and for all intended purposes can be taken to mean their respective standard deviations. In this thesis, a *rough* surface refers to one with lateral roughness scale on the order of or greater than the wavelength. Most real-world surfaces that do not reflect light like a shiny mirror are rough according to this criteria. All “surfaces” mentioned from now on are assumed to be rough unless otherwise stated. A *one-dimensionally rough* surface is one that is smooth along one azimuthal direction and rough along the other.

Scanning through the literature, we found that good quantitative predictions of surface reflectance had only been demonstrated for i) precision-fabricated Gaussian surfaces [O’Donnell and Méndez, 1987] and, to some extent, ii) one-dimensionally rough surfaces, Gaussian [Bruce, 1995] or non-Gaussian [Marx and Vorburger, 1990]. Here “good quantitative prediction” means one that specifies with reasonable accuracy the *entire* shape of the angular reflectance distribution, including the rate of attenuation away from peaks. Refer to Fig. 1.1 for examples of good quantitative predictions. For real-world two-dimensionally rough surfaces, which are rarely Gaussian, previous predictions exhibited notable discrepancies between the shapes of the simulation and measured reflectance distributions, unless the simulation incorporated some free parameter *back-fitted* to the reflectance measurements, which would then make the process non-predictive.

This thesis presents new measuring procedures and computational algorithms for predicting the reflectance distributions of rough surfaces from profilometer measurements. We experimentally evaluated the proposed process on surface samples created via *metallic coating* using commercially-available *leafing* aluminum flakes and polyurethane binder (see Chapter 5). These test surfaces exhibit non-Gaussian near-isotropic two-dimensional roughness characteristics. The
proposed methods achieved much higher prediction accuracy than demonstrated before. Refer to Fig. 1.1 for the kind of achieved reflectance prediction. In particular, for surfaces with the same material composition (and hence identical refractive indices) but different surface height functions, our procedures accurately predicted the relative magnitude of the peak reflectance values for different surfaces; For each individual surface, our procedures accurately predicted the angular shape of the fall-off in reflectance that occurs away from the peak; There was also no back-fitting of free model parameters to the optical reflectance measurements.

![Figure 1.1: Predicted and measured reflectance within the incident plane and for a fixed reflection angle. The three surface samples, denoted by “10um,” “17um,” and “mixture” in the labels, have the same material composition due to their manufacturing process (see Chapter 5). Labels for the optical ground-truth measurements are prefixed by “gonio.”](image)

We computed reflectance using a physical-optics-based reflectance model known as the Kirchhoff approximation (see Chapter 3 for details), and we acquired the profilometer measurements via AFM. Both the Kirchhoff approximation and AFM have served similar purposes before. What differentiates this study from prior ones is the following observation: There exists random and significant misalignment between the one-dimensional scanlines that comprise the two-dimensional topography map produced by AFM; Moreover the usual ad-hoc methods employed in AFM to correct such misalignment, line-flattening and masking (see Chapter 5 for details), are ineffective for most real-world rough surfaces of interest because i) these surfaces have large lateral roughness scale or lateral feature size compared to the dimension
of the AFM scanning window, and ii) these surfaces do not have planar regions that
can be used for calibration.

Based on this observation, we treat the output of AFM as a collection of one-
dimensional height profiles instead of a single two-dimensional height field. However,
according to the Kirchhoff approximation, surface reflectance is defined by the
Kirchhoff integral over the entire two-dimensional surface. We were able to derive the
relationship between the surface reflectance and the one-dimensional height profiles.
The resulting computational problem is equivalent to inverting the projection operator
\((\text{prj } f)(x) = \int f(x, y)\,dy\).

Guided by this analysis, we first introduced a new experimental procedure
for efficiently collecting reliable profile measurements via AFM. This procedure
measures at multiple lateral positions to account for the large lateral roughness scale
of real-world rough surfaces, while exploiting the fact that only one-dimensional
measurements are needed for speedup.

We next designed two different algorithms for solving the ill-posed problem
of inverting \(\text{prj} \). We emphasize that both algorithms require no back-fitting of
free parameters to reflectance measurements, thereby allowing our procedures for
computing reflectance from profilometer measurements to be truly predictive.

The first algorithm treats a rough surface as an ensemble of shaped and oriented
smooth microfacets, each reflecting light diffractively (due to its shape, not roughness)
according to the Kirchhoff approximation. This surface formation model leads to
a least-square optimization on the joint shape-orientation distribution of the facet.
The second algorithm is based on the Fourier slice theorem, which states that the
inverse Fourier transform of the one-dimensional projection of \(f\), \(\mathcal{F}^{-1}\{\text{prj } f\}\), is a
radial slice of the inverse Fourier transform of the two-dimensional \(f\), \(\mathcal{F}^{-1}\{f\}\). Hence
we can estimate \(\mathcal{F}^{-1}\{f\}\) by interpolating \(\mathcal{F}^{-1}\{\text{prj } f\}\), under the assumption that
the azimuthal variation of \(f\) is smooth.
The above duality of using either primal-domain optimization or dual-domain inter-
polation to invert a projection might remind some readers of computed tomography
(CT) [e.g., Deans, 1983], where an unknown spatial density field is sought given its
projections along multiple directions. Two classes of algorithms are well-known in CT
[e.g., Mersereau and Oppenheim, 1974]: algebraic reconstruction techniques, which
solves an optimization problem in the spatial domain, and Fourier reconstruction
techniques, which use interpolation to estimate the unknown signal in the frequency
domain. However, besides the high level goal of inverting a linear projection and the
resulting duality of being able to work either in the primal or the dual domain, our
approaches and CT techniques share little similarity in details, as explained next.

Both reconstruction methods, optimization and interpolation, work by incorpo-
rating prior assumptions about the signal, since the available measurements represent
only sparse projections of the signal in the primal domain, or equivalently sparse
samplings of the signal in the dual domain. But since the underlying unknown
signals sought here and in CT have different physical interpretations, completely
different prior assumptions must be used. In particular, the primal domain in our
case, where the unknown signal is defined over (see Chapter 4) and the optimization-
based algorithm works in (see Chapter 7), is the frequency domain, and the dual
domain is the spatial domain (see Chapter 8). With CT, the primal domain is the
spatial domain, and the dual is the frequency domain. The appropriateness and
implementational specifics of interpolating our spatial signal, which turns out to be
the autocorrelation function of a complex-valued function defined on the surface, have
little to do with the appropriateness and implementational specifics of interpolating
the (frequency-domain) Fourier transform of a CT spatial density field.

Finally we note that while only interfacial reflection is investigated in this thesis,
more complicated light scattering behaviors of solids have interfacial reflection as
a fundamental building block. Hence this study may serve as a first step towards
accurate reflectance prediction for more complex materials from their intrinsic physical characterizations.

1.1 Contributions

The main contributions of this thesis are as follows:

- We demonstrate near-exact prediction of directional surface reflectance from profilometer measurements for real-world non-precision-fabricated two-dimensionally rough surfaces. The achieved accuracy is significantly higher than shown by prior studies of this kind.

- We identify the reason that two-dimensional topography maps produced by AFM and WLI, as used in some prior studies, will likely cause inaccurate reflectance prediction for real-world moderately rough surfaces. For AFM, the misalignment between the one-dimensional scanlines that comprise such a topography map is essentially uncorrectable due to the lateral roughness scale of the surface. WLI on the other hand has difficulty characterizing wavelength-scale microgeometry and tend to return a topography scan smoother than the actual surface.

- We derive how the reflectance of two-dimensional rough surfaces, as defined by the Kirchhoff approximation theory, is related to the one-dimensional height profiles. Recovering surface reflectance from one-dimensional height profiles is shown to be equivalent to inverting a linear projection.

- We propose a new procedure for efficiently collecting reliable one-dimensional profile measurements via AFM for surfaces that are much rougher than traditionally investigated via AFM.
• We propose a surface formation model that treats a rough surface as an ensemble of shaped and oriented smooth microfacets, each reflecting light according to the Kirchhoff approximation. This model can be seen as a natural extension of the traditional geometrical-optics-based microfacet model [Torrance and Sparrow, 1967].

• We propose an algorithm for computing surface reflectance based on the proposed surface formation model. This algorithm represents the unknown function as a linear combinations of elements from a suitable over-complete dictionary, and uses constrained optimization to solve the linear inverse problem.

• We propose a second algorithm for computing surface reflectance based on the Fourier slice theorem. This algorithm uses interpolation to directly reconstruct in the Fourier domain the unknown two dimensional function from its known projections.

• Our theoretical derivation, measurement procedure, and computational algorithms are all formulated for the general case of anisotropically rough surfaces. To handle anisotropy, our methods simply make use of height profiles measured along multiple azimuthal directions.

• We demonstrate that metallic coating is an appealing method for fabricating test surfaces for future studies of this kind. It is easy to create many samples of varying roughness within a single lab session and the produced surfaces exhibit high spatial uniformity. Importantly, our experimental prediction confirms the following hypothesis: When metallic paint is prepared from leafing flakes, meaning flakes that tend to float to near the air-binder interface due to their manufacturing process, once cured the paint will reflect light essentially in the same way roughened metal does, that is, its reflectance distribution
is completely determined by the surface topography. The truthfulness of this hypothesis is not obvious since factors such as the thin layer of residual polymer between the metal flakes and air can potentially affect the reflectance distribution as well.

- Our prediction method, due to its demonstrated accuracy, suggests a practical profilometer-based alternative to optical BRDF measurement for rough metal surfaces and metallic coating made from leafing flakes.

1.2 Limitations

It is useful to keep in mind some limitations of this study.

- Since our test surfaces are mostly isotropically rough, the anisotropic aspect of our methods has not been experimentally evaluated.

- We only collected in-plane reflectance measurements, with the reflection direction fixed at a moderate angle (24.6°) from the macroscale surface normal. The incident light direction however scans through its entire range.

- Our reflectance computation does not model microscale self-shadowing of the surface. Reflectance attenuation due to self-shadowing will be more prominent when both the incident and the reflection directions are at small grazing angles. This situation in general is not investigated by this thesis. Correction terms for self-shadowing are available in the literature [Smith, 1967].

- The classic Kirchhoff approximation, which we use to compute reflectance, is known to have difficulties dealing with very smooth surfaces or very small incident grazing angles [Thorsos, 1988]. Some newer extension [Krywonos, 2006] supposedly overcame these difficulties.
1.3 Organization

The rest of this thesis is organized as follows:

- Chapter 2 reviews related work on physically-based modeling of surface reflection and their experimental evaluation. Previous predictive studies are discussed in extra details and compared to this study.

- Chapter 3 describes the theoretical foundation of this thesis. A reasonably self-contained derivation from Maxwell’s equations all the way to the specification of surface reflection by one-dimensional surface profiles under the Kirchhoff approximation is presented.

- Chapter 4 presents a formal theoretical overview of the problem addressed by this thesis, the experimental measurements collected, and the computation problem involved in predicting reflectance from the collected profilometer measurements.

- Chapter 5 describes the experimental procedures undertaken, including manufacturing of metallic-paint coated test surfaces, optically measuring in-plane surface reflectance via a simple custom-built gonioreflectometer, and measuring surface profiles via AFM.

- Chapter 6 introduces the shaped microfacet model, which models a reflecting surface as an ensemble of shaped and oriented smooth microfacets that reflect light diffractively according to the Kirchhoff approximation theory.

- Chapter 7 presents and evaluates an optimization-based algorithm for computing reflectance from profile measurements, utilizing the proposed shaped microfacet model as the main prior.
• Chapter 8 presents and evaluates an alternative, conceptually simpler algorithm based on Fourier-domain interpolation.

• Chapter 9 summarizes this thesis, discusses main implications, and suggests potential future work.
Chapter 2

Related Work

In this chapter, we review related work on physically-based surface reflectance models and their experimental evaluation. Surface reflectance of opaque surfaces is commonly characterized by the bidirectional reflectance distribution function (BRDF). We review its definition, acquisition, and modeling in Section 2.1. Many BRDF models have appeared in literature over the years. Some of them are specifically formulated based on the laws of physics, to simulate how light bounces off surfaces in the real world. In Sections 2.2 and 2.3, we examine two groups of such models: one derived from microfacet theory and the other Kirchhoff approximation. BRDF models in these two groups have gained much interest and adoption in the graphics community, because they allow practical implementation, produce realistic surface appearances, and arise from first principles. The microfacet theory uses geometrical optics, which abstracts light propagation as rays. The Kirchhoff approximation employs the more comprehensive physical optics, which acknowledges light as propagating electromagnetic waves and accounts for wave effects such as diffraction and interference. Given their adherence to the laws of physics, we naturally ask if these physically-based BRDF models are capable of predicting surface reflectance of real-world materials. In Section 2.4 and Section 2.5, we survey prior experimental studies addressing this
question, and describe our contribution in the context of these work.

2.1 Bidirectional Reflectance Distribution Function

Nicodemus [1965] introduced the bidirectional reflectance distribution function (BRDF), to specify directional light reflectance of opaque surfaces in a manner independent of the illumination and imaging setup. In this section, we review the definition of BRDF and methods for its acquisition and modeling.

2.1.1 Definition

A beam of light hits an area element $dA$, transmitting radiant flux (radiant energy per unit time) $d\Phi$ onto it. See Fig. 2.1. Consider a unit upper hemisphere $H$ over $dA$, which receives all radiant flux reflected back by $dA$. Let $\hat{l} \in H$ be the direction of the incident beam, and $\hat{v} \in H$ an arbitrary reflection direction making angle $\theta_v$ with the surface normal. Let $d\Omega_v \subset H$ be an area element containing $\hat{v}$, and $d\Phi_v$ the radiant flux received by $d\Omega_v$. Since $H$ is a unit hemisphere, $d\Omega_v$ also identifies the solid angle element extended by the corresponding area element.

![Figure 2.1: Geometry for defining the BRDF.](image)

Under this setup, the ratio $d\Phi_v / d\Phi$ is the fraction of incident radiant flux reflected
onto the area element $d\Omega_v$. It relates to our intuition of directional surface reflectance, but is still proportional to $d\Omega_v$, the solid angle extend by the receiving area. Hence we divide by $d\Omega_v$. The resulting quantity, $d\Phi_v / (d\Phi \, d\Omega_v)$, is the directional density of the reflected fraction $d\Phi_v / d\Phi$. It is also known as the angle-resolved scattering (ARS) [Schröder et al., 2011].

The BRDF, as a function of $\hat{l}$ and $\hat{v}$, is defined as the ARS scaled by the constant function $1 / \cos \theta_v$,

$$f(\hat{l}, \hat{v}) = \frac{d\Phi_v}{d\Phi \, d\Omega_v \, \cos \theta_v}. \quad (2.1)$$

We remark the following:

- The term $1 / \cos \theta_v$ does not change what is being characterized, the directional density of the reflected fraction, but alters our intuition for the value. $f(\hat{l}, \hat{v})$ is now proportional to the perceived brightness of the area element $dA$ viewed from direction $\hat{v}$, since the perceived area of $dA$ is scaled by $\cos \theta_v$.

- $f(\hat{l}, \hat{v})$ is dimensionless, since $d\Phi_v / d\Phi$, $\cos \theta_v$, and $d\Omega_v$ are all dimensionless.

- As defined, $f(\hat{l}, \hat{v})$ is a 4D function, since $\hat{l}$ and $\hat{v}$ can be parameterized by two scalars each. In addition, this measured ratio can depend on wavelengths and polarizations (of both the incident and the reflected radiant flux), locations (of both the incident and the reflected radiant flux) on the surface [Nicodemus et al., 1977; Dana et al., 1999], and time [Gu et al., 2006] resulting in even higher dimensional BRDFs.

Some might notice that the above derivation differs from how BRDF is usually defined in literature, namely as the ratio of the reflected radiance to the incident irradiance. Our motive is to dodge the formidable task of explaining and justifying such incantations as “per unit solid-angle-(\Omega)-in-the-direction-of-a-ray per unit projected-area-(A \cos \theta)-perpendicular-to-the-ray” [Nicodemus, 1965]. Nonetheless,
Eq. 2.1 is equivalent to the usual definition: Radiance along a ray (\(\hat{v}\) here) is defined as the radiant flux, per unit solid angle in the direction of the ray, per unit projected surface area perpendicular to the ray. Hence we divide the reflected radiant flux, \(d\Phi_v\), by both the solid angle along \(\hat{v}\), \(d\Omega_v\), and the projected surface area perpendicular to \(\hat{v}\), \(dA\cos \theta_v\), to obtain the reflected radiance, \(d\Phi_v / (dA \, d\Omega_v \, \cos \theta_v)\). The incident irradiance is defined as the incident radiant flux per unit surface area, so it is just \(d\Phi / dA\). Divide the reflected radiance by the incident irradiance, and we obtain the same quantity as in Eq. 2.1.

2.1.2 BRDF Acquisition

A device for measuring BRDFs is called a gonioreflectometer. According to Webster’s New World Dictionary [2016], the prefix “gonio” comes from the Classical Greek word for an angle or corner. Optics researchers have long been building angular-reflectance-meters, although they typically measure only a subset of the 4D BRDF [e.g., O’Donnell and Méndez, 1987]. In comparison, gonioreflectometers in computer graphics usually acquire the full 4D, or even higher-dimensional measurements.

Murray-Coleman and Smith [1990] proposed a design that moved a pair of light source and photometer, each with two degrees of freedom. This setup acquired a single sample of the 4D BRDF at a time, so it was slow. To accelerate the process, later researchers captured multiple BRDF samples simultaneously with a camera. Cleverly arranged mirrors can route BRDF samples for different reflection directions into a single camera shot: Ward [1992] used a hemiellipsoidal mirror; Han and Perlin [2003] used a kaleidoscope. Alternatively, if the measured surface is of some known curved shape, then each camera shot will contain BRDF samples for many known pairs of incident and reflection directions: Lu et al. [1998] used a known cylinder; Matusik et al. [2003] used a known sphere; Marschner et al. [1999] used a range scanner to make any unknown shape known.
Note that the more complex designs, with their mirrors and curved shapes, are also harder to calibrate. In our study, the reflectance measurements served as the ground truth for physical simulation. To avoid unnecessary calibration error, we opted for a simple gonioreflector design, consisting of a statically mounted camera and a light source rotating within the plane of incidence (see Chapter 5).

2.1.3 BRDF Modeling

Applications of BRDFs, such as rendering and photometric stereo, often demand more compact, efficient, and principled representations than a densely sampled, high dimensional table, leading to both data-driven and physically-based approaches to model the BRDF.

Data-Driven Modeling

BRDFs, as bags of numbers, can be treated with the usual tools of data analysis, including dimensionality reduction, wavelet analysis, rank factorization, and function fitting.

Matusik et al. [2003] measured the BRDFs of 100 real-world materials, and applied principle component analysis and nonlinear manifold charting to compute lower dimensional manifolds containing the measured BRDFs, with optional user guidance. Matusik [2003] in his PhD thesis also explored storing BRDFs compactly as wavelet coefficients. Lawrence et al. [2004] applied nonnegative matrix factorization to approximate a BRDF as the sum of products of pairs of 2D functions. Bilgili et al. [2011] applied Tucker decomposition to approximate a BRDF as the sum of tensor products of quadruples of 1D functions. Brady et al. [2014] applied genetic programming to search for analytical functions that best approximate measured BRDFs. These data-driven approaches typically also exploit BRDF properties such as nonnegativity, azimuthal isotropy, Helmholtz reciprocity [Nicodemus, 1965], and
increased uniformity under a half-angle-based reparameterization [Rusinkiewicz, 1998] to further improve efficiency.

Data-driven BRDF models support any material regardless of how complex its BRDF is. In practice, however, factors including acquisition cost, processing overhead, storage requirement, runtime efficiency, and interpolation quality can limit their use in many applications. They also lack principled methods beyond simple blending for synthesizing novel but physically meaningful appearances from existing BRDF data.

**Physically-Based Modeling**

Besides treating BRDFs as opaque data, researchers also theorized and modeled the physical process of light reflection that gives rise to these data.

Torrance and Sparrow [1967] developed the microfacet theory, which spawned many subsequent BRDF models [e.g., Blinn, 1977; Cook and Torrance, 1982]. This theory treats a rough surface as an ensemble of small, mirror-like facets, each reflecting light specularly according to the laws of geometrical optics. Some empirical BRDF models [Phong, 1975; Lafortune et al., 1997] can also be seen as approximate application of the microfacet theory.

Another physically-based theory that inspired many BRDF models [e.g., He et al., 1991; Nayar et al., 1991; Stam, 1999] is the Kirchhoff approximation, first used by Eckart [1953] to model the scattering of acoustic waves, and then systematically developed by Beckmann and Spizzichino [1963] for electromagnetic waves. Based on physical optics, this more complex theory accommodates wave effects such as diffraction and interference.

Several other theories of light reflection and their derivative BRDF models have appeared in literature. Oren and Nayar [1994] introduced a diffuse analogue of the microfacet theory, treating a rough surface as an ensemble of Lambertian facets.

In general, physically-based BRDF models put more restrictions on what materials they can represent, but run efficiently, produce realistic results for supported materials, and expose physically-meaningful parameters for synthesizing novel appearances. Burley [2012] detailed these and other reasons why the film and the game industries currently favor such models.

Our study seeks to predict surface reflectance from surface profilometry measurements alone. We mainly use the Kirchhoff approximation, but also propose a surface formation model based on shaped microfacets as one approach to solve the resulting ill-posed inverse problem. Hence we next review in more details prior work on both the microfacet theory and the Kirchhoff approximation.

2.2 Microfacet Theory

Torrance and Sparrow [1967] originally developed the microfacet theory in the context of applied optics. In this section, we review this theory and summarize the many subsequent BRDF models developed based on it.

2.2.1 Torrance and Sparrow’s Formulation

The microfacet theory assumes that a rough surface consists of small, randomly oriented, perfectly smooth facets, and that surface reflectance arises from specular reflection of these facets, plus a diffuse term. The theory has three main constituents:

First, by the law of reflection from geometrical optics, only facets with normals
in the direction of the *half vector* \( \mathbf{h} \), defined as \( \mathbf{h} = (\mathbf{\hat{l}} + \mathbf{\hat{v}})/2 \), will reflect radiant flux received along the incident direction, \( \mathbf{\hat{l}} \), towards the reflection direction, \( \mathbf{\hat{v}} \). The *microfacet distribution* [Walter et al., 2007], \( D \), characterizes the density of facets over normal directions, and is defined as

\[
D(\omega) = \frac{S(d\omega, dA)}{d\omega dA},
\]  

(2.2)

where \( \omega \) is a unit vector, \( d\omega \) is the solid angle element containing \( \omega \), \( dA \) is the surface area element of the reflectance of which we are modeling, and \( S(d\omega, dA) \) is the total area of facets within \( dA \) whose normals are within \( d\omega \). Note that Torrance and Sparrow [1967] originally argued in terms of facet count, while assuming facets to have equal area. The microfacet distribution of Walter et al. is an equivalent, but more streamlined characterization, forgoing the need for per-facet area to appear in the derivation. Surface reflectance due to specular reflectance by microfacets is then exactly proportional to \( D(\mathbf{\hat{h}}) \), where \( \mathbf{\hat{h}} = \mathbf{h}/\|\mathbf{h}\| \).

Second, radiant flux reflected by a facet with normal \( \mathbf{\hat{h}} \) is attenuated by the *Fresnel factor*, \( F(\theta_{h,l}) \), where \( \theta_{h,l} \) is the angle between \( \mathbf{\hat{h}} \) and \( \mathbf{\hat{l}} \). \( F \) is dependent on both material and wavelength.

Third, radiant flux arriving at or leaving from a facet at a grazing angle can be partially blocked by neighboring facets. The *geometrical attenuation factor*, \( G(\mathbf{\hat{l}}, \mathbf{\hat{v}}) \) models the fraction of unblocked facet surface.

The BRDF is then defined according to the microfacet theory as

\[
f(\mathbf{\hat{l}}, \mathbf{\hat{v}}) = \frac{D(\mathbf{\hat{h}})F(\theta_{h,l})G(\mathbf{\hat{l}}, \mathbf{\hat{v}})}{4 \cos \theta_l \cos \theta_v} + f_d,
\]  

(2.3)

where \( \theta_l \) and \( \theta_v \) are the angles \( \mathbf{\hat{l}} \) and \( \mathbf{\hat{v}} \) make with the macroscale surface normal, the denominator \( (4 \cos \theta_l \cos \theta_v) \) is physically-derived [Torrance and Sparrow, 1967; Walter et al., 2007; Pharr et al., 2017], and \( f_d \) is a diffuse term.

Out of the three, \( F \) is defined by physics, while \( D \) and \( G \) are statistics of the
surface microgeometry. Torrance and Sparrow modeled $D$ as a Gaussian distribution, and developed an ad-hoc formula for $G$ by treating each facet as one side of a V-shaped groove. Later researchers made other choices.

2.2.2 Introduction of Microfacet Theory to Computer Graphics

A decade later, Blinn [1977] introduced Torrance and Sparrow’s theory to the then-fledgling field of computer graphics. Blinn’s choice for the microfacet distribution was the Trowbridge-Rietz distribution [Trowbridge and Reitz, 1975], which arises from an ellipsoid of revolution.

Later, Cook and Torrance himself [1982] proposed a variant that used the Beckmann distribution [Beckmann and Spizzichino, 1963], which arises from Gaussian surfaces (see Chapter 1), as the microfacet distribution.

We remark that while Blinn’s exposition was pragmatic and operational, Cook and Torrance’s paper exhibits more physical rigor: their BRDF formula was normalized correctly; they explained the difference between dielectric and metal reflectance; they explored chromatic variance of the Fresnel factor; and they showed realistic rendering results based on Fresnel factors of real materials. This might be part of the reason why, over the years, the Cook-Torrance model became the poster child of the microfacet theory, whereas Blinn’s earlier work became associated with the Blinn-Phong model, which is really the Phong model [Phong, 1975] shoehorned into the microfacet theory by Blinn, and not the main result of Blinn’s paper.

2.2.3 Zoo of Microfacet-Based BRDF Models

Since then, many additional choices for the three constituents of the microfacet theory have appeared in literature.

For the microfacet distribution, proposals include anisotropic variants of the Beckmann distribution [Ward, 1992; Kurt et al., 2010], cheaper-to-compute rational
fraction approximation to the Beckmann distribution [Schlick, 1994], an anisotropic variant of the Phong (cosine power) distribution [Ashikhmin and Shirley, 2000], anisotropic generalizations of the Trowbridge-Reitz distribution [Walter et al., 2007; Burley, 2012; Dong et al., 2015], the ABC-model [Church et al., 1990; Löw et al., 2012], and nonparametric tabulation [Ashikhmin et al., 2000].

For the Fresnel factor, the rational fraction approximation of Schlick [1994] is often used because of its simplicity and accuracy.

For the geometrical attentuation factor, Smith [1967] developed an analytical formula for 1D Gaussian functions and isotropic Gaussian surfaces. Smith’s approach was extended to arbitrarily distributed 1D random functions by Brown [1980], and to 2D anisotropic random surfaces by Bourlier et al. [2002]. Walter et al. [2007] applied Smith’s approach to derive analytical formulae corresponding to the Phong distribution and the Trowbridge-Reitz distribution.

2.3 Kirchhoff Approximation

The microfacet theory reviewed earlier is based on geometrical optics, which abstracts propagation paths of light as rays. In reality, light propagates as electromagnetic waves. Physical optics is the study of light as waves. A classical wave phenomenon exhibited by light is diffraction, where a surface reflects an incident beam into multiple directions in a wavelength-dependent manner. Geometrical optics cannot explain this dependency. Physical optics can.

In physical optics, objects reflect light by scattering electromagnetic waves. Because of its practical importance, electromagnetic wave scattering is extensively studied in literature, with many theories and simulation methods developed for it. One such theory, the Kirchhoff approximation, not only led to the development of many BRDF models in computer graphics, but also served as the basis of our study.
In this section, we first briefly summarize this theory, and then review the many BRDF models developed based on it. We leave a more technical exposition of the theory to Chapter 3.

2.3.1 Kirchhoff Approximation and Beckmann-Kirchhoff Theory

According to Ogilvy [1987], Eckart [1953] was the first to apply the Kirchhoff approximation in its correct form. Under the Kirchhoff approximation, wave field at each point on a surface is approximated by the sum of the incident wave field and the wave field reflected by the tangent plane at that point. Wave field away from the surface can then be expressed in terms of the wave field on the surface through the Kirchhoff integral [Goodman, 1969].

Beckmann and Spizzichino [1963] developed a Kirchhoff-approximation-based wave scattering theory for random rough surfaces. Their results, sometimes referred to as the Beckmann-Kirchhoff theory, included a simple formula for wave scattering from Gaussian surfaces, in the form of a series parameterized only by the variances of the surface’s height distribution and autocorrelation. In the same vein, Stogryn [1967] developed a vector analogue of the formula supporting polarization effect and finite conductivity.

The Kirchhoff approximation was introduced to the graphics community early on by Kajiya [1985], but more practical analytical BRDF models based on the Beckmann-Kirchhoff theory and its variant appeared much later.

2.3.2 Physical Optics in Early Computer Graphics

During the 1980s, there were several proposals, including Kajiya’s, for incorporating physical optics into computer graphics. Moravec [1981] numerically simulated wavefront propagation within a scene voxelized with sub-wavelength-scale resolution. Kajiya [1985] numerically evaluated the Kirchhoff integral (from the Kirchhoff approx-
imation) on known surface geometry. Juxtaposing his proposal with much cheaper analytical reflectance models, such as that of Torrance and Sparrow, Kajiya remarked that the Kirchhoff approximation was called for only when “one is interested in a particular surface.” Besides these two numerical approaches, Bahar and Chakrabarti [1987] proposed an analytical reflectance model based on the full-wave theory [Bahar, 1987]. All three proposals demonstrated either poor practicality, or little advantage over cheaper geometrical-optics-based alternatives for common scenes.

### 2.3.3 Zoo of Kirchhoff-Approximation-Based BRDF Models

While Kajiya’s early numerical proposal was computationally expensive, later researchers developed cheaper analytical models based on variants of the Beckmann-Kirchhoff theory.

He et al. [1991] introduced an analytical BRDF model based on Stogryn’s vector analogue of the Beckmann-Kirchhoff theory, while incorporating the geometrical shadowing term of Smith [1967] and the effective roughness concept of Beckmann [1965] to model self-shadowing. In the computer vision community, Nayar et al. [1991] presented a similar model based on the original Beckmann-Kirchhoff theory.

Both these physical-optics-based models assume Gaussian surfaces, as required by the Beckmann-Kirchhoff theory. Stam [1999] proved that, for surfaces with roughness scale much larger than the wavelength of light, the series-based formula from the physical-optics-based Beckmann-Kirchhoff theory, as used by He et al.’s model, reduces to the single-term formula from the geometrical-optics-based microfacet theory for Gaussian surfaces, as used by the Cook-Torrance model\(^2\). Physical optics

---

\(^1\)The advantage of Bahar’s full-wave theory over more classic wave scattering theories such as the Kirchhoff approximation is somewhat controversial. See, e.g., Thorsos and Winebrenner [1991].

\(^2\)It is interesting to note that the Beckmann distribution, which the Cook-Torrance model uses as its microfacet distribution, indeed came from the Beckmann-Kirchhoff theory.
only implies that the width of the bell-shaped reflectance lobe varies slightly with wavelength, and that a specular spike is present for very smooth surfaces. Ngan et al. [2005] then showed experimentally that He et al.’s model and the Cook-Torrance model exhibit comparable levels of fidelity when fitted against real-world BRDFs.

Thus it was left to Stam [1999] to unequivocally demonstrate, with his *diffraction shader*, how the Kirchhoff approximation can lead to BRDFs drastically different from microfacet-based ones. Stam reformulated the Kirchhoff integral as the Fourier transform of a simple function of surface height. Using this new formulation, he derived series-based analytical formulae, akin to that of the Beckmann-Kirchhoff theory, for anisotropic Gaussian surfaces and surfaces with periodic microgeometry. The latter class produces *irridescent* BRDFs, where the reflected light changes color periodically like a rainbow, as exemplified by the surface of a CD. Irradecent reflection is due to diffraction and cannot be described using geometrical optics.

Although Stam was mainly interested in developing analytical models, his Fourier formulation provides an alternative numerical method for computing the Kirchhoff integral. It also sheds lights on how surface microgeometry affects this integral. We and several others [Levin et al., 2013; Dhillon et al., 2014] used his results.

### 2.4 Experimental Validation of Reflectance Models

Physically-based reflectance theories such as both the microfacet theory and the Kirchhoff approximation reviewed earlier are still *approximate* descriptions of physical light reflection. For example, the microfacet theory assumes perfectly smooth microfacets and ignores wave effects, while the Kirchhoff approximation uses the tangent-plane approximation to characterize wave scattering at each point. Thus, how well these theories agree with measurements made from real-world materials is an important question relevant to their application. In this and the next section,
we review the many experimental studies that seek to address this question. They can be loosely grouped into three categories according to their main goals: fitting models to measured reflectance, fabricating reflectance, and predicting reflectance from profilometer measurements. Of the three, the goal of predicting reflectance is shared by this thesis, so we leave a detailed survey of prior studies in this last category to the separate Section 2.5.

To help with exposition, we introduce some notations, used only in this and the next section of this chapter: Denote by \( \mathcal{B} \) the space of BRDFs. Let \( d_\mathcal{B} \) be some distance metric defined over \( \mathcal{B} \). A BRDF model \( m \) can be seen as a mapping from its parameter space, denoted by \( \Theta_m \), to \( \mathcal{B} \). The parameters of a physically-based model have precise physical definitions. For example, the Cook-Torrance model has the root-mean-squared roughness of the surface as one of its parameters, whereas the nonparametric model of Ashikhmin et al. [2000] has the tabulated microfacet distribution of the surface as one of its parameters. Hence, given a physical surface, say \( z \), all parameters associated with the model \( m \) should depend deterministically on \( z \), so we can denote them by \( \theta_m(z) \in \Theta_m \). Finally, denote by \( f(z) \in \mathcal{B} \) the measured BRDF of the surface \( z \).

### 2.4.1 Fitting Models to Measured Reflectance

Let \( G \subset \mathcal{B} \) be a library of BRDFs measured from physical surfaces. We can evaluate how well the model \( m \) approximates the BRDFs in \( G \) via the fitting error, \( E_{\text{fit}} \), defined as

\[
E_{\text{fit}}(m) = \sum_{g \in G} d_\mathcal{B}(m(\Theta_m), g) = \sum_{g \in G} \min_{\theta \in \Theta_m} d_\mathcal{B}(m(\theta), g).
\]

Following this paradigm, Ngan et al. [2005] fitted a number of published BRDF models [Phong, 1975; Blinn, 1977; Cook and Torrance, 1982; He et al., 1991; Ward, 1992; Lafortune et al., 1997] to the MERL database [Matusik et al., 2003]. They
observed that, on average, the Cook-Torrance model and the model of He et al., both of which happened to be the most physically-principled of the bunch, produced the lowest fitting error. In addition, they acquired four anisotropic BRDFs using a custom gonioreflectometer, and showed that the nonparametric microfacet model of Ashikhmin et al. [2000], which allows its microfacet distribution to be an arbitrary two-dimensional function, can fit to these measurements much better than parametric anisotropic models of Poulin and Fournier [1990] and Ward [1992]. This is not surprising, since a nonparametric model typically has many more parameters (in this case, a two-dimensional table of scalars instead of a few scalars). The extra degrees-of-freedom allow the model to fit better at the cost of compactness.

Popularized by Ngan et al.’s work, later studies proposing new BRDF models, such as [Löw et al., 2012; Holzschuch and Pacanowski, 2016], often fitted their models to measured reflectance as a form of evaluation. The fitting error measures a model’s representational power, which, together with the model’s complexity, largely determines the model’s usefulness for applications such as rendering.

It is important to note, however, that exhibiting a small fitting error is a necessary but insufficient condition for the model to be physically accurate. To evaluate the physical accuracy of a model for a given surface $z$, the right thing to do is to measure the model parameters $\theta_m(z)$ from $z$, and then compare $m(\theta_m(z))$ to $f(z)$, as opposed to letting the input parameter to the model $m$ be optimized over the entire $\Theta_m$ as in the case of the fitting error. There is no guarantee that the best fitting parameter, $\arg\min_{\theta} d_B(m(\theta), f(z))$, will equal the physical ground truth, $\theta_m(z)$. The reason that $\theta_m(z)$ is often ignored in graphics literature is that (i) $\theta_m(z)$ can lack a precise physical meaning (ii) $\theta_m(z)$ can be difficult to experimentally measure.
2.4.2 Reflectance Fabrication

Fabricating physical surfaces to exhibit a given BRDF, by nature, constitutes as a form of experimental validation. The general idea is as follows: Given a BRDF model \( m \) and a target BRDF \( g \), run optimization to find a (digital) surface \( z^* = \arg \min_z dg(m(\theta_m(z)), g) \) subject to the constraint that \( z^* \) is manufacturable using the proposed technique; Next, fabricate a physical surface \( z \) according to the digital blueprint \( z^* \); Finally, compare the fabricated BRDF \( f(z) \) to the simulated BRDF \( m(\theta_m(z^*)) \) and also the target BRDF \( g \) to evaluate the proposed workflow.

Weyrich et al. [2009] used a computer-controlled milling machine to manufacture heightfields consists of a 2D array of microfacets, with each facet 1mm by 1mm in size. They optimized the orientation of these microfacets based on geometrical optics to approximate a user-provided 2D specular highlight pattern.

Levin et al. [2013] used photolithography to fabricate heightfields consisting of piecewise flat microfacets, with each facet 2um by 2um in size and at a discrete set of heights. They optimized the heights of these flat facets under the Kirchhoff approximation theory to approximate a given BRDF.

Both these works put emphasis on the viability of the fabrication techniques, the optimization algorithms, the types of BRDFs that can be fabricated, and the achievable spatial and angular resolutions. They did not seek or verify quantitative agreement between the fabricated BRDF, \( f(z) \), and the simulated BRDF, \( m(\theta_m(z^*)) \). Moreover, from these works alone we cannot infer how well the corresponding models will work on real-world surfaces with topography vastly different from that of fabricated surfaces.
2.5 Reflectance Prediction from Profilometer Measurements

Given any scientific theory, the ultimate question one can and should ask is: Can it be used for prediction? In the context of physically-based BRDF models, the above question becomes: Can we measure the model parameters $\theta_m(z)$ from a physical surface $z$, and then evaluate the model, $m$, at $\theta_m(z)$ to predict the BRDF of the surface, $f(z)$?

In this section, we survey previous works on reflectance prediction from surface profilometer measurements. Browsing through existing literature, we found that accurate quantitative prediction has been demonstrated only for i) strictly Gaussian surfaces [O’Donnell and Méndez, 1987], and ii) one-dimensionally rough surfaces, Gaussian or non-Gaussian [Bruce, 1995; Marx and Vorburger, 1990]. For real-world two-dimensional rough surfaces (which are rarely Gaussian), existing works [Lettieri et al., 1991; McKnight et al., 2001; Li and Torrance, 2005; Schröder et al., 2011; Dhillon et al., 2014; Dong et al., 2015] exhibit at least one of the following shortcomings:

**Qualitative prediction.** They only demonstrated qualitative agreement between the predicted and measured BRDFs, essentially confirming that rougher surfaces are duller and smoother surfaces are shinier. Fine details of the angle-resolved reflectance, such as the relative magnitudes of the peaks and the shape of reflectance distribution, are not accurately predicted.

**Back-fitted “shape” parameter.** They incorporated in their models a free parameter which was back-fitted from the optical BRDF measurements and not measured from the surface geometry itself. Typically this parameter controls

---

3Here we use the expository notations from Section 2.4
some bandlimiting operation applied to the surface geometry prior to evaluating the actual model. Thus they essentially smoothed the geometry as needed to make the shapes of the computed and measured reflectance distributions agree better, making the procedure non-predictive.

Before we dive into the detailed reviews, there are two practical issues regarding data presentation that warrant some discussions:

**Logarithmic scale plotting.** Some studies plotted the reflectance in logarithmic scale. This makes it somewhat hard to evaluate the agreement between the measured and computed reflectance distributions. Large differences in how fast two signals attenuate away from the peaks will be visually understated by such a plot. Therefore, we will explicitly note the use of logarithmic scaled plotting when mentioning such a study. Interested readers are encouraged to refer to the cited work to decide for themselves.

**Per-sample scaling.** In theory, both the optically measured and the computed BRDFs are unitless, and hence should be compared without any scaling. In practice, the raw data collected by a gonioreflectometer equals the true BRDF scaled by a constant, $\kappa_{\text{opt}}$, determined by the optical setup; Similarly, a BRDF model typically includes a Fresnel term, $F$, which is determined by the optical property of the material and the wavelengths of the incident and reflected light. Measuring both $\kappa_{\text{opt}}$ and $F$ exactly through experimental calibration can be labor-intensive and error-prone, unless high-precision scientific instruments are available. Hence, in practice, $\kappa_{\text{opt}}$ is sometimes determined by fitting the computed BRDF, denoted by $g_{\text{comp}}$, to the measured BRDF, denoted by $g_{\text{meas}}$, i.e., $\kappa_{\text{meas}} = \arg \min_{\kappa} d_g(g_{\text{meas}}, \kappa g_{\text{comp}})$. Likewise, the Fresnel term is often computed from published refractive indices of common materials, even
though the test samples can have slightly different material composition due to oxidization, impurity, manufacturing irregularity, and so on. Sometimes, the Fresnel term is foregone completely, since it is roughly constant within the region of the BRDF domain that is plotted for comparison, and the computed BRDF will be scaled by a constant ($\kappa_{\text{opt}}$) fitted from the optical measurements anyway.

A careful reader might point out that the “scaling factor” $\kappa_{\text{opt}}$ is essentially another free parameter, back-fitted from optical measurements. Note, however, that as long as the optical setup and the material composition do not change, $\kappa_{\text{opt}}$ and $F$ are fixed. Therefore, if multiple test surfaces with the same material composition but different roughness are available, then $\kappa_{\text{opt}}$ can be fixed from a single surface, and it should no longer be considered as a back-fitted free parameter. A few studies, including ours, adopt this approach. On the other hand, if a study uses a different $\kappa_{\text{opt}}$ for each different sample, then it indeed deserves more scrutiny. We will refer to such studies as using per-sample scaling and discuss the implication.

### 2.5.1 Strictly Gaussian Surfaces

O’Donnell and Méndez [1987] were the first to demonstrate accurate quantitative prediction of angle-resolved reflectance for Gaussian surfaces using the Beckmann-Kirchhoff theory. Their key innovation was to meticulously fabricate test surfaces that were strictly Gaussian, employing a carefully-designed lithography process. They found excellent agreement between measurements and predictions, albeit restricted to the regime that the Kirchhoff approximation was known to make sense: Namely, the lateral roughness scale of the surface should not be significantly smaller than the wavelength, and the incident and reflection directions should be away from grazing
angles. Their plots used *per-sample scaling*, but the shapes of the measured and the predicted reflectance distributions matched up so well that little doubt was left concerning the exactness of the prediction.

Knotts and O’Donnell [1994] continued exploring lithography based fabrication techniques, this time creating testing surfaces with one-dimensional Gaussian roughness. Bruce [1995] used Knotts and O'Donnell’s data to validate the Beckmann-Kirchhoff theory in one-dimension. His reflectance plots, which used a *single* scaling factor for different test samples, showed reasonable agreement between measurements and predictions. The predicted peak magnitudes were slightly off for some samples, which might be due to deviation from strict Gaussianity and failing of the Kirchhoff approximation for over-smooth surfaces.

### 2.5.2 Modeling Arbitrary Surfaces as Parametric Random Surfaces

Despite the success with Gaussian surfaces, most real-world surfaces are not Gaussian. To predict the reflectance of a non-Gaussian surfaces, some studies first fit a parametric statistical model of a random rough surface to the profilometer measurements, and then compute the reflectance distribution off this parametric model using either the Beckmann-Kirchhoff theory or some other suitable theory.

Lettieri et al. [1991] computed the reflectance distribution of glossy coating on paper from *one-dimensional* profilometry data using the Beckmann-Kirchhoff theory. Their parametric random surface model has a Gaussian height distribution with variance $\sigma^2$, and an autocorrelation of the form $R(d) = \exp(-|d/T|^a)$, where $\sigma$, $T$, and $a$ are fitted parameters. They plotted using *logarithmic scale*, and the predicted shapes, even in logarithmic scale, were much more dilated than the measured ones. This prompted them to scale $\sigma$ by some constant, or alternatively high-pass filter the profilometry data, in order to contract the computed distribution. Thus either the roughness scaling factor or the bandlimit used for spatial filtering played the role...
of the aforementioned \textit{shape parameter}, and its value was selected to “gave the best visual agreement”. Now they did use a single value for all three test surfaces, but conceded that there was no a priori reason to do so. Moreover, even with the shape parameter applied, the agreement was inexact, especially with respect to how fast reflectance attenuates away from peaks.\footnote{We noticed the inexactness of the match only after overlaying Fig. 9(a) and (b) from [Lettieri et al., 1991]. Visually the two do seem to agree well, but the use of logarithmic scale makes this agreement less meaningful.}

Li and Torrance [2005] computed the reflectance of roughened and aluminum-coated steel and glass substrates using the Beckmann-Kirchhoff theory based BRDF model of He et al. [1991]. Hence their parametric surface model is just the Gaussian surface, with parameters fitted from \textit{one-dimensional} profilometry measurements. Like Lettieri et al., they high-pass filtered the profilometry measurements, in order to achieve reasonable agreement between the computed and the measured reflectance. The bandlimit used for filtering, i.e., the \textit{shape parameter}, was chosen empirically per-sample, so this study is more about model-fitting than prediction.

Schröder et al. [2011] computed the reflectance of steel surfaces with different roughness using the generalized Harvey-Shack theory [Harvey, 1990; Krywonos, 2006]. For their parametric surface model, the height distribution is Gaussian, and the lateral power spectrum, which is the Fourier transform of the lateral autocorrelation and hence an equivalent characterization, follows the ABC-model [Church et al., 1990]. They assembled a “master” power spectrum by combining lateral power spectra computed from \textit{two-dimensional} topography scans with different lateral resolutions, obtained via both white light interferometry and atomic force microscopy. They then reduced this painstakingly-built “master” power spectrum to three numbers for actual use, by fitting the ABC-model to it. They plotted results using \textit{logarithmic scale} in both the reflectance and angle axes. Considerable discrepancies exist between
their computed and measured reflectance distributions.

In summary, parametric modeling has the practical advantage of requiring less extensive profilometer measurements to fit the few parameters needed to characterize the parametric random surface model. Moreover, it is often possible to fit parameters of a two-dimensional random surface from one-dimensional profile data alone, assuming that the surface is isotropically rough. As we will discuss in Chapter 5, it is a formidable if not impossible quest to obtain two-dimensional topography scans of good enough accuracy from real-world surfaces with moderate roughness. Finally it can also be argued that since many measurements are condensed into few parameters for final use, this approach is more robust and less affected by measurement noises and errors.

However, this approach also comes with one major downside: The physical surface and its parametric approximation intrinsically differ, so one would not expect the two to produce identical reflectance, and indeed experimentally they do not. As a result, the shapes of the computed reflectance distributions often differ greatly from those of the measured ones. If researchers desire a more exact match, they invariably have to incorporate into the model the aforementioned shape parameter, to bring the computed shape closer to the measured one. At that point, however, the whole procedure falls into the realm of model-fitting (Section 2.4.1).

2.5.3 Computing Directly from Topography of Arbitrary Surfaces

Previously, we discussed the difficulty associated with accurately predicting reflectance from parametric random surface models fitted to real-world surfaces. Alternatively, it is possible to compute reflectance directly from the surface profilometer measurements represented either as two-dimensional topography maps, i.e., height-fields, or one-dimensional height profiles.
Marx and Vorburger [1990] computed the reflectance distribution of stainless steel surfaces rough only in one-dimension, by numerically evaluating the Kirchhoff integral on one-dimensional profilometry data. The agreement between the computed and the measured distributions looked reasonable, although their results were presented using both per-sample scaling and logarithmic scale. Note that they also studied the inverse problem of inferring surface statistics from optical measurements, by modeling the unknown surface as a parametric random surface and fitting its parameters to the optical measurements. They found that surface roughness (of the parametric random surface) inferred from optical measurements was consistently smaller than one derived from profilometer measurements, which is exactly why studies like [Lettieri et al., 1991; Li and Torrance, 2005] needed to incorporate a shape parameter as we discussed in Section 2.5.2.

McKnight et al. [2001] computed the reflectance of clear dielectric epoxy coating from two-dimensional topography maps measured via interferometric microscopy. They implemented two numerical methods, one based on the Kirchhoff integral, and the other geometrical optics. Even with both per-sample scaling and logarithmic scale, their plots showed considerable discrepancies between the measured and the computed distributions, especially with respect to how fast reflectance attenuates away from peaks, i.e., the shape of the distribution. The use of per-sample scaling also means that we cannot tell if relative peak magnitudes were correctly predicted.

Dhillon et al. [2014] computed the iridescent reflectance of biological grated surfaces, such as snake skin, by evaluating the Kirchhoff integral over two-dimensional topography map measured via atomic force microscopy (AFM). As experimental validation, they took photos of a (gold-coated) sample surface and also rendered it using the computed BRDF. Impressively, the iridescent diffraction patterns due to surface grating appeared at roughly the same locations on both the rendered images and the photos, which indicates that AFM reasonably characterized the lateral
positions of the surface microstructures. Exact agreement between the measured and the computed reflectance distributions was not a goal of the study, and indeed there were notable differences in intensity levels and gradient magnitudes between the rendered images and the photos.

Dong et al. [2015] computed the reflectance of roughened metals from two-dimensional WLI topography measurements, using both geometrical optics and the Kirchhoff approximation. Their study is unique in that they considered both isotropic and anisotropic samples, and made optical measurements not just for a single azimuthal direction. Their anisotropic samples, which were brushed metal, however were so smooth along the brush direction that measured reflectance fell to near zero when the half angle was more than 1° from the normal direction. Their gonioreflectometer had a lower limit of 1.5° when resolving surface normal variations. Taking both into account, their anisotropic samples were effectively one-dimensionally rough. Within a single azimuthal direction, they restricted the incident and reflection directions to be within the plane-of-incidence, forming a fixed 14.14° between them, and let the half angle step through its valid range. This represents a similar amount of optical measurement, i.e., a one-dimensional scan, compared to other studies, where one of the incident or reflection direction stays fixed and the other steps through its range. Due to the high spatial variance of their surface samples, both the measurements and the predictions shown were noisy in the peak region, making it hard to tell if error was due to systematic mis-prediction or just spatial variance. Outside of the peak, the measured and predicted curves were similar in shape but have different attenuation rates. Their geometrical optics based computation produced a somewhat better fit, but it employed weighted least square plane-fitting to estimate facet normals, with the weighting controlled by an empirically-chosen kernel size which serves as the shape parameter.

In summary, the above studies experienced some degrees of success, but an exact
agreement between measurement and prediction such as shown in [O’Donnell and Méndez, 1987] remained elusive, except for one-dimensionally rough surfaces [Marx and Vorburger, 1990]. Some common caveats include: The predicted and measured reflectance curves exhibited correlated but nonetheless different attenuation rates; The prediction was sensitive to some empirical choices, such as how local tangent planes were fitted; Reflectance for different surfaces was plotted using different scales, begging the question whether the procedure, given two surfaces of different roughness but identical material composition, could correctly forecast the ratio between their respective peak reflectance (even just within the incident plane for a fixed incident angle).

2.5.4 Recap and Comparison to This Study

Here is a brief recap of predictive studies reviewed in this section:

- For precision-fabricated strictly Gaussian surfaces, previous studies were largely successful at making accurate predictions.

- Studies that modeled real-world surfaces as parametric random surfaces often exhibit systematic discrepancies between the computed and measured reflectance due to the inherent difference between the physical surfaces and the statistical random surface models, prompting some researchers to massage the profilometer measurements empirically to obtain a better fit.

- Studies that computed reflectance directly from profilometer measurements of real-world surfaces, except in the case of one-dimensionally rough surfaces, could not achieve prediction with the level of accuracy as demonstrated for precision-fabricated Gaussian surfaces, with no clear reason why.

In this thesis, we detail how we were able to predict the reflectance distributions of real-world two-dimensionally rough surfaces with much higher accuracy than
demonstrated before, correctly characterizing both the shapes and the relative peak magnitudes. We mainly used the Kirchhoff approximation theory, and acquired the profilometer measurements via atomic force microscopy.

Our method computed directly from surface profilometer measurements, similar to the studies reviewed in Section 2.5.3 but with one key difference: We used only one-dimensional profile measurements in our computation. We will explain in Chapter 5 our rationals for this choice. Consequently, we suspect that the two-dimensional topography measurements used in prior studies were of quality unsuitable for high Precision reflectance prediction. Our test surfaces, of course, are still two-dimensionally rough, so we had to formulate and solve a different computational problem than these prior studies. Table 2.1 summarizes the studies reviewed in this section along with our work.

2.6 Summary

In this chapter, we reviewed existing literature related to our work. We first introduced the basic radiometric concept of BRDF, and reviewed its acquisition and modeling. We then turned our focus to the two physically-based reflectance models that have achieved great success in computer graphics, computer vision, and applied optics: the microfacet model, and the Kirchhoff approximation model. We described their historical first appearances, their adoption in computer graphics, and their subsequent derivative BRDF models. Finally we reviewed work on experimental validation of these physically-based models. In particular, we went to great length to describe past studies on reflectance prediction from surface profilometer measurements because of their relevance to our work. In the next chapter, we will develop the theoretical foundation of our work, and frame in technical details the problem addressed by this thesis.
<table>
<thead>
<tr>
<th>Surface type</th>
<th>Compute</th>
<th>Measurements</th>
<th>Theory</th>
<th>Notes</th>
<th>Acc. Pr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>O'Donnell and Méndez [1987]</td>
<td>2D iso. Gaussian</td>
<td>param.</td>
<td>1D stylus</td>
<td>KA</td>
<td>scale</td>
</tr>
<tr>
<td>Bruce [1995]</td>
<td>1D Gaussian</td>
<td>param.</td>
<td>1D stylus</td>
<td>KA</td>
<td></td>
</tr>
<tr>
<td>Lettieri et al. [1991]</td>
<td>2D iso. rough</td>
<td>param.</td>
<td>1D stylus</td>
<td>KA</td>
<td>shape, log</td>
</tr>
<tr>
<td>Li and Torrance [2005]</td>
<td>2D iso. rough</td>
<td>param.</td>
<td>1D stylus</td>
<td>KA</td>
<td></td>
</tr>
<tr>
<td>Schröder et al. [2011]</td>
<td>2D iso. rough</td>
<td>param.</td>
<td>2D AFM + WLI</td>
<td>GHS</td>
<td>log</td>
</tr>
<tr>
<td>Marx and Vorburger [1990]</td>
<td>1D rough</td>
<td>direct</td>
<td>1D stylus</td>
<td>KA</td>
<td>scale, log</td>
</tr>
<tr>
<td>McKnight et al. [2001]</td>
<td>2D iso. rough</td>
<td>direct</td>
<td>2D WLI</td>
<td>KA, ray</td>
<td>scale, log</td>
</tr>
<tr>
<td>Dhillon et al. [2014]</td>
<td>2D bio. grated</td>
<td>direct</td>
<td>2D AFM</td>
<td>KA</td>
<td>rendering</td>
</tr>
<tr>
<td>Dong et al. [2015]</td>
<td>2D iso. + 1D rough</td>
<td>direct</td>
<td>2D WLI</td>
<td>KA, ray</td>
<td>MD plot</td>
</tr>
<tr>
<td>This study</td>
<td>2D iso. rough</td>
<td>direct</td>
<td>1D AFM</td>
<td>KA</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.1: Summary of literature on reflectance prediction from surface profilometer measurements. The “Surface type” column describes the type of the physical surface samples, where “iso. rough” means isotropically rough and “bio. grated” means biologically grated; The “Compute” column describes the computational method used, where “param.” represents modeling as parametric random surfaces, and “direct” represents direct calculation from profilometer measurements; The “Measurements” column describes the type of profilometer measurements used in computation, where “1D” or “2D” indicates whether one-dimensional or two-dimensional measurements were used, “stylus” represents mechanical stylus profilometry, “AFM” stands for atomic force microscopy, and “WLI” stands for white light interferometry; The “Theory” column describes the theory used to compute the reflectance distributions, where “KA” stands for Kirchhoff approximation, “ray” stands for geometrical (ray) optics, and “GHS” stands for generalized Harvey-Shack; The “Notes” column lists specifics about the study worth noting, where “scale” indicates the use of per-sample scaling, “shape” indicates the use of a shape parameter, “log” indicates logarithmic scaled plotting, “rendering” indicates that rendered images were shown instead of the actual reflectance distributions, and “MD plot” indicates that microfacet distributions were shown instead of the actual reflectance distribution; Finally, the “Acc. Pr.” column shows whether accurate quantitative prediction was demonstrated according to our evaluation.
Chapter 3

Background

This chapter details the theoretical foundation of this study. In Section 3.1, we review the scalar wave theory of light from classical physics. Based on this theory, in Section 3.2 we next show how to evaluate the so-called Kirchhoff integral to compute the amount of light reflected off a rough surface from its microscopic topography.

We note that most of the materials covered in this chapter can be found scattered across multiple chapters from the classical tomes and several papers [Beckmann and Spizzichino, 1963; Goodman, 1969; Tomiyasu, 1988; Born and Wolf, 1999; Stam, 1999]. However, we find it hard to come by shorter writings that develop from such well-known first principles as the Maxwell equations, all the way to the rather technical physical-optics-based calculation behind the Kirchhoff approximation theory. For example, both Nayar et al. [1991] and Stam [1999] took as given the final form of the Kirchhoff integral with boundary conditions defined by the Kirchhoff approximation (Eq. 3.21). Our goal is thus to provide a succinct and yet self-contained treatise on the derivation in these two sections, built from first principles.

Another important reason for deriving these results from scratch is that it allows us to take note of the approximations taken at various steps of the Kirchhoff approximation theory. When the computed reflectance differs from the optical
measurement, as it invariably does at various point of this study, the discrepancy can be due to coding bugs, experimental measurement error (associated with both the profilometer measurements and the BRDF measurements), or inaccuracy inherent in the Kirchhoff approximation theory. By understanding the exact approximations employed by the theory, we will have better chance at locating the source of the discrepancy and thus fixing it, instead of prematurely attributing all errors to inherent inaccuracy of the theory.

3.1 Physical Optics Theory

In this section, we guide the reader through some concepts and results from classic physical (wave) optics. These results will be used in the next section to derive a formula for light scattering by rough surfaces under the Kirchoff approximation.

We start from the classic Maxwell equations and simplify them into a single scalar equation, known as the wave equation. By restricting our attention to monochromatic waves, we further simplify wave equation into the Helmholtz equation. Finally, we apply Green’s second identity to derive the integral theorem of Helmholtz and Kirchhoff, which expresses the solution to the wave equation (or equivalently the Helmholtz equation) at any point in terms of values of the solution and its normal derivative on an enclosing surface.
3.1.1 Maxwell Equations

Light is an electromagnetic phenomenon. Its propagation in charge-free and current-free space is described by the Maxwell equations,

\begin{align}
\nabla \times \mathbf{E} &= -\mu \frac{\partial \mathbf{H}}{\partial t} \\
\nabla \times \mathbf{H} &= \varepsilon \frac{\partial \mathbf{E}}{\partial t} \\
\nabla \cdot \varepsilon \mathbf{E} &= 0 \\
\nabla \cdot \mu \mathbf{H} &= 0,
\end{align}

where \( \mathbf{E} \) is the electric field, \( \mathbf{H} \) is the magnetic field, \( \mu \) is the permeability of the space, and \( \varepsilon \) is the permittivity of the space. Both \( \mathbf{E} \) and \( \mathbf{H} \) are time-dependent vector fields.

Note that the full Maxwell equations contain extra terms involving the electric charge density and the electric current density, but in the setting of light propagation in free space, both are 0, and hence we do not show the related terms in the above equations. When away from boundaries and in a homogeneous medium such as vacuum and air, we can also assume \( \varepsilon \) and \( \mu \) to be constant through the region of propagation.

3.1.2 Scalar Wave Equations

Using the assumption that the region of propagation is charge-free, current-free, and of constant permeability \( \mu \) and permittivity \( \varepsilon \), we can remove the coupling between the electric field and the magnetic field in the Maxwell equations (Eq. 3.1) and describe both fields with a single scalar equation.

Applying the \( \nabla \times \) operation to both sides of the first equation from Eq. 3.1, and using the vector identity \( \nabla \times (\nabla \times \mathbf{F}) = \nabla(\nabla \cdot \mathbf{F}) - \nabla^2 \mathbf{F} \) for any vector field \( \mathbf{F} \), we
obtain
\[ \nabla (\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} = -\nabla \times \mu \frac{\partial \mathbf{H}}{\partial t}. \]

Now substituting in the second and third Maxwell equation from Eq. 3.1 and making use of the assumption that \( \epsilon \) and \( \mu \) are constant, we obtain
\[ \frac{\partial^2 \mathbf{E}}{\partial t^2} = v^2 \nabla^2 \mathbf{E}, \]
where \( v = \sqrt{\mu \epsilon}^{-1} \). This second order partial differential equation describes wave propagation and is known as the wave equation. In particular, the coefficient \( v \) is the speed of wave propagation. We can repeat this process for the magnetic field to arrive at an identical wave equation for \( \mathbf{H} \):
\[ \frac{\partial^2 \mathbf{H}}{\partial t^2} = v^2 \nabla^2 \mathbf{H}. \]

Since projections of the vector fields \( \mathbf{E} \) and \( \mathbf{H} \) along any direction satisfy the same wave equation by the linearity of \( \partial^2 / \partial t^2 \) and \( \nabla^2 \), we can represent the behavior of both fields with a scalar wave equation:
\[ \frac{\partial^2 u}{\partial t^2} = v^2 \nabla^2 u, \quad (3.2) \]
where the time-dependent scalar field \( u \) is taken to represent any Cartesian component of \( \mathbf{E} \) and \( \mathbf{H} \). From now on, we simply refer to \( u \) generically as the wave field or wave function.

3.1.3 Monochromatic Waves and the Helmholtz Equation

Now that we have a single scalar wave equation for the description of light, down from the four vector-valued Maxwell equations, we will further simplify our description by restricting our attention to a particularly simple form of waves, time-harmonic waves of a single frequency, or monochromatic waves. As we will see, this allows us
to remove the dependency on time from the scalar wave equation.

Monochromatic waves have a wave field \( u(r, t) \) periodic in time with the form:

\[
    u(r, t) = A(r) \cos(2\pi ft + \phi(r))
\]

(3.3)

where \( A(r) \) and \( \phi(r) \) are the amplitude and phase of the wave at position \( r \), and \( f \) is the fixed frequency of the wave, representing the number of vibrations per second of the wave field for any point.

Since calculation with trigonometric functions can be cumbersome, it is customary in the literature to use a notation based on complex exponential by rewriting Eq. 3.3 as

\[
    u(r, t) = \text{Re}(U(r)e^{-i2\pi ft})
\]

(3.4)

where \( \text{Re}(\cdot) \) denotes the real part, and the complex valued function \( U(r) = A(r)e^{-i\phi(r)} \) encodes both the amplitude, \( A(r) \), and the phase, \( \phi(r) \), of the wave. \( U \) is known as the complex amplitude of the wave.

A monochromatic wave field must still satisfy the wave equation (Eq. 3.2). Hence upon substituting the complex amplitude based wave definition (Eq. 3.4) into the scalar wave equation (Eq. 3.2), we obtain

\[
    \frac{\partial^2 u}{\partial t^2} = \text{Re}(-e^{-i2\pi ft}(2\pi f)^2U)
\]

\[
= v^2 \nabla^2 u = \text{Re}(e^{-i2\pi ft}v^2\nabla^2 U).
\]

Since the above holds for any \( t \), it holds if and only if the complex amplitude \( U \) satisfies the so-called Helmholtz equation:

\[
    (\nabla^2 + k^2)U = 0, \quad k = \frac{2\pi f}{v},
\]

(3.5)

where \( k \) is called the wave number. Note that the time parameter \( t \) no longer appears in the equation. This is expected since we have fixed the time dependence a priori.
The Helmholtz equation (Eq. 3.5) thus gives a time-independent description of a monochromatic wave field.

3.1.4 The Integral Theorem of Helmholtz and Kirchhoff

Our eventual goal for this section is to express the value of the complex wave amplitude $U$ at any observation point in terms of the value of $U$ and its normal derivative on the boundary of an enclosing volume. We will need a mathematical relation known as Green’s second identity, stated without proof here:

**Theorem** (Green’s second identity). Let $V$ be a volume bounded by a closed surface $S$. Let $U$ and $G$ be complex-valued functions with continuous first and second-order partial derivatives within and on $S$, then

$$
\iint_V \left( U \nabla^2 G - G \nabla^2 U \right) \text{d}v = - \iint_S \left( U \frac{\partial G}{\partial n} - G \frac{\partial U}{\partial n} \right) \text{d}S,
$$

(3.6)

where $\partial/\partial n$ denotes differentiation along the inward normal\(^1\) of $S$.

Refer to Fig. 3.1 for the following setup. Now let $v$ be the viewpoint. Let the complex wave amplitude $U$ satisfy the Helmholtz equation (Eq. 3.5). Let $S$ be a closed surface containing $v$ and on which $U$ and $\partial U/\partial n$ are known. Define an auxiliary function $G$, often referred to as the Green’s function, for use with Green’s second identity as

$$
G(r) = \frac{e^{ik\|r-v\|}}{\|r-v\|},
$$

where $\|\cdot\|$ specifies the length of a vector and $k$ is the wave number from Helmholtz’s equation (Eq. 3.5). Finally surround $v$ by a sphere $S_\epsilon$ of radius $\epsilon$ that lies completely inside $S$, so $G$ is defined for any point outside $S_\epsilon$.

We will apply Green’s second identity to $U$ and $G$ on the volume bounded by $S$

\(^1\)Green’s second identity is often stated in terms of an outward normal, but the inward normal is more convenient for its application here.
and \( S_\epsilon \). The definition of \( G \) implies that, similar to \( U \), \( G \) also satisfies the Helmholtz equation (Eq. 3.5), so when substituted into Green’s second identity (Eq. 3.6), the integrand on the left side of the equation vanishes for every point in the volume bounded by \( S \) and \( S_\epsilon \), and we have

\[
\left\{ \int_S + \int_{S_\epsilon} \right\} \left( U \frac{\partial G}{\partial n} - G \frac{\partial U}{\partial n} \right) dS = 0
\]

Move the terms corresponding to integration over \( S_\epsilon \) to the other side and let \( \epsilon \to 0 \), we will eventually, via calculation, arrive at

\[
U(v) = \frac{1}{4\pi} \int_S \left[ U \frac{\partial}{\partial n} \left( e^{ik\|r-v\|} \frac{r-v}{\|r-v\|} \right) - e^{ik\|r-v\|} \frac{\partial U}{\partial n} \right] dS(r), \quad (3.7)
\]

where \( \partial/\partial n \) denotes differentiation along the inward normal, which points from points on \( S \) towards \( v \). This result is known as the integral theorem of Helmholtz and Kirchhoff. As promised, it expresses the complex wave amplitude \( U \) in terms of values of both \( U \) and \( \partial U/\partial n \) on the boundary surface \( S \). The integral in Eq. 3.7 is often referred to simply as the Kirchhoff integral.

### 3.1.5 Recap

Now is a good time to pause and recap the sequence of derivations in this section.

We first simplified the Maxwell equations into the scalar wave equation, a second
order partial differential equation satisfied by any Cartesian component of both the 
electric and magnetic vector fields. Without loss of generality, we pick one such 
component as our wave field, \( u(\mathbf{r}, t) \).

The wave field \( u \) is a real valued spatially and temporally varying function, but 
by restricting our consideration to \textit{monochromatic waves}, we can represent \( u(\mathbf{r}, t) \) 
by the \textit{complex amplitude}, \( U(\mathbf{r}) \), a complex valued function of space only. Under 
this representation, the wave equation reduces to the time-independent \textit{Helmholtz} 
equation constraining \( U \).

Finally with a clever application of Green’s second identity, we arrived at the 
\textit{integral theorem of Helmholtz and Kirchhoff}, which shows that if \( U \) satisfies the 
Helmholtz equation, then \( U(\mathbf{v}) \) for a given observation point \( \mathbf{v} \) can be expressed 
using values of \( U \) and \( \partial U / \partial \mathbf{n} \) on the boundary of some volume enclosing \( \mathbf{v} \).

### 3.2 Reflection from a Rough Surface

In this section, our goal is to use the Kirchhoff integral developed in the previous 
section to compute the light reflectance distribution of rough surfaces. Doing so 
requires knowing the wave field on the surface boundary. However, calculating the 
precise wave field on the boundary of an arbitrary rough surface is hard, so the 
Kirchhoff approximation, also known as the tangent plane approximation, was often 
used [Eckart, 1953; Beckmann and Spizzichino, 1963]. Namely, for each point on 
the surface, the wave field at that point is assumed to be what it will be if the 
neighborhood around that point is a local tangent plane.

Hence we start off by investigating how light waves reflect off a smooth plane. 
We then use this knowledge to define suitable boundary conditions for the Kirchhoff 
integral, resulting in a canonical formula for the wave field reflected by a rough 
surface under the Kirchhoff approximation [Beckmann and Spizzichino, 1963]. We
next describe a more convenient specialization of this formula by Stam [1999] for two-dimensional heightfield. Finally, we derive how the reflected wave field relates to the BRDF as customarily used in graphics and vision.

### 3.2.1 Plane Waves

One particular simple family of solutions to the scalar wave equation (Eq. 3.2) is the family of plane waves, which can be written as

$$u(r, t) = F(t - r \cdot \hat{k}/v),$$

(Eq. 3.8)

where $F(t) = u(0, t)$ describes the time-dependent behavior of the wave at the origin, $\hat{k}$ is the unit vector in the direction of wave propagation, and $v$ is the speed of wave propagation as in the scalar wave equation (Eq. 3.2). Note that for any plane perpendicular to the propagation direction $\hat{k}$, the wave field is constant on the plane, hence the name “plane wave”.

When a plane wave is also monochromatic with frequency $f$, we can adopt the complex amplitude notation (Section 3.1.3) for monochromatic waves, $u(r, t) = \text{Re}\{U(r)e^{-i2\pi ft}\}$. Let the scalar $U_0$ be the complex amplitude of the wave at the origin, then $F(t)$ can be written as

$$F(t) = u(0, t) = \text{Re}\{U_0e^{-i2\pi ft}\}.$$

Upon substituting the above into the definition of plane waves (Eq. 3.8), we have an expression for $u$ at an arbitrary point in terms of $U_0$ and a so-called wave vector, $k = 2\pi f\hat{k}/v$, pointing in the direction of propagation,

$$u(r, t) = \text{Re}\{U_0e^{-i2\pi f(t - r \cdot \hat{k}/v)}\}$$

$$= \text{Re}\{U_0e^{i2\pi fr\hat{k}/v}e^{-i2\pi ft}\}$$

$$= \text{Re}\{U_0e^{ikr}e^{-i2\pi ft}\},$$

48
Comparing with the definition of complex amplitude (Eq. 3.4), we arrive at an explicit expression for the time-independent complex amplitude of the monochromatic plane wave

$$\psi(r) = U_0 e^{ik\cdot r}.$$  \hfill (3.9)

We remark the following:

- Recall from Section 3.1.3 that any monochromatic solution to the wave equation (Eq. 3.2) must also satisfy the Helmholtz equation (Eq. 3.5). It is a straightforward computation to check that the above definition does so indeed. Also notice that the length of the wave vector, $\|k\|$, is exactly the wave number, $k$, from the Helmholtz equation (Eq. 3.5).

- A useful fact that follows is that $\psi(r)$ is completely determined by the wave vector $k$ and $\psi(x)$ for any point $x$ in space since

$$\psi(r) = \psi(x) e^{ik\cdot (r-x)}. \hfill (3.10)$$

Thus far, our discussion on the scalar wave field $u$ of a plane wave is applicable to any Cartesian component of both the electric and the magnetic fields. Next we state, without proof, some properties concerning the vector nature of these fields. All are direct consequences of Maxwell’s equations, and interested readers should refer to [Born and Wolf, 1999, § 1.4.1] for the derivation. The electric field, $E$, and the magnetic field, $H$, associated with a plane wave, not necessarily monochromatic, satisfy the following:

- $E$ and $H$ are perpendicular to each other at every point in space.

- The cross product $E \times H$ points in the direction of wave propagation.

- The magnitudes of $E$ and $H$ are related by the equality $\sqrt{\mu} ||H|| = \sqrt{\varepsilon} ||E||$, $\varepsilon_\mu$.
where $\mu$ and $\epsilon$ are the permeability and permittivity constants from the Maxwell equations (Eq. 3.1).

We will make use of these properties in the next sections.

### 3.2.2 Reflection of Plane Waves at a Smooth Plane

We are interested in the representation of monochromatic plane waves because the light wave arriving at a reflecting surface from a distant point light source can be approximated as a superposition of independent monochromatic plane waves. An exact analysis of what happens when a train of plane waves arrives at the boundary between two media of different optical properties from the physical optics point-of-view becomes quite technical, so we only summarize the classical results [Born and Wolf, 1999, §1.5 and §14.2] here.

The incident monochromatic plane wave, denoted by

$$
\psi_1(r) = U_1 e^{ik_1 \cdot r},
$$

is split into two waves, a reflected monochromatic plane wave bouncing back from the boundary, denoted by

$$
\psi_2(r) = U_2 e^{ik_2 \cdot r},
$$

and a transmitted plane wave propagating into the second medium. In the case of conductive surfaces, the transmitted wave attenuates rapidly, so its penetration into the medium is superficial.

The reflected plane wave, which we are interested in, has the same frequency as the incident wave, so $\|k_1\| = \|k_2\|$. Its propagation direction follows the well-known law of reflection from geometrical optics: it lies in the plane formed by the propagation direction of the incident wave and the surface normal, and the angle of incidence equals the angle of reflection. Hence, let the unit surface normal be $\hat{n}$, and
we have
\[ \mathbf{k}_2 = \mathbf{k}_1 - 2(\mathbf{k}_1 \cdot \hat{n})\hat{n}. \]  
(3.13)

**Fresnel’s Equations**

The characterization of the complex amplitude of the reflected wave is more complicated, and is dependent on the vector nature of electromagnetic waves. Hence, we need to carefully specify the coordinate system and the associated Cartesian components of the incident and reflected waves being characterized.

We only describe the relationship between the complex amplitude of the incident and reflected electric fields, denoted by \( \mathbf{E}_1 \) and \( \mathbf{E}_2 \) respectively. The relationship between that of the magnetic fields follows accordingly, as they are functions of the electric fields as outlined at the end of Section 3.2.1. See Fig. 3.2 for an illustration of the geometric configuration used in the following exposition.

![Geometric configuration used in Fresnel’s equations.](image)

**Figure 3.2:** Geometric configuration used in Fresnel’s equations.

First define two Cartesian bases, \( \mathbf{f}_1 \) and \( \mathbf{f}_2 \), as
\[ \mathbf{f}_j = (\hat{\mathbf{e}}_{\perp}, \hat{\mathbf{e}}_{\parallel,j}, \hat{\mathbf{k}}_j), \]
\[ \hat{\mathbf{e}}_{\perp} = \hat{\mathbf{k}}_1 \times \hat{n} = \hat{\mathbf{k}}_2 \times \hat{n}, \]  
(3.14)
\[ \hat{\mathbf{e}}_{\parallel,j} = \hat{\mathbf{k}}_j \times \hat{\mathbf{e}}_{\perp}, \]

where \( j \in \{1, 2\} \). Thus \( \hat{\mathbf{e}}_{\perp} \) is perpendicular to the plane of incidence; \( \hat{\mathbf{e}}_{\parallel,1} \) and \( \hat{\mathbf{e}}_{\parallel,2} \)
are each parallel to the plane of the incidence, and perpendicular to the propagation directions of the incident and reflected waves.

We will represent $E_j$ with respect to the basis $f_j$, $j = 1, 2$. Now recall from Section 3.2.1 that the vector values of the electric and the magnetic fields of a plane wave lie in the plane perpendicular to the propagation direction. Hence, $E_j$ lies in the plane spanned by $\hat{e}_\perp$ and $\hat{e}_\parallel,j$, and can be seen as the superposition of two fields of mutually perpendicular polarizations,

$$E_j = E_{\perp,j}\hat{e}_\perp + E_{\parallel,j}\hat{e}_\parallel.$$

We are now ready to state the relationship between the incident electric field $E_1$ and the reflected electric field $E_2$. Let $\theta$ be the incident angle formed between $\hat{n}$ and $-\hat{k}_1$. For any point $s$ on the surface boundary, we have the following Fresnel’s equations:

$$E_{\perp,2}(s) = F_{\perp}(\theta)E_{\perp,1}(s)$$
$$E_{\parallel,2}(s) = F_{\parallel}(\theta)E_{\perp,2}(s),$$

where $F_{\perp}(\theta)$ and $F_{\parallel}(\theta)$ are the so-called Fresnel coefficients, dependent only on the optical properties of the media and the incident angle $\theta$. Note from the above Eq. 3.15 that the perpendicular (or parallel) polarized field is reflected as a perpendicular (or parallel) polarized field, and there is no contribution from $E_{\perp,1}$ towards $E_{\parallel,2}$, nor from $E_{\perp,2}$ towards $E_{\parallel,1}$.

To define the Fresnel coefficients, let $n_1$ be the refractive index of the medium that the incident and reflect waves travel in. Let $n_2$ be the refractive index of the surface material. Let $m = n_2/n_1$. Then the Fresnel coefficients are given by:

$$F_{\perp}(\theta) = \frac{\cos \theta - \sqrt{m^2 - \sin^2 \theta}}{\cos \theta + \sqrt{m^2 - \sin^2 \theta}}$$
$$F_{\parallel}(\theta) = \frac{m^2 \cos \theta + \sqrt{m^2 - \sin^2 \theta}}{m^2 \cos \theta + \sqrt{m^2 - \sin^2 \theta}}.$$
A few remarks regarding the refractive index and Fresnel’s equations are in order:

1. For dielectric materials, the refractive index, denoted by $n$, is always real, and satisfies $n = c/v$ where $v$ is the propagation speed of the electromagnetic wave in the medium, and $c$ is its speed in vacuum. In particular, the refractive index is 1 for vacuum, and is close to 1 for air.

2. For conductive materials such as metal, the refractive index is a complex number, $\tilde{n} = n + i\kappa$. The imaginary component $\kappa$ is known as the extinction coefficient and serves to model amplitude attenuation during electromagnetic wave propagation because of nonzero conductivity. It must be stressed that the complex refractive index is a piece of neat mathematical notation, allowing formulae developed for dielectric media (with real refractive indices) to work analogously in the presence of nonzero conductivity. It arises partly because we use complex exponential to represent propagating waves (which itself is a clever piece of mathematical notation). The complex refractive index does not represent some “imaginary” physical quantity that is in effect.

3. The characterization using Fresnel coefficients, Eq. 3.15 and Eq. 3.16, is physically correct for media of both real and complex refractive indices. Thus, with a conductive medium, the corresponding Fresnel coefficients can be complex valued, implying a phase change in the reflected field. With dielectric media, the resulting Fresnel coefficient is always real, but can be negative, corresponding to a phase change by $\pi$. The moduli of the Fresnel coefficients dictate amplitude attenuation of the wave field due to reflection.

4. The refractive index, both its real and complex components, is wavelength dependent. Tables exist that list the physically measured refractive index for different materials at different wavelength.
Relating Reflected Wave to Incident Wave

Recall that before we delved into Fresnel’s equations, we denoted the incident and reflected electromagnetic waves simply by the scalar wave fields $\psi_1$ and $\psi_2$. Now we need to be more specific in order to apply Fresnel’s equations (Eq. 3.15). We state that the incident electric fields is the superposition of two vector wave fields, $E_1 = E_{\perp,1}\hat{e}_\perp + E_{\parallel,1}\hat{e}_\parallel$. The perpendicular polarized vector wave field, $E_{\perp,1}\hat{e}_\perp$, is reflected as a perpendicular polarized vector wave field, $E_{\perp,2}\hat{e}_\perp$. The parallel polarized $E_{\parallel,1}\hat{e}_\parallel$ is reflected as the parallel polarized $E_{\parallel,2}\hat{e}_\parallel$. The total reflected vector field is the superposition of the two, $E_2 = E_{\perp,2}\hat{e}_\perp + E_{\parallel,2}\hat{e}_\parallel$. The scalar wave field $\psi_1$, used before to denote the incident wave field, must be understood to mean either $E_{\perp,1}$ or $E_{\parallel,1}$, and $\psi_2$ should refer to either $E_{\perp,2}$ or $E_{\parallel,2}$ accordingly. To avoid confusion, we continue our exposition in this subsection using $E_{\perp,j}$ and $E_{\parallel,j}$ in place of $\psi_j$ to emphasize the Cartesian bases used to represent the incident and reflected waves.

Consider the perpendicular polarized fields first. The wave vector of the reflected field $E_{\perp,2}$ is $k_2$, as defined in Eq. 3.13. We also know from Fresnel’s equations (Eq. 3.15) that the value of $E_{\perp,2}$ at any point $s$ on the surface is $F_\perp(\theta)E_{\perp,1}(s)$. Using the fact that the complex amplitude of a monochromatic plane wave is completely determined by its wave vector and the complex amplitude value for any known point in space Eq. 3.10, we can write out the expression for $E_{\perp,2}$ directly as

$$E_{\perp,2}(\mathbf{r}) = F_\perp(\theta)E_{\perp,1}(s)e^{ik_2 \cdot (\mathbf{r} - s)},$$

(3.17)

where $s$ is any point on the reflecting plane. It is straightforward to check that this definition is independent of the choice of $s$.

For the parallel polarization, we have the analogous expression for $E_{\parallel,2}$,

$$E_{\parallel,2}(\mathbf{r}) = F_\parallel(\theta)E_{\parallel,1}(s)e^{ik_2 \cdot (\mathbf{r} - s)}.$$

(3.18)
3.2.3 The Kirchhoff Approximation

We are now ready to compute the wave field reflected by a rough surface patch. Under the Kirchhoff approximation [Beckmann and Spizzichino, 1963], the wave field at each surface point is approximated by the sum of the incident wave field and the wave field reflected by the tangent plane at that point. Wave field away from the surface can then be expressed in terms of the wave field on the surface and its normal derivative through the *Kirchhoff integral*

See Fig. 3.3. Let $S$ be a rough surface patch centered around the origin $0$. The viewpoint $v$ is at distance $R = ||v||$ from the origin, inducing the viewing direction $\hat{v} = v/R$. The incident wave is given by $\psi_1(r) = U_1 e^{ik_1 \cdot r}$, with the wave number $k$ and the (unit) propagation direction $\hat{k}_1$. Let $U$ be the resulting wave field (due to both the incident wave and scattering by the rough surface).

Now consider a bounding volume $V$ such that $V$ encloses the observation point $v$, and the surface patch $S$ is part of its boundary $\partial V$, i.e., $S \subset \partial V$. The bounding surface in the context of the Kirchhoff integral is taken to be $\partial V$, but we evaluate the integral over $S$ only, assuming the wave field $U$ to be 0 everywhere else on $\partial V$.

![Figure 3.3: Setup for evaluating the Kirchhoff integral under the Kirchhoff approximation.](image)

Let $s \in S$ be a point on the surface. Let $n(s)$ be the (unit) surface normal at $s$. We can approximate the surface neighborhood around $s$ by a tangent plane.
Let $\theta(s)$ be the incident angle formed between $n(s)$ and $-\hat{k}_1$. The reflection of the incident wave at this tangent plane is described by Fresnel’s equations as detailed in Section 3.2.2. In particular, if $\psi_1$ corresponds to the perpendicular polarized component of the incident electric field, i.e., $\psi_1\hat{e}_\perp$, it undergoes modulation by $F_\perp(\theta(s))$ as it is reflected back. If $\psi_1$ corresponds to the parallel component, i.e., $\psi_1\hat{e}_\parallel$, it undergoes modulation by $F_\parallel(\theta(s))$.

But here lies a complication: the tangent plane normal $n(s)$ is slightly different for each surface point. As a result, all following quantities discussed in Section 3.2.2 are different per point: the incident angle $\theta$, the specular reflection direction $\hat{k}_2$, the pair of perpendicular-parallel Cartesian bases for incident and reflected waves $f_1 = (\hat{e}_\perp, \hat{e}_\parallel, \hat{k}_1)$ and $f_2 = (\hat{e}_\perp, \hat{e}_\parallel, \hat{k}_2)$, and the Fresnel coefficients $F_\perp(\theta)$ and $F_\parallel(\theta)$. Moreover, since the Cartesian bases used to define the perpendicular and parallel polarization directions are different per-point, for the same incident electric field, its decomposition into perpendicular and parallel components is slightly different at each $s$. In other words, the scalar field $\psi_1$ itself is different at each $s$. Such pointwise variation will greatly complicate later derivation.

Therefore we make another approximation here. The per-point tangent plane normal $n(s)$ should be close to the average surface normal, denoted by $n_\ast$. The Fresnel coefficients are also smooth functions. Hence, for the purpose of calculating the Fresnel coefficients and decomposing the incident electric field into its perpendicular and parallel components, and only for those two purposes, we use the average surface normal $n_\ast$ in place of the varying tangent plane normal $n(s)$.

Under this approximation, for a given incident electric field, a single pair of perpendicular-parallel bases, $f_1$ and $f_2$, and a single pair of Fresnel coefficients, $F_\perp$ and $F_\parallel$, derived form the average surface normal $n_\ast$, are used for the entire surface. We can now properly qualify the scalar fields $\psi_1$ and $U$ in the context of the vector nature of the electric fields.
Consider the perpendicular case first. The scalar wave field $\psi_1$ then represents the perpendicular component, with respect to $f_1$, of the incident electric field. The scalar wave field $U$ represents the perpendicular component, with respect to $f_2$, of the overall unknown electric field to be solved for. Define $F = F_\perp$ for convenience.

At each surface point $s$, denote by $\psi_{2,s}$ the perpendicular component, with respect to $f_2$, of the electric field reflected by the tangent plane at $s$. Then $\psi_{2,s}$ is related to $\psi_1$ via the explicit expression Eq. 3.17 derived in Section 3.2.2. The value of $U$ on the surface is simply $U(s) = \psi_1(s) + \psi_{2,s}(s)$. Explicitly evaluating $U$ and $\partial U/\partial n$ at $s$ then gives:

$$U(s) = (1 + F)U_1 e^{ik_1 \cdot s}$$

$$\frac{\partial U}{\partial n}(s) = i(1 - F)(k_1 \cdot n(s))U_1 e^{ik_1 \cdot s}.$$ (3.19)

Note that the per-point normal $n(s)$ is used here, not the average surface normal $n_*$. It is also worth pointing out that the tangent plane approximation is employed only for estimating the gradient term $\partial U/\partial n$, and is not need for $U$. Indeed the expression for $U$ is independent of the normal. It is simply the sum of the incident wave field value at $s$, $\psi_1(s)$, and its product with the Fresnel coefficient, $F\psi_1(s)$.

The other ingredient of the Kirchhoff integral is the auxiliary function $G(r) = e^{ik\|r-v\|/\|r-v\|}$ in Eq. 3.6. Recall $R = \|v\|$. Using Taylor expansion, we can approximate $\|s-v\|$ by $(R - s \cdot \hat{v})$. Using the fact that $R \gg s \cdot \hat{v}$, we can approximate $G$ and $\partial G/\partial n$ as

$$G(s) \approx \frac{e^{ikR}}{R} e^{-iks \cdot \hat{v}}$$

$$\frac{\partial G}{\partial n}(s) \approx -ik(\hat{v} \cdot n(s))G(s).$$ (3.20)

Finally we substitute Eq. 3.19 and Eq. 3.20 back into the Kirchhoff integral Eq. 3.7 to get the formula for rough surface scattering under the Kirchhoff approximation
[Beckmann and Spizzichino, 1963]:

\[
U(\mathbf{v}) = \frac{1}{4\pi} \iint_S \left[ U\frac{\partial G}{\partial \mathbf{n}} - G\frac{\partial U}{\partial \mathbf{n}} \right] \, ds
= \frac{ikU_1 e^{ik\|\mathbf{v}\|}}{4\pi \|\mathbf{v}\|} \left[ F(\hat{k}_1 - \hat{\mathbf{v}}) - (\hat{k}_1 + \hat{\mathbf{v}}) \right] \cdot \iint_S e^{ik(\hat{k}_1 - \hat{\mathbf{v}}) \cdot \mathbf{n}(s)} \, ds.
\]  

(3.21)

The formula for the parallel polarization is identical, except with \( \psi_1 \) and \( U \) representing the parallel components (with respect to the Cartesian bases \( f_1 \) and \( f_2 \) respectively), and with \( F = F_\parallel \). Hence, from now on we will not explicitly mention the polarization choice when using Eq. 3.21 unless it becomes relevant in the discussion.

At this point, it is instructive to explicitly list the main approximations taken during the derivation of the Kirchhoff-approximation-based formula for light reflection (Eq. 3.21):

- At each surface point, the wave field is assumed to reflect in a manner as if the local surface neighborhood is replaced by the tangent plane at the surface point.

- The present derivation ignores shadowing and multiple scattering between different parts of the surface. Shadowing can be (partially) accounted for by letting the incident wave field \( U \) and its normal derivative vanish in areas shadowed by other parts of the surface [Brockelman and Hagfors, 1966; Wagner, 1967], although this simple geometrical treatment does not consider effects of diffraction. Multiple scattering can be simulated by taking the reflected plane wave at each surface point, propagating it again to another surface point taking self-shadowing into account, and propagating the secondarily-reflected wave at the new surface point to the viewpoint [Bruce, 1995].

- The wave field \( U \) and its normal derivative are assumed to vanish on the bound-
ary of the bounding volume outside of the surface area being integrated, that is, \((\partial V \setminus S)\). This is an oversimplification of the physical reality. Mathematically, if a solution of a three-dimensional wave equation vanishes on a finite surface element, it must in fact vanish in all space [Goodman, 1969, § 3.5].

- The Green’s function \(G\) is evaluated using the following two approximations: (i) The distance between the surface point \(s\) and the viewpoint \(v\) is approximated by \((\|v\| - s \cdot \hat{v})\), which is the first-order Taylor expansion of the function \(r \mapsto \|r - v\|\) around \(0\); (ii) The fraction \(1/\|s - v\|\) is approximated by \(1/\|v\|\). Both approximations are good as long as the surface patch is small compared to the viewing distance, i.e., \(\|s\| \ll \|v\|\).

- The average surface normal \(n_*\) is used in place of the pointwise-varying surface normal \(n(s)\) when computing the Fresnel coefficients and decomposing the incident electric field into its perpendicular and parallel components.

Finally it is worth noting that the Kirchhoff integral is only meaningful over an area in which the incident wave is approximately a coherent plane wave. The actual coherence area depends on the lighting configuration, and results such as the van Cittert-Zernike theorem [e.g., Born and Wolf, 1999, § 10.4.2] can be used to estimate the coherence area of an incoherent light source. On the other hand, the irregularity of a rough surface implies that regions far apart are unlikely to scatter waves in a coherent manner such that their interferences contribute in any meaningful manner to the final reflectance, so integrating over a window larger than the actual coherence area will introduce little error.

### 3.2.4 Stam’s Specialization for Heightfield

When the rough surface is represented as a height field, Stam [1999] developed a more convenient specialization of Eq. 3.21 using Fourier analysis. We next briefly
derive Stam’s formula. For somewhat improved semantics, we adopt the notation from [Levin et al., 2013] by defining the light direction vector \( \hat{\mathbf{l}} = -\hat{\mathbf{k}}_1 \), so \( \hat{\mathbf{l}} \) and \( \hat{\mathbf{v}} \) point away from the surface towards the light source and the viewpoint respectively. Also define the half vector \( \mathbf{h} = (\hat{\mathbf{v}} + \hat{\mathbf{l}})/2 = [h_x, h_y, h_z]^T \). \( \hat{\mathbf{l}} \) and \( \mathbf{h} \) will be used in place of \( \hat{\mathbf{k}}_1 \) where appropriate.

Let \( z : X \to \mathbb{R} \) be a heightfield defined over a rectangular domain \( X \subset \mathbb{R}^2 \). The corresponding parametric surface is given by \( \mathbf{s}(\mathbf{x}) = [x; z(x)] \in \mathbb{R}^3, \mathbf{x} \in X. \)

We can express the differential element \( (n(\mathbf{s}) \, ds) \) in the Kirchhoff approximation formula (Eq. 3.21) using partial derivatives of \( z \) and the area element \( dx = dx \, dy \),

\[
\mathbf{n}(\mathbf{s}) \, ds = [-\nabla z(\mathbf{x}); 1] \, d\mathbf{x},
\]

where \( \nabla f = [-\partial f/\partial x; -\partial f/\partial y] \) is the gradient vector of \( f \). The reflected wave under the Kirchoff approximation (Eq. 3.21) for the heightfield can thus be written as

\[
U(\mathbf{v}) = \frac{ikU_1 e^{ik\|\mathbf{v}\|}}{4\pi \|\mathbf{v}\|} [(1 - F)\hat{\mathbf{l}} - (1 + F)\hat{\mathbf{v}}] \cdot \mathbf{U}(\mathbf{v})
\]

\[
= \int_X e^{-2ik\mathbf{h} \cdot [\mathbf{x}, z(\mathbf{x})]^T} [-\nabla z(\mathbf{x}); 1] \, d\mathbf{x}.
\]  (3.22)

Next define the surface modulation function \( m \),

\[
m(\mathbf{x}) = \begin{cases} 
e^{-2ikh_z z(\mathbf{x})}, & \text{if } \mathbf{x} \in X \\
0, & \text{otherwise.} \end{cases} \]  (3.23)

Then the integral \( U(\mathbf{v}) \) from Eq. 3.22 can be expressed as a Fourier transform,

\[
U(\mathbf{v}) = \int_{\mathbb{R}^2} e^{-2ik\mathbf{h}\cdot \mathbf{x}} m(\mathbf{x}) [-\nabla z(\mathbf{x}); 1] \, d\mathbf{x} \]

\[
= \mathcal{F}\{m[-\nabla z; 1]\}(2k\mathbf{h}_{xy}), \]  (3.24)

where \( \mathbf{h}_{xy} = [h_x, h_y]^T \) is the two-dimensional vector consisting of the X and Y

\^Note that we use the semi column “;” to represent the vertical concatenation of column vectors or scalars into a column vector, in order to avoid writing the transposition sign twice, i.e., \( [x^T, z(x)]^T \).
components of the half vector $h$, and $F\{f\}(\omega) = \iint_{\mathbb{R}^2} f(x)e^{-i\omega \cdot x}dx$ is the Fourier transform of a (scalar or vector-valued) function $f$ evaluated at $\omega$.

We can further simplify the above expression by noting the follows: First, without loss of generality we can assume that $z(x)$ has some defined value for $x$ outside of $X$. Then by the definition of $m$, the following holds almost everywhere over the entire $\mathbb{R}^2$,

$$\nabla m = -2ikh_z m \nabla z. \tag{3.25}$$

Second, the Fourier transform of the derivative of a function satisfies that

$$F\{\nabla f\}(\omega) = i\omega F\{f\}(\omega). \tag{3.26}$$

We can now write the integral $U$ from Eq. 3.24 as

$$U(v) = F\left\{ \frac{\nabla m}{2ikh_z} ; m \right\}(2kh_{xy}) = \frac{h}{h_z} F\{m\}(2kh_{xy}), \tag{3.27}$$

where the first equality is due to Eq. 3.25 and the second to Eq. 3.26.

Substituting the above back into the Kirchhoff approximation formula for height-fields, Eq. 3.22, and noting that

$$[(1 - F)\hat{l} - (1 + F)\hat{v}] \cdot h = -F(1 + \hat{l} \cdot \hat{v}),$$

we arrive at Stam’s formula,

$$U(v) = -ikU_1 e^{ik|v|} F(1 + \hat{l} \cdot \hat{v}) \frac{F\{m\}(2kh_{xy})}{4\pi|v|/h_z}, \tag{3.28}$$

which relates the reflected wave field to the Fourier transform of the surface modulation function $m$, which depends on $h_z$.\(^3\)

We remark that in the above derivation, because $m$ is defined to be zero outside

\[^3\text{Stam evaluated the Fourier transform at } -2kh_{xy} \text{ instead of } 2kh_{xy}. \text{ The sign flip is due to the choice of Fourier transform convention: Stam uses the convention that } F\{f\}(\omega) = \iint f(x)e^{i\omega \cdot x}dx, \text{ whereas we use the convention that } F\{f\}(\omega) = \iint f(x)e^{-i\omega \cdot x}dx.\]
of \( X \), the extension of the integration domain from \( X \) (Eq. 3.22) to the entire plane \( \mathbb{R}^2 \) (Eq. 3.24) is well-defined and keeps intact the value of \( U(v) \).\(^4\) Computing \( \mathcal{F}\{m\} \), e.g., via the *discrete Fourier transform* (DFT), on the other hand does involve the usual well-understood trade-offs between accuracy and computational complexity such as sampling rate versus aliasing.

### 3.2.5 From Reflected Wave to the BRDF

The graphics and vision community customarily describes the reflectance property of opaque materials using the *bidirectional reflectance distribution function* (BRDF) as reviewed in Section 2.1.1. We next show how the reflected wave field is related to the BRDF. The reasoning follows that of Tomiyasu [1988], who derived the relationship between a different radiometric concept, the *differential scattering coefficient*, and the BRDF.

Electromagnetic radiation can be seen as carrying energy in its oscillating fields. The term *energy flux density*, or simply *energy flux*, refers to the rate of energy flow through a unit area perpendicular to the direction of energy flow. The energy flux density is measured in watt \( \cdot \) m\(^{-2} \). For a monochromatic plane wave \( \psi(r) = U e^{ik \cdot r} \), the energy flow direction is the same as the wave propagation direction \( k \), and the energy flux density equals \( Q|U|^2 \), where \( Q \) is a constant dependent on the wavelength. In other words, the energy flux density is proportional to the squared amplitude of the wave,

Of course the electromagnetic wave is defined by two propagating vector fields, so what do we mean when referring to the “amplitude of the wave”? A more formal

\(^4\)Stam remarked that the extension used “the common assumption (e.g., [2, 8]) that the integration can be extended over the entire plane” and that “[t]his assumption is usually justified on the grounds that the surface detail is much smaller than the distances over which the surface is viewed.” Based on our derivation, in particular the choice of defining \( m \) to be 0 outside of \( X \), we do not see the need for such an assumption.
description of the energy flux density utilizes the concept of the Poynting vector [Born and Wolf, 1999, p. 9], defined as \( \mathbf{E} \times \mathbf{H} \) scaled by a unit dependent constant, where \( \mathbf{E} \) and \( \mathbf{H} \) are the electric and magnetic field respectively. The Poynting vector points in the direction of the energy flow, and its length equals the energy flux density.

For monochromatic plane waves, recall from Section 3.2.1 that \( \mathbf{E} \) and \( \mathbf{H} \) are mutually orthogonal, \( \mathbf{E} \times \mathbf{H} \) points in the direction of wave propagation, and \( \sqrt{\mu} \| \mathbf{H} \| = \sqrt{\varepsilon} \| \mathbf{E} \| \).

It follows that the time-average of the Poynting vector has length proportional to the squared amplitude of either \( \mathbf{E} \), \( \mathbf{H} \), or any of their non-zero Cartesian components. Thus we are justified in simply stating that the energy flux density is proportional to the squared amplitude of the wave as in the previous paragraph.

Refer back to the geometry for defining the BRDF (Fig. 2.1). Let the incident wave field be a monochromatic plane wave \( \psi_1(\mathbf{r}) = U_1 e^{i\hat{k}_1 \cdot \mathbf{r}}, \hat{k}_1 = -\hat{l} \) as in the setup for evaluating the Kirchhoff integral (Section 3.2.3). Recall from Eq. 2.1 that the BRDF is defined as \( f(\hat{l}, \hat{v}) = d\Phi_v / (d\Phi \, d\Omega_v \cos \theta_v) \), where \( d\Phi \) and \( d\Phi_v \) are the flux received by the surface area elements \( dA \) and \( d\Omega_v \) respectively. The incident wave field has energy flux density of \( Q\|U_1\|^2 \), and its propagation direction forms angle \( \theta_l \) with the surface normal. Hence the flux received by the surface area element \( dA \) is just \( d\Phi = Q\|U_1\|^2 \cos \theta_l \, dA \). The receiving area element \( \Omega_v \) lies in the unit hemisphere and is orthogonal to the propagation direction of the reflected wave field (assuming \( dA \) is small). Hence the flux received by \( d\Omega_v \) is just \( d\Phi_v = Q\|U(\hat{v})\|^2 \, d\Omega_v \), where the reflected wave field \( U \) is as defined by the Kirchhoff approximation (Eq. 3.21).

Substituting \( d\Phi \) and \( d\Phi_v \) back into the definition of \( f(\hat{l}, \hat{v}) \), we obtain

\[
f(\hat{l}, \hat{v}) = \frac{|U(\hat{v})|^2}{\cos \theta_v \cos \theta_l \, dA \, |U_1|^2}.
\] (3.29)

We remark the following:

- The surface element \( dA \) is assumed to be small compared to the distance
between the surface patch and the viewpoint $\hat{v} \in d\Omega_v$, which is 1.

- $|U(\hat{v})|^2$ as defined per the Kirchhoff approximation (Eq. 3.21) is proportional to $|U_1|^2$. Hence the resulting BRDF $f$ is independent of $|U_1|^2$.

- The relationship between $f$ and $dA$ is nuanced. $U$ is defined as an integral over the surface patch $dA$. Given patches $dA$ and $dA'$, each resulting in the reflected wave field $U$ and $U'$. It is possible for the surface geometry over the patches $dA$ and $dA'$ to be correlated in such a way that $|U + U'|$ ranges from 0 to $(|U| + |U'|)$ for different regions in space, because of constructive and destructive interferences of the wave fields $U$ and $U'$. This means that a linear increase in $dA$ can result in at most a quadratic increase of $|U|^2$, or at worst the total annihilation of $|U|^2$ for some directions. However, when the surface is randomly rough, the geometry corresponding to patches $dA$ and $dA'$ should result in largely uncorrelated reflectance fields $U$ and $U'$, in which case we expect $|U + U'|^2 = |U|^2 + |U'|^2$. In other words $|U|^2$ is proportional to $dA$.

### 3.2.6 Recap

Let us take a pause and recap the sequence of derivations in this section.

In order to analyze how light waves reflect off a patch of rough surfaces, we first described how the simplest monochromatic plane waves reflect off a smooth plane, a process defined by the Fresnel equations. More complex incident waves can simply be treated as superposition of these simpler waves.

For a rough surface patch, the wave field at a surface point can be approximated by the sum of the incident plane wave and a plane wave reflected by the tangent plane at that surface point. Since the tangent plane varies across the surface, so does the reflected plane wave. We thus have a varying complex amplitude field defined on the entire surface patch.
We next evaluated the Kirchhoff integral (Section 3.1) to obtain the wave field at any point away from the surface, using the wave field defined on the surface and its normal derivative. We also derived Stam’s specialization for two-dimensional height fields.

Finally we showed how the reflected wave field is related to the BRDF.

### 3.3 Summary

In this chapter, we derived the results from physical optics that are required to properly specify the problem addressed by this thesis and the methods employed. Starting from the Maxwell equations, we eventually arrived at the Kirchhoff approximation based formula for wave scattering by rough surfaces, and Stam’s specialization for heightfields. Having developed enough physical-optics related machinery, we are now ready to formally describe in the next chapter the experimental measurements we collect and the computation we perform to predict reflectance from profilometer measurements.
Chapter 4

Problem Statement

In this chapter we present a formal theoretical overview of the problem addressed by this thesis. In particular, we derive how reflectance of a two-dimensionally rough surface is specified by the one-dimensional surface profiles under the Kirchhoff approximation. We also specify at a theoretical level the experimental measurements we collect, the computation we perform to predict surface reflectance, and how we evaluate the predicted reflectance using optically-measured ground-truth.

4.1 Characterizing Surface Using One-Dimensional Profiles

The input to Stam’s specialization of the Kirchhoff integral (Eq. 3.28) is the surface geometry represented a two-dimensional heightfield. Since the Kirchhoff approximation takes into account interferences of reflected wave field originating from every point on the surface, ideally the heightfield should be sampled with sub-wavelength resolution and accuracy for the computation to be meaningful.

Atomic force microscopy (AFM) can measure with nanoscale resolution the two-dimensional topography of a surface, and in theory is ideal for acquiring the data
needed to evaluate the Kirchhoff integral. AFM works by moving a tiny mechanical probe across the measured surface sample. A feedback mechanism ensures that the probe is always barely “touching” the sample. AFM scans through a predefined scanning window, one scanline at a time. By recording the vertical position of the probe, AFM is able to produce a two dimensional grid of height values.

Experimentally, we however observed that for the kind of rough surfaces we investigate there exist significant and uncorrectable misalignment between the one-dimensional scanlines that comprise the two-dimensional scan from the AFM. A detailed discussion of this phenomenon is provided in Chapter 5. Because of this phenomenon, the output of the AFM is modeled as providing a set of independent one-dimensional scans, \( \{z_j\} \),

\[
z_j(x) = z(x, y_j) + d_j,
\]

for some fixed \( \{y_j\} \) and some unknown displacements \( \{d_j\} \).

### 4.2 Computing Reflectance from One-Dimensional Profiles

We now derive how surface reflectance can be specified by one-dimensional profiles. From Stam’s formula (Eq. 3.28), the intensity at the viewpoint \( v \) can be written as

\[
|U(v)|^2 = g(v, \hat{l}) P(2k \mathbf{h}_{xy}),
\]

where \( P = |\mathcal{F}\{m\}|^2 \) and the coefficient \( g(v, \hat{l}) \) depends only on the refractive index of the material, due to \( F(\theta_l) \), and the geometry of the setup,

\[
g(v, \hat{l}) = \frac{k^2|U_1|^2|F(\theta_l)|^2(1 + \hat{l} \cdot \hat{v})^2}{16\pi^2|\mathbf{v}|^2h_x^2}. \quad (4.2)
\]
Since \( P \) is the power spectral density (PSD) of \( m \), which itself is the surface modulation function of \( z \), we will refer to \( P \) as the modulated power spectral density (MPSD) to avoid confusion with the regular PSD of a function.

Surface reflectance is thus completely determined by the MPSD \( P \). The two-dimensional function \( P \) is related to the one-dimensional (displaced) profiles \( \{ z_j \} \) as follows: From \( \{ z_j \} \), we can compute the one-dimensional PSDs, \( \{ p_j \} \), of horizontal slices of \( m \), \( \{ m(\cdot,y_j) \} \),

\[
p_j = |\mathcal{F}_x\{e^{-2ikh_2z_j(x)}\}|^2 = |\mathcal{F}_x\{e^{-2ikh_2z(x,y_j)}\}|^2
= |\mathcal{F}_x\{m(x,y_j)\}|^2.
\]

Next define a projection operator, \( \text{prj} \), that takes a bivariate function \( f \) defined on \( \mathbb{R}^2 \) and projects it along its second variable,

\[
(\text{prj} f)(x) = \int f(x,y)dy.
\]

We can then express \( \text{prj} P \), the projection of the two-dimensional MPSD \( P \), in terms of one-dimensional PSDs of horizontal slices of \( m \):

\[
(\text{prj} P)(\omega_1) = \int |\mathcal{F}_y\{\mathcal{F}_x\{m(x,y)\}(\omega_1)\}|^2(\omega_2)d\omega_2
= \int |\mathcal{F}_x\{m(x,y)\}(\omega_1)|^2 dy,
\]

where the last equality is due to Parseval’s theorem.

If we assume that \( \{ y_i \} \) sample the scanning range in \( Y \) uniformly, then the last integral can be approximated by the uniformly-weighted average, denoted by \( p \), of the computed one-dimensional PSDs \( \{ p_j \} \),

\[
p(\omega) = \frac{W}{N} \sum_{j=1}^{N} p_j(\omega) = \frac{W}{N} \sum_{j=1}^{N} |\mathcal{F}_x\{m(x,y_j)\}|^2 \approx \int |\mathcal{F}_x\{m(x,y)\}(\omega)|^2 dy,
\]

where \( N \) is the number of profiles measured, and \( W \) is the vertical dimension of the window over which \( m \) and \( z \) are defined.
In summary, from the one-dimensional displaced height profiles \( \{ z_j \} \) alone, we can compute a numerical approximation to \( \text{prj} P \) that will converge to the true value as the number of measured profiles increases.

Finding \( P \) given \( \text{prj} P \) is an ill-posed linear inverse problem. The estimated \( \text{prj} P \) will also be noisy because the surface is rough and only a limited number of profiles can be acquired. For example, Fig. 4.1 shows the typical estimated MPSD projections computed from measured one-dimensional profiles. Thus we need additional priors and constraints to discard unlikely and undesirable candidates that satisfy the estimated \( \text{prj} P \) by, for example, over-fitting to its noises.

\[ \begin{array}{c}
\text{Figure 4.1: Estimated MPSD projections (for some fixed half vector) computed from one-dimensional profiles. The labels “10um,” “17um,” and “mixture” identify different surface samples (detailed in Chapter 5).}
\end{array} \]

### 4.3 Assumptions on Surface Topography

One strategy for constraining the unknown MPSD \( P \) is to make additional assumptions regarding the surface topography, e.g., on its level of azimuthal anisotropy. To demonstrate how this assumption can be used, we next describe a naive algorithm which assumes that \( P \) is fully rotational symmetric and hence leads to fully isotropic reflectance.
Let $\text{rot}_\beta: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be the two-dimensional counterclockwise rotation,

$$
\text{rot}_\beta \mathbf{x} = \begin{bmatrix}
\cos \beta & -\sin \beta \\
\sin \beta & \cos \beta
\end{bmatrix} \mathbf{x}.
$$

Let $\text{rvl}$ be an operator that takes a univariate function $f: [0, \infty) \rightarrow \mathbb{R}$ and revolves it around the origin to produce a bivariate function,

$$(\text{rvl} f)(\text{rot}_\theta [r, 0]^T) = f(r) \text{ for } \theta \in [0, 2\pi) \text{ and } r \geq 0.$$ 

If the two-dimensional MPSD $P$ is rotationally symmetric, then $P = \text{rvl} q$ for some univariate $q$. We only need to solve for $q$, instead of $P$, using the linear equation

$$\text{prj} \circ \text{rvl} q = p. \quad (4.5)$$

Because $q$ and $p$ are both univariate functions, we expect the above equation to be more tractable than directly solving for $P$.

Note however that the revolve-and-project operator $(\text{prj} \circ \text{rvl})$, when discretized, is essentially a blur operator with a large and variably-sized kernel and hence extremely ill conditioned. Intuitively, if we denote by $\phi_x$ some bell shaped basis function of narrow support centered at $x$, then $(\text{prj} \circ \text{rvl})$ will "splatter" $\phi_x$ across the entire range of $[-x, x]$. Hence, $\|\text{prj} \circ \text{rvl} \phi_x\|$ for small $|x|$ is going to be significantly smaller than for large $|x|$. Any residue in the linear system, $(\text{prj} \circ \text{rvl} q - p)$, will cause a least-square solver to over-enthusiastically make large adjustment to $q$ for small $|x|$, i.e., around the origin. This will lead to large oscillation around the peak of both $q$ and $P = \text{rvl} q$. Moreover, a nontrivial residue will always be present due to measurement noises and intrinsic surface anisotropy, however small.

To remove such oscillation, we can adopt priors that prefer smoother functions over noisier ones even though the latter achieve smaller residues: One possibility is to choose basis functions (used for representing the unknown $q$) to have wider support.
and attenuate slower; Another is to add an explicit smoothness regularization term; Yet another is to preemptively smooth out the noisy projected MPSD $p$ computed from measured profiles.

The danger of adopting arbitrary priors is that, by themselves, these priors are purely numerical and do not have physical interpretation. The set of prior parameters that work best for one test surface might not be good for another. Given our goal of quantitative prediction, we must caution against introducing arbitrary parameters to be fitted against the optical ground truths.

In Chapter 6, we introduce the shaped microfacet model, which treats a rough surface as an ensemble of shaped and oriented smooth microfacets, each reflecting light diffractively (due to its size, not roughness) per the Kirchhoff approximation. Under this model, the MPSD $P$ can be expressed as the linear combination of elements from a “dictionary” of suitably constructed basis functions. Moreover, physically meaningful parameters of the model, which can be chosen based on the profilometer measurements alone, translate in a principled way to numerical properties of these basis functions, such as their widths and shapes as mentioned earlier for the naive algorithm. In this way, the shaped microfacet model provides the prior for tackling the linear inverse problem. Unlike the naive algorithm, this model also allows various degrees of surface anisotropy to be expressed.

### 4.4 Profile Measurement Along Multiple Azimuthal Directions

Further constraints on the MPSD $P$ can come from measuring surface profiles along additional azimuthal directions. We extend $\text{prj}$ to integrate along the Y axis rotated
counterclockwise by an angle $\beta$,
\[
\text{prj}_\beta P = \text{prj}(P \circ \text{rot}_{-\beta}),
\]
so that the previously defined $\text{prj}$ is now $\text{prj}_0$. If we can measure profiles along an azimuthal direction $\beta$, we can estimate $\text{prj}_\beta P$ and use it to constraint $P$. Hence the measurement procedure and the algorithm presented in this thesis easily support anisotropic materials if profiles along more azimuthal directions are measured.

For this particular study, our AFM hardware does not have a precision-controlled rotation stage, so we obtained $\text{prj}_\beta P$ for roughly $\beta = 0$ and $\frac{\pi}{2}$ by manually rotating the surface samples. Moreover, our surface samples exhibit high isotropy, so the estimated $\text{prj}_0 P$ and $\text{prj}_{\frac{\pi}{2}} P$ are very close anyhow. As we will discuss in Section 7.3.2, noises in the estimated $\text{prj}_\beta P$ is the main obstacle to good recovery of $P$. Hence we compute a single estimated $\text{prj} P$ from profile measurements along both $\beta = 0$ and $\beta = \frac{\pi}{2}$ to reduce noise, and use it for reflectance computation. To simplify notation, unless otherwise noted we omit the subscript $\beta$ and just write $\text{prj}$ for the rest of the thesis.

4.5 Comparison of Predicted and Optically Measured Surface Reflectance

The Kirchhoff approximation specifies the wave field reflected by a rough surface. Hence we can compute the radiant flux incident on a receiving surface due to this wave field. A light sensor such as a camera can also measure the radiant flux incident on a receiving surface. We will compare the computed and the optically measured radiant flux to evaluate our prediction.

We have built a simple gonioreflectometer consisting of a statically mounted camera that serves as the light sensor, a collimated LED light source mounted on
a motorized rotating arm that produces the incident light waves, and a fixed base to house the surface sample. We leave the details of this setup to Chapter 5, but will focus on describing exactly what physical quantities are measured optically by the gonioreflectometer. The same quantity is then computed from profilometer measurements under the Kirchhoff approximation and compared to the optical ground truth.

Refer back to the formulae relating the MPSD to the intensity of the reflected wave field Eq. 4.1 and Eq. 4.2. Our gonioreflectometer steps the reflection direction $\hat{l}$ through its entire range within the plane of incidence, while keeping the viewpoint $v$ and the intensity and the wavelength of the incident wave field, $|U_1|^2$ and $k$, constant (since the camera is statically mounted and the light source is at a constant distance from the origin). For each $\hat{l}$, the gonioreflectometer records the radiant flux, denoted by $\Phi(\hat{l})$, reflected to the camera aperture within a small fixed solid angle along $\hat{v}$.

We can also compute $\Phi(\hat{l})$ from the profilometer measurements via Eq. 4.1 to obtain,

$$\Phi(\hat{l}) = G \left[ \frac{(1 + \hat{l} \cdot \hat{V})}{h_z} \right]^2 |F(\theta_l)|^2 P(2k h_{xy}), \quad (4.6)$$

where $G$ is a constant dependent only on the illumination and geometry setups, e.g., $\|v\|, |U_1|^2$, and $k$. We remark the following:

- The Fresnel coefficient $F(\theta_l)$ depends on the material composition of the sample surface. It also takes on two different values, $F_{\parallel}(\theta_l)$ and $F_{\perp}(\theta_l)$, depending on if the incident and reflected fields represent the parallel or perpendicular polarized components (see Section 3.2.3). In this study, the incident light is unpolarized, so we can assume that the incident electric field decomposes into a parallel and perpendicular components of equal strength. It follows that we should evaluate Eq. 4.6 for both the parallel and perpendicular polarizations and average the results, which has the same effect as replacing the term $|F(\theta_l)|^2$
above by \( \frac{|F_{\parallel}(\theta)|^2 + |F_{\perp}(\theta)|^2}{2} \).

- We assume that the Kirchhoff integral is over the maximal AFM scanning window, 80um by 80um in size, even though the actual coherence area might be smaller than that. See Section 3.2.3 for discussion on the coherence area.

- If we have multiple surfaces with the same material composition, then we can use the optical measurements of a single surface to fix \( G \). For all other surfaces, \( \Phi(\hat{l}) \) should be computable solely from the profilometer measurements.

We will compare the computed and the measured versions of \( \Phi(\hat{l}) \), and colloquially refer to this quantity as “reflectance” or “intensity.”

### 4.6 Summary

The Kirchhoff approximation expresses reflectance as an integral over the two-dimensional surface. However, using AFM we were only able to obtain reliable measurements of one-dimensional surface profiles. Hence in this chapter we first derived the relationship between this integral and the one-dimensional surface profiles, resulting in a computational problem equivalent to inverting the projection operator \( \text{prj} \). We then briefly described our strategies for solving this ill-posed linear inverse problem, by making prior assumptions on surface properties and by acquiring profile measurements along more azimuthal directions. Finally, we described what physical quantity corresponds to the “reflectance” optically measured by our gonioreflectometer, and how the same quantity can be computed from profilometer measurements using results developed in this chapter.

In the next chapter, we will detail the experimental portion of this study, namely, the fabrication of surface samples via metallic coating, the gonioreflectometer setup, and the procedures for measuring one-dimensional surface profiles via AFM. Then
in Chapters 6 to 8, we tackle the linear inverse problem of computing the MPSD from its projections, which then yields the surface reflectance.
Chapter 5

Experimental Procedures

This chapter details the experimental portion of our study. In Section 5.1, we describe how we manufactured our surface samples via metallic coating. In Section 5.2 we describe the simple gonioreflectometer we built for measuring the reflectance distribution of our surface samples. In Section 5.3 we present a new procedure for accurately measuring via AFM the one-dimensional profiles of surfaces much rougher than traditionally investigated via AFM.

5.1 Metallic Paint Production

Metallic coating is a popular exterior finish for automotive applications and consumer goods [Sung et al., 2002] such as electronics and sporting equipments. Metallic paint is made by mixing metallic pigments with a binder medium. Dyes can be added to the binder to produce colored appearance.

The most commonly used metallic pigments are (Al) flakes and powders, although other metals and compounds such as copper (Cu), silver (Ag), gold (Au), and titanium dioxide (TiO₂) are also employed to achieve different surface appearances. The choice of paint binder depends on the intended application. Automotive exterior coating
typically uses polyurethane-based binder. Metallic printing ink employs water or paste-based binder.

We manufactured our surface samples by airbrushing metallic paint on smooth glass substrates. The metallic paint is made from commercially available aluminum flakes and a polyurethane-based automotive clear coat formulation. In the rest of this section we detail the materials and the procedures used to produce our surface samples, and discuss the rational behind our choices.

5.1.1 Aluminum Flakes

Commercially available aluminum flakes come in different morphological shapes and size distributions, determined by the manufacturing process. They can also differ in their “leafing” properties (explained later).

Morphological Types

The metallic pigment industry commonly manufacture aluminum flakes in three morphological types: *cornflake*, *silverdollar*, and *VMP* (Vacuum Metallized Pigments) [Carl Schlenk AG, 2016]. Fig. 5.1 shows images of these three types of flakes captured using scanning electron microscopy. Cornflake pigments resemble leaves or flakes, and have jagged irregular boundary. Silverdollar pigments are lenticular in shape, have smoother boundaries, and look like pancakes. VMP pigments are ultra thin and smooth, and can produce mirror-like reflection.

The size distribution of metallic pigments are characterized using $D_{50}$ values as follows: Half of the particles, as measured by cumulative mass, have diameters smaller than the $D_{50}$ value, while the rest have diameters larger than it. $D_{50}$ values are typically measured using optical scattering models which approximate individual flakes as spheres irrespective of their true shape. Hence $D_{50}$ values provide only a rough characterization of the size distribution of the flakes. In practice, commercial
flake products can contain flakes both much smaller and larger than specified by their D50 values.

**Leafing Property**

Metallic flakes used for coating also come in leafing and nonleafing variants. See Fig. 5.2. Leafing flakes tend to float to the surface of the liquid binder, much like how leaves float on water surface. Nonleafing flakes will disperse within the liquid binder. The leafing property is induced by adding oleic acid to the flakes during its manufacturing (milling) process [Oil and Colour Chemists’ Association, 2012].

With nonleafing flakes, the cured paint contains flakes embedded inside the polymer film. There have been some attempts at estimating the reflectance of metallic paint made from nonleafing flakes [e.g., Sung et al., 2002]. Measuring the
exact topography of flakes embedded in the polymer film is a formidable task. A destructive method such as cross-cut analysis [Kettler and Richter, 1997] is labor-intensive and has low data-acquisition bandwidth. A nondestructive alternative is confocal laser scanning microscopy (CLSM) as used by Sung et al. [2002] and Kirchner and Houweling [2009]. CLSM detects light reflected by surface points lying on a single focus plane. A focal sweep at discrete steps allows recovery of the surface topography, one levelset at a time.

The lateral and vertical resolution of CLSM are restricted by the diffraction limit of the optical setup, and are hard to push beyond the scale of the wavelength of the laser light source. For example Sung et al. used oil immersion objective and a scanning laser wavelength of 543 nm to achieve a theoretical vertical (Z) resolution of 286 nm. By itself this resolution is already too low to faithfully capture the surface detail of our surface samples, which as we will show later in Section 5.3.5 have root-mean-square roughness, i.e., height variance, ranging from 0.28 μm to 0.40 μm.

More damning is the fact that this theoretical resolution is derived from the maximal optical resolving power under optimal assumptions, e.g., two distinct diffuse surface point located spatially apart, with no other noise or background light transmission. Such assumption is violated in practice: Light refraction and transmission by the polymer film contributes to overall reflected light; If the distribution of flakes is sparse enough to allow incident light to penetrate beyond a superficial layer, multiple scattering among the flakes will contribute to the overall light reflection; Finally geometric imperfections on the flake surface will contribute to the light reflection diffractively. Hence reflectance prediction based on CLSM scans alone seems unlikely to yield accurate result.

With leafing flakes, the cured paint is essentially covered on the surface by a layer of aluminum flakes, permitting very little light to penetrate beyond the surface. We therefore hypothesize that metallic paint prepared using leafing flakes should reflect
light in the same way roughened metal does, that is, with its reflectance completely
determined by surface topography.

This hypothesis is by no means a given or obvious fact. Factors such as the
thin layer of residual polymer between the metal flakes and air can potentially
affect reflectance as well. Our experimental prediction (described in Chapter 7 and
Chapter 8) thus helped validate this hypothesis, demonstrating the suitability of
metallic coating for future studies of this kind.

Because reliable topography measurements seemed more achievable for paint
made from leafing flakes than nonleafing flakes, we settled on using leafing flakes to
produce our sample surfaces. Schlenk, a commercial metallic effect pigment vendor,
kindly provided us with samples of various metallic flakes in their inventory. The
only leafing flakes available from Schlenk happened to be of the cornflake morphology
type, so we simply used those. Cornflake-shaped pigments are also the roughest and
most irregular of the three, and represent the most challenging type to characterize
via profilometry.

5.1.2 Polyurethane Binder

We used a Nason branded automotive clear coat formulation as the paint binder.
Known as a “2K” system, the formulation consists of two components: the clear and
the activator. Both are liquid solutions, packaged separately. After the clear is mixed
with the activator under a recommended proportion (4:1 for the product used), the
resulting solution will gradually cure into hardened semi-transparent polymer film.

5.1.3 Paint Production and Application

The exact procedure for producing and applying the metallic paint is as follows:

1. Mix 4ml clear and 1ml activator to create the binder solution.
2. Add a few mg leafing cornflake-shaped aluminum flakes to the solution. Use a vortex mixer to disperse the flakes within the solution.

3. Transfer the solution to the fluid cup of a pen-sized airbrush. Airbrush a glass slide for one or two passes.

4. Let the coated slide cure in the fume hood for an hour, and transfer it to an open area to dry overnight.

This procedure has low cost and quick turn-round. We were able to create in one lab session a large batch of samples corresponding to different flake types and concentration.

The coated surfaces exhibited high spatial homogeneity, since the metallic pigments are free to move around a little and reorient themselves during paint curing, smoothing out any initial non-uniformity during airbrushing. This property is useful to studies of our kind since surfaces with high spatial variance require more profile and optical measurement to meaningfully characterize.

5.1.4 Results

While we created many surface samples using various flakes types and concentration, we made extensive profilometry measurements only for a number of them, due to the time and effort required (detailed in Section 5.3). Throughout this thesis, we make reference to the following three samples:

**10um sample** 0.3g Schlenk EM 130 flakes (D50 10um), 5ml binder solution, airbrushed in two passes.

**17um sample** 0.3g Schlenk EM 90 flakes (D50 17um), 5ml binder solution, airbrushed in one pass.
Mixture sample mixture of 0.15g Schlenk EM 130 flakes and 0.15g Schlenk EM 90 flakes, 5ml binder solution, airbrushed in one pass.

Out of the three, the 10um sample is the glossiest. The 17um sample is duller. The mixture sample looks similar to the 17 sample, but is the dullest. This is interesting considering that the metallic pigments used to create the mixture sample is a blend of the ones used to create the other two. Fig. 5.3 shows photos of the three samples.

![Figure 5.3: Photographs of the 10um, 17um, and mixture surface samples. The photographs are not calibrated. See Fig. 5.6 for their reflectance measurements. In particular, the peak reflectance value of the 10um sample is about 30% higher than that of the 17um sample.](image)

### 5.2 Optical Measurement of Surface Reflectance

We built a simple automated gonioreflectometer to measure surface reflectance within the incident plane. See Fig. 5.4 for a schematic diagram of the optical geometry and a photograph of the physical setup. A rigidly mounted digital single-lens reflex (DSLR) camera (Nikon D7000) using a telephoto lens (Nikon 105mm f/2.8D AF Micro Nikkor) served as the light sensor; A fiber-coupled LED module (Thorlabs MCWHF1) with its output collimated by a reflective collimator (Thorlabs RC04FC-P01) served as the light source; An iris diaphragm was attached to the collimator to control the diameter of the light beam; The light source was attached to a motorized continuous
rotation stage (Thorlabs CR1-Z7), driven by a motor control module (Thorlabs TDC001); Finally a horizontal platform with adjustable height housed the surface sample. Both the DSLR camera and the motor control module were connected to a computer, allowing the measuring process to be automated.

![Optical geometry](image1)

![Physical setup](image2)

**Figure 5.4:** Our gonioreflectometer setup.

### 5.2.1 Details of Operation

We calibrated the setup such that the rotating incident directions and the optical path of the camera stayed in the same incident plane, perpendicular to the horizontal plane. The viewing angle $\theta_v$ was measured to be 24.6°, with 0° corresponding to the normal of the horizontal plane. The incident angle $\theta_l$ steps from $-90^\circ$ to $+90^\circ$ in variable increments to allow denser sampling around the specular reflection direction $\theta_l = -\theta_v$. At each step, raw photos are taken at different exposure levels, demosaiced to pick out the green channel, filtered for saturated pixels, summed up, and divided by the (per-pixel) exposure time, resulting finally in a linear intensity map.

The value in the intensity map should be proportional to the radiant flux transmitted to the camera aperture from the surface region defined by the corresponding pixel and weighted by the spectral response of the green color filter of the camera. Jiang et al. [2013] measured the spectral response curves of a large selection of
commercially available digital cameras. Most Nikon DSLR cameras seem to have peak green spectral response at about 0.5\,\mu m. \(^1\) The spectral intensity curve of our light source module [Thorlabs, Inc., 2013] also has a local maximum around that wavelength. Hence, as an approximation, we assume the measured radiant flux to result solely from reflected light of the wavelength \( \lambda = 0.5\,\mu m \). We also use this wavelength for our numerical simulation.

Fig. 5.5 shows two examples of the processed intensity images, for two different incident angles. The incident light beam had a diameter of a few millimeters. Depending on the incident angle, the shape of the lit area in the captured intensity map ranges from circular to elliptical. To average out speckles and surface inhomogeneity, we specify a small square region-of-interest (ROI) of fixed size, located at the center of the beam highlight. This ROI is the same for all intensity maps, irrespective of the incident light angles. For each incident direction \( \hat{l} \), the intensity values within the ROI is summed to arrive at a single value, \( \Phi(\hat{l}) \), corresponding to the radiant flux transmitted to the camera aperture from the surface region defined by the image-space ROI.

5.2.2 Relating Measured Flux to Reflected Wave

Let \( A_v \) be the surface corresponding to part of the camera lens not blocked by the aperture. Then the measured radiant flux \( \Phi(\hat{l}) \) should be proportional to the intensity of the reflected wave field \( U \) integrated over \( A_v \),

\[
\Phi(\hat{l}) \propto \iint_{A_v} |U(\mathbf{v})|^2 d\mathbf{v},
\]

\(^1\)Strictly speaking, the measurements provided in Jiang et al. [2013] do not contain that of D7000, which we use. However, the selection of Nikon DSLR cameras measured seemed to have similar green spectral responses, so we assume the same for D7000.
where \( |U|^2 \) is given by Eq. 4.1. For our gonioreflectometer, the camera is about 1m away from the surface sample, the lens is 105mm in diameter, and images are captured using a high f/stop number (corresponding to a smaller aperture). Hence the solid angle extended by \( A_v \) is small enough that we can assume the terms \( \|v\| \) and \( \hat{1} \cdot \hat{v} \) from Eq. 4.1 to be constant for \( v \in A_v \). The light source is also at a constant distance from the surface sample, so the intensity of the incident wave, \( |U_1|^2 \) in Eq. 4.1, stays constant for different \( \hat{1} \). Thus we arrive at Eq. 4.6, which relates the optically measured \( \Phi(\hat{1}) \) to the MPSD derived from the profilometer measurements.

5.2.3 Results

We used our gonioreflectometer setup to measure the reflectance of our surface samples within the plane of incidence and with a fixed reflection angle \( \theta_r = 24.6\degree \). To account for azimuthal variation, we measured at four azimuthal directions, 0\degree, 90\degree, 180\degree, and 270\degree. Since we rotated the samples by hand, each measurement corresponds to a slightly different spatial location on the sample surfaces. Fig. 5.6 shows the measured reflectance. As the samples exhibit high spatial and azimuthal uniformity, we will use the mean reflectance, averaged over the different azimuthal
directions and locations, as the ground truth to compare the prediction against. Note the small difference between the 17um and the mixture sample. As we show later, the methods proposed in this thesis is able to predict this tiny difference quantitatively and accurately.

![Graphs showing optically measured surface reflectance](image)

**Figure 5.6:** Optically measured surface reflectance in the plane of incidence, with the reflection angle $\theta_v = 24.6^\circ$. Measurements were made for four differential azimuthal angles, denoted by “N”, “E”, “S”, and “W” in the labels. For (d) the shaded envelopes are spanned by the minimum and the maximum reflectance values.

### 5.3 Profile Measurement via Atomic Force Microscopy

Atomic force microscopy (AFM) can measure the two-dimensional topography of a surface sample with nanoscale resolution, and in theory is ideal for acquiring the data needed to evaluate the Kirchhoff integral (Eq. 3.28) or other reflectance models. Indeed, several existing studies on reflectance prediction [Zhu, 2004; Schröder et al., 2011; Dhillon et al., 2014] treated AFM as a straightforward black-box procedure, without devoting much texts towards its nuances. Inaccuracies in the returned data
were treated as inherent and unavoidable. Schröder et al. [2011] for example stated that “[b]oth AFM and WLI measurement results are influenced by measurement uncertainties that are seldom given explicitly but can be assumed to be approximately 10-30%.”

Experimentally, we found that for our surface samples, which are much rougher than those traditionally investigated via AFM, performing scans according to the standard AFM procedure yielded measurements with the following defects:

- There exist random, uncorrectable misalignments between the one-dimensional scanlines that comprise the two-dimensional output of the AFM.

- The maximal scanning window allowed by the AFM hardware is too small to fully capture the roughness characteristics of our test surfaces.

Such defects would naturally lead to inaccurate reflectance prediction. To understand the source of these AFM-specific difficulties and potentially overcome them, we must examine how an AFM system operates.

### 5.3.1 Basics of AFM Operation

A high-level schematic description of some user-accessible components of the AFM system is as follows. See Fig. 5.7. The base contains a sample holder. The head unit with length-adjustable supporting legs is mounted on top of the base. A cantilever-probe assembly is installed at the bottom of the head unit. This assembly is subject to wear and tear during scanning, and needs to be replaced once worn. It can only be replaced when the head unit is unmounted from the base.

During a typical scanning session, the head unit is initially unmounted. The cantilever-probe assembly is first installed on the head unit. Next, with the surface sample placed into the sample holder, the head unit is mounted on the base, and manually lowered towards the sample via its adjustable supporting legs. Once the
Figure 5.7: AFM setup.

cantilever-probe on the head unit is close enough to the sample, an automated engagement process can be initiated to gently lower the scanning tip further until it “touches” the sample. Scanning can then ensue. Once done scanning, the probe can be disengaged, which raises it to a safe distance from the sample. When the probe is disengaged, an XY translation stage built into the sample holder can be used to move the sample laterally, allowing a new surface region to be scanned without the user dismounting the head unit in order to adjust the sample by hand.

During scanning, depending on the mode of AFM operation, the tip of the probe is either tapping the surface at high frequency or constantly touching the surface. Hence the probe will be blunted after continuous operation. The probe can also be damaged if, while scanning very rough surfaces, the control loop fails to pull it back away from the surface to avoid a sudden peak, or if the probe collides with some debris resting on the surface. Once the tip becomes blunt or damaged, the cantilever-probe assembly will need to be replaced, necessitating a remount of the head unit.

The vertical and lateral movements of the AFM tip are controlled by piezoelectric actuation. The time required to acquire each scanline is mainly limited by how fast the probe tip can safely trace across the surface. The lateral movement speed is a
user controller parameter, and must be set to give the feedback control loop enough
time to raise or lower the probe in response to surface concavities. Consequently,
the rougher the surface is, the slower the tip has to travel, and the longer it takes to
measure a single scanline. Likewise, the larger the scanning window is (in terms of
physical dimension, not sampling resolution), the longer it takes to acquire a single
scanline.

5.3.2 Inter-Scanline Drift

The scanning probe undergoes constant drifting in the vertical (Z) direction as
it traces across the sample surface. This drift is due to a number of different
factors, such as thermal expansion of the mechanical actuating components, ambient
vibration, and temperature variation [Meyer et al., 2014].

The variance of the accumulated drift thus increases with time. For us, the drift
within a single scanline is less of a concern because surface reflectance of randomly
rough surfaces depends more on the correlation between points closer to each other,
where the effect of drifting is not prominent, than those that are far apart. In fact,
later we show in Section 8.4 that for our test surfaces, the joint height distribution
of pairs of points that are more than 6\(\mu\)m apart do not contribute much to the
reflectance distribution at all.

On the other hand, the delay between the time when successive scanlines are
measured implies that there is significant random drift between them, even though
the lateral distance between consecutive scanlines is minuscule. It follows that
inter-scanline statistics computed from such topography measurements, such as the
two-dimensional MPSD (Eq. 4.1) will be heavily corrupted.

Traditionally, the issue of inter-scanline drift in AFM is addressed via line-
flattening: A best-fitting low-order polynomial is computed for each individual
scanline, and subtracted off.
In addition, if some region of the surface is known to be roughly planar beforehand, as is often the case with surfaces precision-fabricated via lithography or biological samples, only that region can be used to compute the best-fitting polynomial. This trick is known as masking, and when applicable, produces very good result.

Clearly line-flattening will not produce meaningful correction for a randomly rough surface with lateral roughness scale larger than the dimension of the AFM scanning window.

Our AFM system has a maximal scanning window dimension of 80um by 80um, and our test surfaces exhibit lateral roughness scale larger than that (shown later in Section 5.3.5). Hence line-flattening will mischaracterize inter-scanline statistics. The surface is also randomly rough so there is no regions usable for masking.

Given there is no reliable way to remove the inter-scanline drift from the two-dimensional AFM scans of our randomly rough surface samples, we are forced to treat each scan as a collection of independent and randomly displaced one-dimensional profiles as outlined in Section 4.1. We must then solve an inverse linear problem to recover the two-dimensional MPSD from these profiles.

5.3.3 Head Tilt

Whenever the head unit is remounted, the reference plane for height readout, i.e., the surface for which the height readout from AFM is 0, forms a new unknown tilt angle with the macroscale XY plane that the surface sample is assumed to lie in. This tilting is due mostly to manual adjustment of the supporting legs of the head unit during mounting.

Traditionally, the tilt angle is corrected via surface flattening: A best-fitting plane or low-order polynomial surface is computed and subtracted off the entire two-dimensional height field. Masking can be used if some regions of the surface are known to be planar beforehand.
Similar to the case of line-flattening, surface-flattening will not produce meaningful correction for rough surfaces with lateral roughness scale exceeding the scanning window size.

One obvious strategy for handling such surfaces is to scan at multiple lateral locations without a head unit remount, so tilt keeps constant for all these scans. Then a single tilt can be estimated to correct all the scans. We call this no-remount multipoint measuring.

**New Profile Measurement Procedure for Rough Surfaces**

Based on the above analysis, we propose a new experimental procedure for reliably measuring one-dimensional profiles of surfaces with much larger roughness scale than typically examined via AFM.

The key trick is to decouple the number of scanlines (resolution of the scan in Y) from the resolution of each scanline (resolution of the scan in X) during scanning. This significantly reduces the amount of time required for each scan, thereby making no-remount multipoint measuring feasible, as we explain now.

Our AFM hardware has a maximum scanning window size of 80μm by 80μm, and a maximal sampling resolution of 4096 by 4096. Covering 80μm with 1024 samples corresponds to an inter-sample distance of 0.078μm. This is anywhere between 1/5 to 1/9 of the wavelength of visible light, and serves as a reasonable lower limit on the scanning resolution that still allows a good-fidelity representation of the wave field on the surface. Recall that the maximal lateral traverse speed of the AFM probe tip is limited by the roughness of the sample. For our surface samples, performing a 1024 by 1024 scan over a 80μm by 80μm window requires at least three hours for the 10μm sample, which is the smoothest, and even longer for the rougher 17μm and mixture samples.

In comparison, when the number of scanlines per scan is reduced to 64 (the
lowest allowed by the AFM software) with an inter-scanline distance of 1.25\,um, each scan only takes 10 to 12 minutes. Moreover, each scanline can be measured at the full sampling resolution of 4096, since scanning time is mostly determined by the physical distance traveled by the probe.

Reducing the time needed for an individual AFM scan from 3 hours to 10 minutes is what allows no-remount multipoint measuring to be feasible. The AFM probe tip undergoes constant wear-and-tear during scanning since it physically taps the surface. For our surface samples, the probe would wear out after two or three 1024 by 1024 scans, necessitating a remount of the head unit to replace the cantilever-probe assembly. The probe can also be damaged due to collision with the surface when tracing steep profiles, necessitating a remount, and the probability of a scanning session being accident-free decreases exponentially as the length of the scanning session increases. When so few scans can be made after each remount, correcting head tilt from them will always underestimate surface roughness.

In contrast, with our new strategy of performing 4096 by 64 (64 scanlines each sampled at 4096 points) scans, we can reliably collect a batch of around 16 to 25 scans, corresponding to lateral positions located on a 4 by 4 or 5 by 5 equispaced grid with 1\,mm inter-grid spacing, in a single 3 to 4 hours session without a remount. At each grid point, we engage the AFM tip, perform a scan covering an area of 80\,um×80\,um, disengage the AFM tip, and translate the sample using the built-in translation stage so that the area around the next grid point can be scanned. The head unit is kept still during the entire process. Finally, all the scanlines contained in a batch are corrected using a single best-fitting slope.

In case of a probe tip collision early on, e.g., before 10 scans have been acquired, at most two hours of work are wasted. We can also afford to moderately decrease the probe traverse speed to further reduce both the likelihood of a collision accident and the wear rate of the probe.
From the above discussions, we can also conclude that, for surfaces with large lateral roughness scales, even if inter-scanline drift can somehow be completely eliminated from the measurements, performing high-resolution two-dimensional scans would still be incompatible with no-remount multipoint measuring, and the latter is required to avoid systematic underestimation of surface roughness.

5.3.4 Comparison with White Light Interferometry

A few existing studies [McKnight et al., 2001; Dong et al., 2015] computed surface reflectance from two-dimensional topography scans acquired via white light interferometry (WLI). Compared to AFM, WLI has the following advantages:

- WLI has a higher data-acquisition speed, independent of the physical dimension of the scanning window.
- WLI is easier to operate as it lacks consumable parts that need to be replaced after periods of operation.
- The two-dimensional topography returned by WLI does not have inter-scanline misalignment.

WLI does have its disadvantages. For example, its lateral resolution, which for many setups is claimed to be around 0.1um, is significantly worse than what is achievable by AFM. Nonetheless, a lateral resolution of 0.1um would seem borderline sufficient for performing physical-optics-based computation for visible light. What really makes WLI measurements unappealing to our goal of reflectance prediction is that WLI works off the interference pattern created by a beam reflected by the surface and a reference beam. The reflected beam is already affected diffractively by wavelength-scale surface details, but the demodulation algorithm employed by WLI can only assume that the reflected beam originates from a single surface “point”.

93
The algorithm has no way of recovering the exact wavelength-scale geometry around that point.

An exact technical analysis of this situation is beyond the scope of this thesis, and different vendors likely tune their algorithms differently. Nonetheless we refer to [Sosale, 2007, Fig. 2.11] for a visual demonstration of the difficulties associated with WLI. There profile scans of known gratings measured using both WLI and AFM were plotted. Denote by $w$ the traverse spacing and $h$ the step height of the gratings. The plots can be summarized as follows:

- With $w = 15\text{um}$, $h = 550\text{nm}$, both WLI and AFM characterized the profile reasonably well.

- With $w = 1.5\text{um}$, $h = 100\text{nm}$, the WLI output appeared as if the ground truth profile, which is a square wave, was first smoothed by a kernel spanning at least $0.5\text{um}$, and then scaled in the vertical direction by a factor of 2. The AFM output stayed reasonably accurate.

- With $w = 1.5\text{um}$, $h = 18\text{nm}$, the WLI output appeared as an oscillating polyline that looked very different from the ground truth, while the AFM output again stayed accurate.

We therefore hypothesize that when wavelength-scale surface details contribute nontrivially to the overall reflectance, WLI will fail to characterize surface topography with sufficient fidelity to allow accurate reflectance prediction. Indeed the reflectance curves computed by Dong et al. from two-dimensional WLI scans under the Kirchhoff approximation is consistently less diffuse than the optical ground truth, and one possible explanation is that its input heightfield is systematically smoother than the actual surface.
5.3.5 Results

Using the procedure described in this section, we acquired 2048 one-dimensional height profiles for the 10um sample, 3328 for the 17um sample, and 896 for the mixture sample. On average every 1000 profiles took three hours to acquire. The measurements were along two azimuthal directions, 0° and 90°. Fig. 5.8 shows typical AFM typography scans. We do not “flatten” the raw AFM measurements using conventional AFM processing algorithms. Note the visible discontinuities in the Y direction for the 10um and mixture samples. Such discontinuities across scanlines are common in raw AFM readout. In fact, varying amounts of drift exist across any two successive scanlines.

Fig. 5.9 plots the (unnormalized) autocorrelations of the three surface samples along with best fit Gaussians. The autocorrelations are computed after the AFM head tilt has been corrected. They are clearly non-Gaussian, and highly nonzero at the maximum scanning window size, 80um, implying that the lateral roughness scale of our surfaces exceed that. The root-mean-squared (RMS) roughness of the three samples, which is computed from the 80um window and equals the square root of the autocorrelation value at 0, are 0.28um, 0.43um, and 0.40um respectively.

5.4 Summary

In this chapter, we described the physical experimental procedures undertaken in this study: We first manufactured surface samples by airbrushing metallic paint made from off-the-shelf commercially-available leafing aluminum flakes and automotive clear coat formulation; We next built a simple gonioreflectometer to optically measure

\[2\text{Note that values of the autocorrelations closer to 80um are less reliable due to aliasing. Nonetheless, with at least 800 scanlines each sampled at 4096 points available for each surface, the autocorrelation value at, say 75um, would be derived from over 200 thousand pairs of surface points.}\]
Figure 5.8: Typical AFM measurements for the 10um, 17um, and mixture samples. Each scan is 4096 by 64 in resolution, and covers an area of 80um by 80um. Heightfields on the left column are rendered using the correct aspect ratio between height and lateral dimensions. Heightfields on the right column are rendered with ten-fold Z-exaggeration to make topographical features more visually-pronounced.
Figure 5.9: (Unnormalized) height autocorrelation of the height profiles and best fit Gaussians. The computed RMS roughness of the three samples are 0.28μm, 0.43μm, and 0.40μm respectively.

the reflectance of our surface samples; Finally we obtained one-dimensional profile measurements of our surface samples using a new AFM-based procedure specifically designed to handle surfaces rougher than typically investigated via AFM.

In the next chapters, we will present algorithms for reflectance prediction from one-dimensional surface profiles and evaluate these algorithms using the experimental measurements collected here.
Chapter 6

Shaped Microfacet Model

Chapter 3 established that reflectance prediction from one-dimensional profile measurements reduces to estimation of the MPSD $P$ from its projection $\text{prj} P$. Solving this ill-posed linear inverse problem requires additional priors and constraints to discard unlikely or undesirable candidates that satisfy $\text{prj} P$ by, for example, over-fitting to its noises.

In this chapter, we present the \textit{shaped microfacet model}, to introduce physically-meaningful priors into the linear inverse problem. This model also happens to be a natural extension to the classic microfacet model [Torrance and Sparrow, 1967] under the Kirchhoff approximation, and can conceivably be of use to other applications.

In Section 6.1 we first describe how a surface can be modeled as a collection of planar microfacets, each parameterized by its location, shape, and orientation. Then in Section 6.2 we derive a remarkably simple expression for the MPSD of such a surface, namely, as the \textit{expected} MPSD of its constituting microfacets. The expectation is taken over the joint distribution on facet shape and orientation. In Section 6.3, we model the expected MPSD (of microfacets) nonparametrically, expressing it as the linear combination of elements from a suitably constructed dictionary. Finally in Section 6.4 we give a probabilistic interpretation to the coefficients that appear...
in the linear combination, and analyze how physical characteristics of the actual surface translate to properties of the dictionary elements.

### 6.1 Surface Formation by Shaped Microfacets

The surface height field \( z(x), x \in \mathbb{R}^2 \) is seen as a collection of \textit{shaped} and \textit{oriented planar} microfacets, or just “facets” for short. Each facet will reflect light diffractively (due to its shape, not roughness) according to the Kirchhoff approximation.

A facet \( f \) is specified by the parameters \( c_f, z_f, \Omega_f, \theta_f, \) and \( \alpha_f \) as follows: Let \( c_f \in \mathbb{R}^2 \) and \( z_f \in \mathbb{R} \) be the lateral (XY) and vertical (Z) locations of the \textit{centroid} of the facet. The \textit{shape function} \( \Omega_f : \mathbb{R}^2 \to \{0, 1\} \) is an indicator function characterizing the lateral domain of the facet, that is, points on \( f \) are given by \( \{(x, z(x)) : \Omega_f(x - c_f) = 1\} \). The \textit{slant angle} \( \theta_f \in [0, \pi/2) \) is the angle between the facet normal and the positive Z axis. It is also the positive angle that the planar facet makes with the XY plane. Let \( w \) be the projection of a vector along the positive Z direction onto the facet. Then the projection of \( w \) onto the XY plane forms an angle \( \alpha_f \in [-\pi/2, \pi/2) \) with the X axis. We refer to \( \alpha_f \) as the \textit{azimuthal angle} of the facet. Fig. 6.1 illustrates \( \theta_f \) and \( \alpha_f \).

![Figure 6.1: The slant angle, \( \theta_f \), and the azimuthal angle, \( \alpha_f \), of a facet \( f \). For illustration, the facet is assumed to be centered at the origin. \( l \) is the intersection between the facet and the XY plane.](image)

99
Points on the planar facet $f$ can also be conveniently described through the plane equation as

$$z(x) = s_f \cdot (x - c_f) + z_f \text{ for } x \text{ such that } \Omega_f(x) = 1,$$  \hspace{1cm} (6.1)

where $s_f \in \mathbb{R}^2$ consists of the partials of $z(x)$ restricted to the domain of the facet, so we refer to $s_f$ as the slope vector of $f$. Also note that $(-s_f, 1) \in \mathbb{R}^3$ lies in the direction of the facet normal. $s_f$ is related to the slant angle $\theta_f$ and the azimuthal angles $\alpha_f$ via a function $s$,

$$s_f = s(\theta_f, \alpha_f) = \tan \theta_f [\cos \alpha_f, \sin \alpha_f]^T.$$ \hspace{1cm} (6.2)

In other words, $\theta_f$ and $\alpha_f$ determine the magnitude and the direction of $s_f$ respectively. While the slope vector is notationally useful, the angular parameters are useful for discussing surface anisotropy.

Joining together the piecewise definition for each facet (Eq. 6.1), we can write surface height for the entire domain as

$$z(x) = \sum_f [z_f + s_f \cdot (x - c_f)]\Omega_f(x).$$ \hspace{1cm} (6.3)

Note that our definition does not require the surface to be continuous across facets boundaries.

### 6.2 Expected Modulated Power Spectral Density

Instead of considering a single surface and calculating its reflectance, we opt to consider a fixed probability space $\mathcal{S}$ of surfaces. The unknown surface whose reflectance we are predicting is modeled as drawn from $\mathcal{S}$. Surfaces drawn from $\mathcal{S}$ are assumed to consist of $M$ shaped facets whose facet parameters \{\(X_f = (c_f, z_f, \Omega_f, \theta_f, \alpha_f) : f = 1, \ldots, M\)\} are independent and identically distributed (i.i.d.), i.e., $X_f \sim p_X$ i.i.d.
for some distribution $p_X$.

We now seek to express the expected MPSD of surfaces in $S$. Let $z \in S$ be parameterized by $\{X_f = (c_f, z_f, \Omega_f, \theta_f, \alpha_f) : f = 1 \ldots M\}$. $z$ can be written as in Eq. 6.3, and its surface modulation function (Eq. 3.23) is

$$m(x) = \sum_{f=1}^{M} e^{ihz_f} e^{-ihs_f c_f} e^{ihs_f} x \Omega_f(x),$$

where the scalar $h$ is introduced to simplify notation,

$$h = -2kh_z = -4\pi/(\lambda h_z).$$

We next take the Fourier transform of the surface modulation function $m$, and use the Fourier shift theorem $\mathcal{F}\{e^{id \cdot x} f(x)\}(u) = \mathcal{F}\{f\}(u - d)$ to obtain the following,

$$\mathcal{F}\{m\}(u) = \sum_{f} e^{ihz_f} e^{-ihs_f c_f} \mathcal{F}\{\Omega_f\}(u - h s_f).$$

The expected MPSD $\mathbb{E}[P]$ is then just the expected value of the squared magnitude of $\mathcal{F}\{m\}$,

$$\mathbb{E}[P(u)] = \mathbb{E}[\mathcal{F}\{m\}(u)\mathcal{F}\{m\}(u)] = \mathbb{E}\left[ \sum_{f} |\mathcal{F}\{\Omega_f\}(u - h s_f)|^2 \right] + K,$$

where $K$ contains the cross terms,

$$K = \sum_{j \neq k} \mathbb{E}[e^{ih(z_j - z_k - s_j c_j + s_k c_k)} \mathcal{F}\{\Omega_j\}(u - h s_j) \mathcal{F}\{\Omega_k\}(u - h s_k)]. \quad (6.4)$$

We will assume that $z_f$ is independently distributed from the other facet parameters, i.e., knowing a facet’s vertical location tells nothing about its lateral position, shape, or orientation, and vice versa. Then the expectation terms in $K$ can be factored as

$$K = \sum_{j \neq k} \mathbb{E}[e^{ih(z_j - z_k)}] \mathbb{E}[w(u, h, s_j, c_j, \Omega_j, s_k, c_k, \Omega_k)],$$

where the function $w$ does not depend on $\{z_f\}$. 

101
Here we make another assumption: The surface is sufficiently rough such that $(h(z_j - z_k) \mod 2\pi)$ approaches a uniform distribution over $[0, 2\pi)$, which implies that $\mathbb{E}[e^{ih(z_j - z_k)}] = 0$ and $K = 0$. And since the facet parameters $\{(c_f, z_f, \Omega_f, \theta_f, \alpha_f)\}$ are i.i.d., the expected MPSD is just

$$\mathbb{E}[P(u)] = M\mathbb{E}_{\Omega, \theta, \alpha} \left[ |\mathcal{F}\{\Omega\}(u - hs(\theta, \alpha))|^2 \right],$$

where the expectation is taken over the joint distribution on facet shape and orientation, $(\Omega, \theta, \alpha)$.

Eq. 6.5 is the main result of this section. The expression inside the expectation is simply the PSD of the shape function $\Omega$ shifted by $hs$. It is convenient that the expected MPSD depends on the distribution on $(\Omega, \theta, \alpha)$ only, and not on the positional distribution of the facets. We need to keep in mind however that this result depends on the assumptions we made, recapped as follows:

- The facets are i.i.d.
- Facet height, $z_f$, is independently distributed from other facet parameters.
- The surface is sufficiently rough such that $(h(z_j - z_k) \mod 2\pi)$ approaches a uniform distribution over $[0, 2\pi)$.

When a surface does not admit these assumptions, e.g., if it is periodically grated or very smooth, we cannot discard the cross terms (Eq. 6.4) as easily, and must analyze and model them suitably. We also note that these assumptions are not necessary conditions for the cross terms to vanish. Some different set of assumptions can conceivably lead to the same conclusion.

The traditional microfacet model, parametrized only by the facet orientation (normal) distribution, can now be seen as a specialization of our model that assumes a fixed facet shape so large in area that $|\mathcal{F}\{\Omega\}|^2$ is effectively a delta spike.

102
6.3 Nonparametric Modeling of the Expected MPSD

In this section, we model the expected MPSD nonparametrically as the linear combination of basis functions. The expected MPSD (Eq. 6.5) can be rewritten using conditional expectation as

\[
\mathbb{E}[P(u)] = M \mathbb{E}_{\theta, \alpha}[\mathbb{E}[|F\{\Omega\}|(u - hs(\theta, \alpha))^2 | \theta, \alpha]]
\]

\[
= M \int_{\theta} \int_{\alpha} W_{\theta, \alpha}(u - hs(\theta, \alpha)) p_{\theta, \alpha}(\theta, \alpha) \, d\alpha \, d\theta,
\]

where \( p_{\theta, \alpha} \) is the probability density function (pdf) of the marginal distribution of \((\theta, \alpha)\), and \( W_{\theta, \alpha} \) is the expected PSD of shape functions conditioned on the given \((\theta, \alpha)\),

\[
W_{\theta, \alpha} = \mathbb{E}[|F\{\Omega\}|^2 | \theta, \alpha].
\]

Note that \( W_{\theta, \alpha} \) is a plain nonnegative function defined on \( \mathbb{R}^2 \). Hence we can approximate it by the linear combination of a suitably chosen set of basis functions, \( \{B_j : \mathbb{R}^2 \to \mathbb{R}\} \),

\[
W_{\theta, \alpha} = \sum_j b_j(\theta, \alpha) B_j,
\]

where \( b_j(\theta, \alpha) \) is the linear weight associated with \( B_j \) and is dependent on \( \theta \) and \( \alpha \). For example, \( \{B_j\} \) can simply be a collection of two-dimensional Gaussians of different means and variances. We will refer to \( \{B_j\} \) as the shape basis.

Combining Eqs. 6.6 to 6.8, we can write the expected MPSD in terms of the basis elements \( \{B_j\} \),

\[
\mathbb{E}[P(u)] = M \sum_j \int_{\theta} \int_{\alpha} b_j(\theta, \alpha) p_{\theta, \alpha}(\theta, \alpha) B_j(u - hs(\theta, \alpha)) \, d\alpha \, d\theta,
\]

Note that the product \( b_j(\theta, \alpha) p_{\theta, \alpha}(\theta, \alpha) \) is a plain scalar function of \( \theta \) and \( \alpha \). We can approximate it as follows: Define a set of basis functions \( \{\Theta_k\} \) for approximating functions of the slant angle \( \theta \); Define another set of basis functions \( \{A_l\} \) for approximating functions of the azimuthal angle \( \alpha \); Finally, form the tensor product of these
two functional spaces to approximate \((b_j p_{\theta,\alpha})\),

\[
b_j(\theta, \alpha)p_{\theta,\alpha}(\theta, \alpha) = \sum_{k,l} d_{j,k,l} \Theta_k(\theta)A_l(\alpha),
\]

where \(d_{j,k,l}\) is the linear coefficient associated with the basis element \(\Theta_k \otimes A_l : (\theta, \alpha) \mapsto \Theta_k(\theta)A_l(\alpha)\). We will refer to \(\{\Theta_k\}\) and \(\{A_l\}\) as the slant basis and azimuthal basis respectively.

From the above, we can now write the expected MPSD \(\mathbb{E}[P]\) as a linear combination of elements from a suitably constructed dictionary, \(\{\Psi_{j,k,l}\}\),

\[
\mathbb{E}[P] = \sum_{j,k,l} c_{j,k,l} \Psi_{j,k,l},
\]

where \(c_{j,k,l} = Md_{j,k,l}\) and

\[
\Psi_{j,k,l}(u) = \int_{\theta} \int_{\alpha} B_j(u - hs(\theta, \alpha))\Theta_k(\theta)A_l(\alpha)d\alpha d\theta.
\]

As illustrated by Fig. 6.2, each dictionary element \(\Psi_{j,k,l}\) is the “sum” of many copies of \(B_j\), and each copy, indexed by \((\theta, \alpha)\), is translated by \(hs(\theta, \alpha)\) and weighted by \(\Theta_k(\theta)A_l(\alpha)\).

Figure 6.2: Each MPSD dictionary element \(\Psi_{j,k,l}\) is the “sum” of many copies of the shape basis \(B_j\). Each copy, indexed by \((\theta, \alpha)\), is translated by \(hs(\theta, \alpha)\) and weighted by \(\Theta_k(\theta)A_l(\alpha)\).
6.4 Discussions

In this section, we present more discussions on the interpretation and choices of basis functions introduced for modeling the expected MPSD.

6.4.1 Probabilistic Interpretation of the Basis Coefficients

Consider the following way of defining the shape basis: Assume the set of possible facet shapes to be finite, denoted by \( \{\Omega_j\} \). Define the shape basis simply to be \( \{B_j = |F\{\Omega_j}\}|^2 \} \). Then the expected PSD of shape functions conditioned on a given orientation \((\theta, \alpha)\), is by definition the follows,

\[
E[|F\{\Omega}\|^2 | \theta, \alpha] = \sum_j \Pr(\Omega_j | \theta, \alpha) B_j.
\] (6.13)

Compare with Eq. 6.8 and it becomes clear that \( b_j(\theta, \alpha) \), the linear weight associated with \( B_j \) for representing \( E[|F\{\Omega}\|^2 | \theta, \alpha] \), is just \( \Pr(\Omega_j | \theta, \alpha) \), that is, the probability that the facet shape is \( \Omega_j \) conditioned on the given orientation \((\theta, \alpha)\). It follows that \( b_j(\theta, \alpha) \) should be nonnegative and sum to 1 over \( j \).

In the above discussion, the discrete probability distribution on facet shapes can also be replaced by a mixture distribution. Each shape basis element should then be set to the expected PSD of facets within a mixture component, but the analysis otherwise stays the same.

Denote by \( \mathcal{P} \) the space of all possible expected PSD of shape functions conditioned on any orientation, i.e., the space of all possible \( E[|F\{\Omega}\|^2 | \theta, \alpha] \). Under the above definition of shape basis, \( \mathcal{P} \) corresponds to the convex hull of the shape basis. To properly model \( \mathcal{P} \) of real-world surfaces, it is conceivable that many shapes need to be included in \( \{\Omega_j\} \), resulting in an equally large \( \{B_j\} \).

Alternatively, \( \mathcal{P} \) can be modeled as the linear span of the shape basis, as we did in Section 6.3. Under such a choice, there is no one-to-one correspondence between
a shape basis element $B_j$ and a possible facet shape. $\{B_j\}$ is chosen, for example, to numerically best represent $\mathcal{P}$, and the basis coefficients can be any real numbers. It is conceivable that although $\mathcal{P}$ is the convex hull of the PSD of many facet shapes, $\mathcal{P}$ is also (approximately) contained in the linear span of a much smaller set of functions which form the shape basis. One can even use a numerical algorithm such as principle component analysis to compute a smaller linear basis from a large collection of shape PSDs. Of course a linear span will likely contain elements not found in $\mathcal{P}$ as well.

When $b_j(\theta, \alpha)$ does represent $\Pr(\Omega_j \mid \theta, \alpha)$, the term $(b_k(\theta, \alpha)p_{\theta, \alpha}(\theta, \alpha))$, which was approximated by the linear combinations $\sum_{k,l} d_{j,k,l} \Theta_k(\theta) A_l(\alpha)$ in Eq. 6.10, also takes on a probabilistic interpretation: Define the function $p_{\Omega, \theta, \alpha}$ as follows,

$$p_{\Omega, \theta, \alpha} : (j, \theta, \alpha) \mapsto b_j(\theta, \alpha)p_{\theta, \alpha}(\theta, \alpha).$$

The by definition $p_{\Omega, \theta, \alpha}$ is the pdf of the joint distribution on $(\Omega, \theta, \alpha)$, i.e., facet shape and orientation. It follows that $(b_j p_{\theta, \alpha})$ should be nonnegative and integrate to 1 over $(j, \theta, \alpha)$. In addition a similar dichotomy on the interpretation of the slant and azimuthal bases now exists: We can treat $\{\Theta_k\}$ and $\{A_l\}$ as the pdfs of the mixture components of a mixture distribution, in which case the coefficients $\{d_{j,k,l}\}$ in Eq. 6.10 take on probabilistic interpretation and constraints, i.e., nonnegativity and sum-to-unity; Or, we can treat them as plain functions whose linear span contains $p_{\Omega, \theta, \alpha}$. Of course, if $b_j(\theta, \alpha)$ does not admit a probabilistically interpretation, then the above dichotomy will not exist in the first place, and $\{d_{j,k,l}\}$ will just be plain linear coefficients.

### 6.4.2 Properties of a Shape Basis

We now discuss the choice of the basis $\{B_k\}$ for approximating $W_{\theta, \alpha} = E[|\mathcal{F}\{\Omega}\}|^2 \mid \theta, \alpha]$, the expected shape PSD conditioned on a given orientation (Eq. 6.7). Because $W_{\theta, \alpha}$ is the average of the PSD of many facet shapes, it is instructive to examine the
PSD of a few example facet shapes.

Fig. 6.3 shows the PSD of a disk and a triangle shape functions. The indicator function of a unit disk is the *cylinder function*, \( \text{cyl} \),

\[
\text{cyl}(x) = \begin{cases} 
1, & \text{if } \|x\| \leq 1, \\
0, & \text{otherwise}.
\end{cases}
\] (6.15)

Its Fourier transform is the *sombrero function*, \( \text{somb} \), scaled by \( \pi \),

\[
\mathcal{F}\{\text{cyl}\}(u) = \pi \text{somb}(u) \\
\text{somb}(u) = \frac{2J_1(\|u\|)}{\|u\|},
\] (6.16)

where \( J_1 \) is the *Bessel function* of the first kind. The PSD of a disk is hence proportional to \( |\text{somb}|^2 \). For the triangle shape, while it is possible to derive a closed form expression of the Fourier transform of a general polygon shape function [Lee and Mittra, 1983], we simply evaluated its PSD numerically. From the plots we can observe that in general the PSD of a shape will look like a two-dimensional “blob” centered at the origin, similar in shape to \( |\text{somb}|^2 \). Straight-line edges of the shape will exhibit as “ridges” across the origin in the power spectrum. For example, there are three ridges in the PSD of the triangle shape.

![Figure 6.3: PSD of a disk (left) and a triangle (right) shape function.](image)

Applying a linear transformation to the shape has the following effect on its
Fourier transform: Let $\Omega_2(x) = \Omega_1(Mx)$, then

$$\mathcal{F}\{\Omega_2\}(u) = |M|^{-1}\mathcal{F}\{\Omega_1\}(M^{-T}u).$$  \hspace{1cm} (6.17)

Hence we make the following observations:

- Rotation in spatial domain corresponds to rotation by the same amount in the Fourier domain. Given a polygon with many linear ridges in its PSD, averaging together the PSD of many rotated copies of it will blur out the ridges.

- Scaling in the spatial domain corresponds to inverse-scaling in the Fourier domain. Averaging together the PSD of many slightly dilated and contracted copies of a disk will blur out the circular ripples in its PSD.

- If, conditioned on a fixed orientation, the facets are distributed over a range of shape, size, and azimuthal orientation, then the conditional expected PSD should be smooth and devoid of circular ripples or linear ridges.

Based on the above, we conclude that a suitable shape basis for real-world rough surfaces should be chosen according to the following criteria:

- The shape basis should contain functions with different “widths,” since the facets will likely exhibit varying sizes. One simple example is a collection of blurred sombrero functions corresponding to the PSD of differently sized disks. Another example is a collection of two-dimensional Gaussians with different variances.

- The minimal “width” of basis functions, i.e., how narrow and spiky they can be, should be chosen according to the geometric properties of the surface. For example, very rough surfaces cannot not have large planar facets, so their constituting facets cannot have a spiky PSD.
• The aspect ratio of basis functions should also be set based on the geometric properties of the surface. For example, the planar facets constituting the surface of a piece of brushed metal will exhibit elongated shape, leading to shape PSD with large aspect ratios.

6.4.3 Azimuthal Basis for Modeling Anisotropy

Anisotropic reflectance can arise from the combination of two factors: (i) The (conditional) expected shape PSD, \( \mathbb{E}[|\mathcal{F}\{\Omega\}|^2 | \theta, \alpha] \) can be anistropic; (ii) The joint distribution on facet shape and orientation, \( p_{\Omega,\theta,\alpha} \), can be nonuniform with respect to the azimuthal angle \( \alpha \). The azimuthal basis, \( \{A_l\} \), allows the latter source of anisotropy to be expressed.

Under the probabilistic interpretation in Section 6.4.1, the term \( b_j(\theta, \alpha)p_{\theta,\alpha}(\theta, \alpha) \) (Eq. 6.11) corresponds to \( p_{\Omega,\theta,\alpha} \) (Eq. 6.14). The azimuthal basis, \( \{A_l\} \), determines how fast \( p_{\Omega,\theta,\alpha} \) can vary with respect to the azimuthal angle \( \alpha \). For a completely isotropic surface, \( p_{\Omega,\theta,\alpha} \) is likely to be constant over \( \alpha \), which suggests that \( \{A_l\} \) should consist of a single element: the uniform distribution. Alternatively, a surface can exhibit a low level of anisotropy, which suggests that its azimuthal basis functions should have wide support and slow decay, so that \( p_{\Omega,\theta,\alpha} \) changes slowly with respect to \( \alpha \).

6.5 Summary

In this chapter we described the shaped microfacet model, which treats a surface as an ensemble of shaped and oriented microfacets. Under suitable assumptions, the MPSD of such a surface, which determines its reflectance under the Kirchhoff approximation, is proportional to the expected MPSD of the microfacets over the joint distribution on their shape and orientation. We then modeled this expected
MPSD nonparametrically as the linear combination of elements from a dictionary, built from three sets of basis functions: a shape basis, a slant basis, and an azimuthal basis. We finally discussed how physical characteristics of the surfaces translate to properties of these basis functions.

In the next chapter, we will tackle the actual problem of reflectance prediction from one-dimensional profile measurements. To solve the ill-posed linear inverse problem of estimating the surface MPSD from its projections, we will presume the shaped microfacet model and choose its defining bases suitably.

We conclude this chapter by noting that the nonparametric approach was undertaken because a parametric model would often introduce too strong an assumption, e.g., Gaussianity, which would be violated by real-world physical surfaces such as our samples. Such violation would hinder accurate reflectance prediction. For other potential use of the shaped microfacet model including rendering, parametric modeling is an appealing alternative.
Chapter 7

Estimation of Model Parameters

Chapter 4 showed that the MPSD, $P$, determines surface reflectance, and an estimate on its projection, $p \approx \text{prj} P$, can be computed from the one-dimensional surface profile measurements. In this chapter, we describe an algorithm for estimating $P$ given $p$. To tackle this ill-posed linear inverse problem, we employ the shaped microfacet model proposed in Chapter 6. Under this model, the unknown surface is assumed to be drawn from a probabilistic space, consisting of surfaces formed from i.i.d. shaped and oriented microfacets. The expected MPSD of surfaces in this space, $\mathbb{E}[P]$, can be modeled nonparametrically as the linear combination of elements from a suitably constructed dictionary, $\{\Psi_{j,k,l}\}$. We thus have the following linear equation,

$$\text{prj} \mathbb{E}[P] = \sum_{j,k,l} c_{j,k,l} \text{prj} \Psi_{j,k,l} = p. \quad (7.1)$$

Solving for $\{c_{j,k,l}\}$ gives $\mathbb{E}[P]$ and hence the surface reflectance.

Recall from Chapter 6 that the MPSD dictionary elements $\{\Psi_{j,k,l}\}$ arise from three building blocks: the shape basis $\{B_j\}$, the slant basis $\{\Theta_k\}$, and the azimuthal basis $\{A_l\}$. In Section 7.1 we specify our choices for these bases. Then in Section 7.2 we detail our numerical algorithm for estimating $\{c_{j,k,l}\}$ via least-square optimization.
Finally in Section 7.3, we evaluate the proposed algorithm using experimental measurements collected in Chapter 5.

7.1 Bases Choice for Shaped Microfacet Model

Under the probabilistic interpretation (Section 6.4.1), our choice of the bases will correspond to modeling the joint distribution on facet shape and orientation as follows: The sets of possible shapes and slant angles, denoted by $\{\Omega_j\}$ and $\{\theta_k\}$, are assumed to be finite; The distribution on the azimuthal angle, conditioned on shape and slant angle, is assumed to be a mixture distribution with $\{A_l\}$ as the pdfs of its mixture components.

However we do not enforce nonegativity or sum-to-unity when we estimate the coefficients $\{c_{j,k,l}\}$ (described later in Section 7.2). As it turns out, enforcing nonnegativity (without even the sum-to-unity constraint) slightly decreases the agreement between the predicted and measured reflectance. This implies that the linear span of our simple shape basis (described later) is good enough for representing the expected shape PSD for our test surfaces, but its convex cone, i.e., linear span with nonnegative coefficients, is too restrictive. We have verified that even though some of the fitted coefficients $\{c_{j,k,l}\}$ can be negative, the resulting (conditional) expected shape PSD, $E[|\mathcal{F}\{\Omega\}|^2 | \alpha, \beta]$, is nonnegative as expected. Because the prediction turns out to be good as is, we do not pursue a more sophisticated shape basis for use with nonnegative coefficients.

7.1.1 Shape Basis: Scaled Sombreros

We let $\Omega_j$, $j = 1, \ldots, N_\Omega$ be the indicator function of a two-dimensional disk with radius $r_j$, i.e., $\Omega_j(u) = \text{cyl}(u/r_j)$ where cyl is the (unit-radius) cylinder function. The shape basis $\{B_j\}$ is defined to be the PSD of these disk shapes, normalized to
have the same $L^2$ norm for better numerical stability,

$$B_j(u) = |r_j \text{somb}(r_j u)|^2 = \frac{r_j^2 |J_1(r_j \|u\|)|^2}{\|u\|^2}, \quad (7.2)$$

where $\text{somb} = \mathcal{F}\{\text{cyl}\}$ is the sombrero function and $J_1$ is the Bessel function of the first kind. The shape basis thus consists of many dilated sombreros.

We let the radii $\{r_j\}$ be distributed within the interval $[\lambda, r_{\text{max}}]$ with $r_{\text{max}}$ as a user-tunable parameter. Hence the smallest facet size has a radius equal to the wavelength, which gives a very dilated PSD and hence diffuse reflection. It is not physically meaningful to set $r_{\text{max}}$ arbitrarily large since the underlying surface is rough. The inter-spacing among $\{r_j\}$ is such that the sequence $\{r_j^{-1}\}$ is logarithmically-spaced, which implies that the “widths” of $\{B_j\}$, loosely defined as the diameter of the circle formed by the first zero crossing of the dilated sombrero, are logarithmically-spaced within their range.

Recall from Chapter 5 that our surface samples are fabricated using metallic paint produced from leaking flakes. The top surface of the cured paint is essentially covered by a layer of flakes. One might wonder if there is a one-to-one correspondence between a metallic flake and a planar facet in our model. The answer is no: The flakes used were of the cornflake type, with jagged irregular boundary and rough topography. Flakes also stack on top of each other, resulting in occlusion. Thus we cannot hope to model each flake as a single planar facet, as should be apparent from the AFM scans, Fig. 5.8. The suitable $r_{\text{max}}$ will be much smaller than the average flake radius. One possible method for estimating $r_{\text{max}}$ is to scan for the longest straight line segment in the measured profiles.

For our test surfaces, the expected shape PSD of its constituting facets should be roughly isotropic, because without external factors affecting the deposition of flakes, a planar microfacet, which in general is a part of a larger flake and has irregular anisotropic shape, can exhibit any azimuthal orientation with equal probability.
Hence we are justified in using isotropic shape basis functions. For surfaces whose constituting microfacets have anisotropic shapes and, in addition, are oriented azimuthally with nonuniform probability, anisotropic shape basis functions should be used.

7.1.2 Slant Basis: Shifted Dirac Deltas

We let the possible slant angles \( \{ \theta_k : k = 1, \ldots, N_\theta \} \) be distributed within the interval \([0, \theta_{\text{max}}]\) with \( \theta_{\text{max}} \) corresponding to the steepest facet considered. \( \theta_{\text{max}} \) has little effect on the computed reflectance beyond a certain value, so we simply set \( \theta_{\text{max}} \) to \( \pi/4 \). The inter-spacing among \( \{ \theta_k \} \) is such that the lengths of the corresponding slope vectors, \( \| s(\theta_k, \cdot) \| = h \tan \theta_k \), form an equally-spaced sequence. The slant basis \( \{ \Theta_k \} \) is simply many shifted Dirac deltas,

\[ \Theta_k(\theta) = \delta(\theta - \theta_k). \]

This choice of the slant basis is mostly for computational convenience, as integration over \( \theta \) (Eq. 6.12) now reduces to a single evaluation. More sophisticated choices using smoother basis functions and more efficient sampling, e.g., more densely near 0, are clearly possible.

7.1.3 Azimuthal Basis: Piecewise-Linear “Hats”

We define \( \{ A_l : l = 1, \ldots, N_\alpha \} \) with \( N_\alpha \) as a user-tunable parameter that controls the support and rate of decay over the angular domain of the basis functions: When \( N_\alpha = 1 \), we set \( A_1(\alpha) \equiv 1 \), so \( p_{\Omega, \theta, \alpha} \) would be constant with respect to \( \alpha \). When \( N_\alpha > 1 \), we use piecewise-linear “hats” as the basis functions,

\[
A_l(\alpha) = \begin{cases} 
1 - \frac{N_\alpha |\alpha - \alpha_l|}{2\pi} & \text{if } \alpha \in [\alpha_l - \frac{2\pi}{N_\alpha}, \alpha_l + \frac{2\pi}{N_\alpha}], \\
0 & \text{else},
\end{cases}
\]
where $\alpha_l = 2\pi(l - 1)/N_\alpha$. More sophisticated choices such as the spherical harmonics or higher order B-spline bases are clearly possible.

Recall from Section 4.4 that if profile measurements along some angle $\beta$ is available, $\text{prj}_\beta P$ can be computed from these measurements to constrain $P$. In such a case, the azimuthal basis should include the pair of elements centered at $\beta$ and $\beta + \pi$, since they are the most effective at resolving the asymmetry present in $\text{prj}_\beta P$. In contrast, a basis function centered at $\beta + \pi/2$, being mirror symmetric over a line along the direction of the projection, always results in a completely symmetric image under $\text{prj}_\beta$, and hence, is completely ineffective at resolving any asymmetry present in $\text{prj}_\beta$.

### 7.1.4 Dictionary for the Expected MPSD

With the shape, slant, and azimuthal bases as specified above, the expected MPSD (Eq. 6.11) is given by

$$
\mathbb{E}[P] = \sum_{j=1}^{N_\Omega} \sum_{k=1}^{N_\theta} \sum_{l=1}^{N_\alpha} c_{j,k,l} \Psi_{j,k,l},
$$

where

$$
\Psi_{j,k,l}(u) = \int B_j(u - hs(\theta_k, \alpha)) A_l(\alpha) d\alpha.
$$

The construction of the dictionary $\{\Psi_{j,k,l}\}$ incorporates our prior assumptions on the unknown rough surface, including the presumed shaped microfacet model, an estimated upper bound on facet size ($r_{\text{max}}$), and the amount of surface anisotropy ($N_\alpha$).

### 7.2 Least-Square Optimization

This section details how we solve the main linear equation, Eq. 7.1 using least-square optimization.
7.2.1 Discretization

We discretize the relevant frequency-domain functions as follows: Recall from Section 5.3 that each height profile has a physical extent of \( W = 80 \mu m \) and is sampled at \( L = 4096 \) points. The DFT of such a discrete signal corresponds to approximating its continuous Fourier transform at angular frequencies \( \{ \omega_n = \frac{2\pi n}{W} : n = -L/2, \ldots, L/2 - 1 \} \). This is the grid we use to discretize \( E[P], p, \Psi_{j,k,l}, \text{prj}\Psi_{j,k,l} \) and other frequency-domain functions. Because MPSDs attenuates rapidly as frequency increases, we can even truncate this grid, for example to \( |n| < 256 \), when computing with no ill effect. We use numerical integration to evaluate \( \Psi_{j,k,l} \) and the projection operator \( \text{prj} \).

7.2.2 Solving at Multiple \( h_z \)

Notice the following ambiguity when fitting a joint shape and orientation distribution to the MPSD (and its projection): A small facet exhibits a “fat” MPSD; Meanwhile, a collection of suitably oriented larger facets can exhibit the same “fat” MPSD, even though they individually contribute a “thin” MPSD.

There is a subtle difference between the two configurations though. Refer to Eq. 6.12. Consider a single unslanted small facet. Its MPSD, say \( g \), is determined completely by the PSD of its shape function, which does not vary with the half angle. Consider now a collection of larger facets suitably oriented so their MPSDs together sum to \( g \). When the half angle, or more specifically \( h_z \) (the Z component of the half vector), varies, the MPSD of a larger facet will be translated according to the \( hs(\theta, \alpha) \) term in Eq. 6.12 because of its non-zero slant angle. Hence the overall MPSD of the collection of larger facets will change.

In other words, the ambiguity is resolved if we enforce the constraints on \( \{c_{j,k,l}\} \) (Eq. 7.1) simultaneously for multiple \( h_z \). In practice we pick 8 samples evenly
distributed from its range of $[0, 1]$. We do not find the exact choices for these values to matter much to the computed reflectance.

### 7.2.3 Smoothness Regularization

When the physical surface is rough, and the profile measurements are too sparse and incomplete to cover the full space of possible surface profiles, the estimated MPSD projection, $p$, can be noisy. With a noisy $p$, we found it necessary to incorporate smoothness regularization to make the computed MPSD well-behaved around the peak region. We define an energy functional $E_s$,

$$ E_s\{f\} = \iint_{\mathbb{R}^2} |D_{\hat{u}}^2 f(u)|^2 du, $$  

where $\hat{u} = u/\|u\|$ is the outward radial direction at $u$, and $D_{\hat{u}}^2 f(u)$ is the second-order (directional) derivative of $f$ along the direction $\hat{u}$, evaluated at $u$. $E_s$ favors functions that are smooth along radial directions. We include $E_s\{E[P]\}$ in the total energy functional to be minimized.

### 7.2.4 Anisotropy Regularization

Our azimuthal bases provides somewhat coarse-grained control on the level of anisotropy via the integral parameter $N_\alpha$. To have finer level of control, we define an energy functional $E_a$,

$$ E_a\{f\} = \iint_{\mathbb{R}^2} |D_{\hat{u}^\perp} f(u)|^2 du, $$  

where $\hat{u}^\perp = \text{rot}_{\hat{z}} (u/\|u\|)$ is the circumferential direction at $u$, and $D_{\hat{u}^\perp} f(u)$ is the first-order (directional) derivative of $f$ along the direction $\hat{u}^\perp$, evaluated at $u$. $E_a$ favors functions with low rate of change along circumferential directions, i.e., isotropic functions. We include $E_a\{E[P]\}$ in the total energy functional to be minimized.
7.2.5 Total Energy Functional

The overall quadratic energy functional on \( \{ c_{j,k,l} \} \) to be minimized is then

\[
\sum_{h_z, \beta} \left\{ \| p_{h_z}^h - \sum_{j,k,l} c_{j,k,l} \text{pr}_j \Psi_{h}^{h_z} \|_2^2 \right. \\
+ \left. \kappa_s E_s \left\{ \sum_{j,k,l} c_{j,k,l} \Psi_{h}^{h_z} \right\} \right. \\
\left. + \kappa_a E_a \left\{ \sum_{j,k,l} c_{j,k,l} \Psi_{h}^{h_z} \right\} \right\},
\]

(7.7)

where the range of \( h_z \) is as discussed in Section 7.2.2, the range of \( \beta \) corresponds to the azimuthal directions along which profile measurements are available (Section 4.4), the superscript \( h_z \) on \( p \) and \( \Psi_{j,k,l} \) indicates their dependency on \( h_z \), the subscript \( \beta \) on \( p \) and \( \text{pr}_j \) indicates their dependency on \( \beta \), \( \kappa_s \) is a user-supplied smoothness regularization weight, and \( \kappa_a \) is a user-supplied anisotropy regularization weight.

The coefficients \( \{ c_{j,k,l} \} \) can be constrained to be nonnegative if the probabilistic interpretation (Section 6.4.1) is to be enforced, though we do not enforce it. We note that because of the piecewise-linear azimuthal basis used, \( E_s \) and \( E_a \) can be computed as the weighted sum of one-dimensional integrals evaluated at \( \{ \alpha_l \} \) (Section 7.1.3) instead of performing full two-dimensional integration.

7.3 Results and Discussion

In this section we experimentally evaluate the proposed reflectance prediction algorithm on the 10um, 17um, and mixture surface samples. Refer to Chapter 5 on how these samples are fabricated and collected their profile and reflectance measurements.

Our bases choice and optimization procedure are affected by the following parameters:

- \( \kappa_s \), the smoothness regularization weight,
- \( r_{\text{max}} \), the estimated upper bound on facet radius,
- \( N_\alpha \), the number of azimuthal angular basis functions, and
• \( \kappa_a \), the anisotropy regularization weight.

Other parameters such as the number of discrete shapes and slant angles, \( N_\Omega \) and \( N_\theta \), can be safely set to some large numbers. Lowering them to reduce computational cost will not be explored here.

We first present results under some suitably chosen parameter values: \( \kappa_s = 5 \cdot 10^3 \), \( r_{\text{max}} = 3\text{um} \), \( N_\alpha = 2 \), and \( \kappa_a = 0 \). The absolute scale of \( \kappa_s \) and \( \kappa_a \) are not important as they depend on discretization. The effects of the parameters will be explored after. Note that reflectance plots presented in Sections 7.3.1 to 7.3.5 all use a single scale to match prediction to measurements for all samples and for all parameter values explored.

7.3.1 Overall Prediction

The input to our algorithm are the MPSD projections estimated from the one-dimensional profiles. We plot them in Fig. 7.1a for \( h_z = 1 \). These estimates are noisy, due to the roughness of our surfaces and the limited number of profiles we could acquire. In particular, the mixture sample, for which we had the fewest measurements, also has the noisiest estimate. Fig. 7.1b shows the reflectance computed from the estimated MPSD projections together with the optical ground-truths. Excellent agreement between prediction and measurement can be observed. Both the relative magnitudes of the peaks and the attenuation rates away from peaks are accurately specified. In particular the small difference between the 17um sample and the mixture sample is correctly and quantitatively characterized.

7.3.2 Effect of Smoothness Regularization Weight (\( \kappa_a \))

Fig. 7.2 plots the reflectance computed for different values of \( \kappa_a \). For our samples, values within the window \([3 \cdot 10^3, 10^4]\) all produce reasonable results, so the effects of
\( \kappa_s \) is relatively stable. When \( \kappa_s \) is too large, the curves are expectedly over-flattened. When \( \kappa_s = 0 \) the overall shape and peak magnitudes are still somewhat accurately computed, although the shape around the peak regions are poorly characterized for the 17um and the mixture samples. Fig. 7.1a confirms that the estimated MPSD projections for these two are indeed more noisy than for 10um sample. These noises are unevenly magnified around the peak region in the solution to the linear inverse problem. Moreover, the non-averaged optical reflectance measurements of these samples plotted in Figs. 5.6b and 5.6c also exhibit high variance around the peak, which indicates that the physical surfaces themselves exhibit high spatial variance for reflectance around the peaks. Noises in the estimated MPSD projections imply that profile sampling was too sparse. It is conceivable that if more measurements are available to make the estimated MPSD projections less noisy for the rougher 17um and mixture samples, their predictions would be more well-behaved around the peak region even without smoothness regularization, just like the 10um sample.

Note that the effect of \( \kappa_s \) is markedly different from that of the empirical parameters in the predictive studies that assume parametric surface models (Section 2.5.2) and the empirical kernel size parameter that controls tangent-plane-fitting in the geometrical-optics-based prediction of [Dong et al., 2015]. There the empirical param-
Figure 7.2: Reflectance computed using different smoothness regularization weights ($\kappa_s$).

Parameters could not be properly chosen without consulting the optical measurements no matter how much additional profilometer measurement were made available. Those parameters also directly control the peak magnitude and the shape of the predicted reflectance distribution. For example, if the kernel size parameter of [Dong et al., 2015] is set too small, the predicted reflectance will be too diffuse; if it is set too large, the predicted reflectance will be too specular. The optimal parameter value is likely slightly different for different surfaces, but because the peak magnitude of the reflectance is sensitive to this choice, when a single parameter value is used for multiple surfaces with identical material composition but different roughness, the relative peak magnitudes of the predicted reflectance will likely be inaccurate.

In our case, we can choose $\kappa_s$ using the following simple heuristic: Let $\kappa_s$ increase from 0; the peak value of the MPSD should first increase (and possibly fluctuate) a little, then stabilize over a large window, and eventually fall off; now just set $\kappa_s$ to some value in the middle of its stabilizing window.
7.3.3 Effect of Estimated Upper Bound on Facet Radius ($r_{\text{max}}$)

Fig. 7.3 plots reflectance computed using $r_{\text{max}}$ set to 2um, 4um, and 12um. Also plotted are the estimated MPSD projections, $p \approx \text{prj } P$, and projections of the MPSD dictionary elements, i.e., $\text{prj } \Psi_{j,k,l}$, constructed from the first six slant angles, $\theta_1 (= 0), \theta_2, \ldots, \theta_6$, and the narrowest shape basis function, which is the PSD of a disk with radius $r_{\text{max}}$. A fixed $h_z (= 1)$ is used for plots of the projections. The overall shapes and the relative peak magnitudes are accurately predicted for all different settings of $r_{\text{max}}$. What differs is the sharpness of the peaks. $r_{\text{max}}$ specifies the width of the narrowest shape basis function. A large $r_{\text{max}}$ introduces sharper basis elements into the dictionary $\{\Psi_{j,k,l}\}$ which then allow a sharper reflectance curve. When estimated MPSD projections contain noises, these sharper basis elements allow the noises to be over-fitted, leading to inferior predictions. Nonetheless the prediction quality is clearly not very sensitive to the choice of $r_{\text{max}}$. As mentioned in Section 7.1.1, a simple heuristic for setting $r_{\text{max}}$ is to scan the profilometer measurements for the longest straight-line segment (within some threshold).

![Figure 7.3: Reflectance (top row) and projections of MPSD dictionary elements (bottom row) computed using different upper bounds on facet radius ($r_{\text{max}}$).](image-url)
7.3.4 Effect of Azimuthal Basis Size ($N_{\alpha}$)

Fig. 7.4 plots reflectance computed using $N_{\alpha} = 1$ for two different choices of $r_{\text{max}}$, along with projections of the MPSD dictionary elements. As expected, the curves become more symmetrical as the surface is assumed to be isotropic. The asymmetry present when $N_{\alpha} = 2$ arises from the asymmetry of the estimated MPSD projections $p \approx \text{prj } P$, i.e., Fig. 7.1a. A single azimuthal basis function cannot express this asymmetry, but two azimuthal basis functions centered at 0 and $\pi$ can. Because our sample materials are nearly isotropic, the quality of approximation stays roughly the same for both settings.

![Graphs showing reflectance and projections for different $r_{\text{max}}$ values](image)

(a) $r_{\text{max}} = 3\text{um}$  
(b) $r_{\text{max}} = 12\text{um}$

**Figure 7.4:** Reflectance (top row) and projections (bottom row) of MPSD dictionary elements computed using a single azimuthal basis ($N_{\alpha} = 1$) and different upper bounds on facet radius ($r_{\text{max}}$).

One might wonder if the isotropy assumption removes the need for smoothness regularization. This is not the case, as shown in Fig. 7.5. Without smoothness regularization, the noises in the estimated MPSD projections for the 17$\text{um}$ and the mixture samples flatten the peaks in the computed reflectance.

123
7.3.5 Effect of Anisotropy Regularization Weight ($\kappa_a$)

Compared to setting $N_\alpha = 1$, anisotropy regularization controlled by $\kappa_a$ is a more flexible tool for controlling the level of reflectance isotropy. Fig. 7.6 plots reflectance computed with different $\kappa_a$.

7.3.6 Effect of Using a Single Shape Basis Function ($N_\Omega = 1$)

The classic microfacet model can be simulated by a shaped microfacet model that uses a single shape basis corresponding to a facet size significantly greater than the wavelength. Fig. 7.7 plots predicted reflectance along with projections of MPSD dictionary elements computed using $N_\Omega = 1$ and various values for $r_{\text{max}}$. Unlike plots in earlier sections, which all use the same scaling constant to fit predictions to measurements irrespective of parameters, here the scaling constant is optimally-fitted for each different $r_{\text{max}}$. Predictions for different samples within a single plot are still scaled using a single constant.
Figure 7.7: Reflectance and projections of MPSD dictionary elements computed from a single shape basis element ($N_{\Omega} = 1$) and different values for $r_{\text{max}}$. 
From the plots, we observe that as $r_{\text{max}}$ increases beyond 6\,um, the predicted reflectance stabilizes to three fixed curves. These curves should correspond to solving the linear inverse problem using the classic microfacet model as the prior. The relative ratios of the peak magnitude of the curves are predicted with less accuracy: The peaks of the 17\,um and the mixture sample are no longer distinct, and the ratio between the peaks of the 10\,um and the 17\,um sample is noticeably higher than indicated by the optical measurements. The overall shapes of the three curves are also more concentrated than that of the ground truth.

Note that setting $r_{\text{max}} = 2\,\text{um}$ and $r_{\text{max}} = 3\,\text{um}$ results in predictions that are more accurate than when $r_{\text{max}} = 12\,\text{um}$, though still inferior to those made using multiple shape basis functions. The implication is that if the classic microfacet model is enhanced with even just a single fixed facet size, it will be able to represent our test surfaces with much greater accuracy.

### 7.4 Summary

In this chapter we presented and experimentally evaluated an algorithm that predicts surface reflectance from one-dimensional profile measurements by fitting a (nonparametric) shaped microfacet model to the estimated MPSD projection. We specified our choices for the shape, slant and azimuthal bases used for constructing the MPSD dictionary $\{\Psi_{j,k,l}\}$ according to the shaped microfacet model. We next formulated a quadratic minimization for estimating the coefficients associated with these dictionary elements. These coefficients determine the MPSD and hence the surface reflectance.

Our optimization-based algorithm uses quite a few parameters, which can be a worrisome sign that the prediction procedure requires per-material empirical tuning. However, our experimental evaluation demonstrated that all the parameters have
stable and intuitive effects on the prediction and can be set in an automatic manner based on the profile measurements alone. Nonetheless, among the parameters, the smoothness regularization weight admittedly has no intrinsic physical interpretation, and is introduced solely to combat noises in the estimated MPSD projections. In the next chapter, we present a completely different algorithm for estimating the MPSD. In particular, it leads to a principled method for denoising the estimated MPSD projections, thereby rendering smoothness regularization less necessary.
Chapter 8

Modulated Autocorrelation: A Dual Perspective

Chapter 6 and Chapter 7 introduced an optimization-based algorithm for recovering the MPSD from estimates on its projection. The algorithm incorporates many prior assumptions on the unknown surface, including the presumed shaped microfacet model, an estimated upper bound on facet size, and the expected level of surface anisotropy.

We are, in some sense, estimating the MPSD via a circuitous route: Instead of finding the solution directly in the space of possible bivariate functions, we introduced a different problem space with a lot of extra structures and solved in that space. The mapping from the new search space, that is, the space of joint facet shape-and-orientation distribution, back to the space of MPSDs is by no means injective. Hence much recovered information will be discarded. A potential benefit of taking this route is that we now know more about the geometry of the surface (or rather, a surface satisfying both the projected MPSD and our priors), and can potentially repurpose this knowledge for other applications, such as editing the reflectance function in
a physically meaningful way, or fabricating surfaces to achieve the look of a given material.

It is also logical to ask if it is possible to recover the MPSD without hallucinating the extra information. Given a function $f$, the inverse Fourier transform of its PSD is known as the \textit{autocorrelation function} (ACF) of $f$. In this chapter, we will investigate the ACF of the surface modulation function, instead of its PSD. Since the Fourier transform does not destroy or add information, we are, in some sense, taking a more direct approach than before and staying within the same search space.

In Section 8.1 we first introduce the \textit{modulated autocorrelation function} (MACF), which is the inverse Fourier transform of the MPSD and hence an equivalent characterization capable of fully deciding surface reflectance. Then in Section 8.2, we take a small digression to properly describe how statistics such as MACF and MPSD computed from surface measurements within multiple small windows (corresponding to the AFM scanning window) are related to the \textit{average} statistics of the entire surface. Next in Sections 8.3 and 8.4 we describe an algorithm that estimates the MACF and hence surface reflectance using simple linear interpolation. Finally in Section 8.5 we evaluate the proposed algorithm using experimental measurements collected in Chapter 5.

\section{Modulated Autocorrelation Function}

Let $f(x)$, $x \in \mathbb{R}^n$ be a function with compact support. Define the operator $\mathcal{R}$ as follows,

$$\mathcal{R}\{f\}(d) = \int_{\mathbb{R}^n} f(x)\bar{f}(x-d) \, dx = (f * \bar{f}_\leftarrow)(d)$$

where $*$ denotes the convolution operation, and $g_\leftarrow(x) = g(-x)$ for any function $g$. From the above definition, we can see that $\mathcal{R}\{f\}(d)$ relates to the notion of the “average” correlation between function values evaluated at two points that differ by
d, with the caveat that the integral is not scaled by the inverse of the total “area” of the set \( \{ x : \text{both } f(x) \text{ and } f(x - d) \text{ lie in the support of } f \} \) as one would expect for an “average” quantity. As explained later in Section 8.2, it is in fact possible to make the above connection exact using the language of stochastic processes. Hence we will refer to \( R\{f\} \) as the autocorrelation function (ACF) of \( f \).

Recall from Section 4.2 that the power spectral density (PSD) of a function \( f \), denoted by \( S\{f\} \) from now on, is defined as

\[
S\{f\} = |\mathcal{F}\{f\}|^2.
\]  
(8.2)

By applying the Fourier transform identities (i) \( \mathcal{F}\{f_\perp\} = \mathcal{F}\{f\}_\perp \) (ii) \( \mathcal{F}\{\bar{f}\} = \overline{\mathcal{F}\{f\}} \) and (iii) \( \mathcal{F}\{f * g\} = \mathcal{F}\{f\}\mathcal{F}\{g\} \), we can obtain the following equality,

\[
\mathcal{F}\{R\{f\}\} = S\{f\}.
\]  
(8.3)

In other words, the ACF and the PSD of a function \( f \) form a Fourier transform pair and hence encode an equal amount of information.

Under the Kirchoff approximation, surface reflectance is determined by the modulated power spectral density (MPSD), which is the PSD of the surface modulation function \( m \) (Section 3.2.4). It follows that the ACF of the surface modulation function, \( R\{m\} \), provides an equivalent characterization of surface reflectance. We will refer to \( R\{m\} \) as the modulated autocorrelation function (MACF). While the MPSD is defined over the somewhat elusive frequency domain, the MACF is defined over the more tangible spatial domain. Examining our core computational problem, i.e., recovering the MPSD \( S\{m\} \) given its projection, from the dual perspective of the MACF \( R\{m\} \) can provide insights to the problem.
8.2 Interpretation Using Stochastic Processes

Surface reflectance as determined by the MPSD accounts only for the area over which the Kirchhoff integral is taken. An optical gonioreflectometer typically measures the reflectance averaged over an area much larger than the AFM scanning window. Using AFM we could not hope to measure the surface modulation function defined over the entire surface, say $A_{\text{opt}}$, whose reflectance is optically measured, nor is necessary to do so as the area over which the incident light stays coherent is much smaller than $A_{\text{opt}}$ anyway. We instead use AFM to measure surface profiles within much smaller windows, 80um in width and height, located at multiple lateral positions, and assume that the “average” projected MPSDs computed from these windows determine the expected projected MPSDs of all such windows. In this section, we use the language of stochastic processes to make the above notion precise.

The PSD and the ACF defined so far all operate on a single function with compact support. They can also be viewed as estimates on the power spectrum and the autocorrelation of a wide-sense stationary (WSS) stochastic process [Papoulis and Pillai, 2002; Oppenheim and Verghese, 2015] under suitable conditions, as explained next.

Given a random process $f(x)$, $x \in \mathbb{R}^2$, its autocorrelation is defined as

$$R_{ff}(x_1, x_2) = \mathbb{E}[f(x_1)f(x_2)].$$

(8.4)

Note that the double $f$ in the subscript indicates that the autocorrelation of $f$ is the cross-correlation between $f$ and itself. Unsurprisingly, the cross-correlation between two processes $f$ and $g$ is just $R_{fg}(x_1, x_2) = \mathbb{E}[f(x_1)g(x_2)]$, though we will not make use of it in the present work.

If the process $f$ is in addition WSS, its autocorrelation $R_{ff}(x_1, x_2)$ will depend
only on \((x_1 - x_2)\), and hence we can write

\[
R_{ff}(d) = R_{ff}(x, x - d) \text{ for any } x \\
= \mathbb{E}[f(x)f(x - d)]
\]  

(8.5)

It is a more involved task to define the power spectrum of a WSS process. A realization of the WSS process is a plain function defined over \(\mathbb{R}^2\) and in general has infinite support. We can, however, force a realization to be more well-behaved by point-multiplying with a window function \(\Pi_W\),

\[
\Pi_W(x) = \begin{cases} 
1, & \text{if } x \in [-W/2, W/2] \times [-W/2, W/2] \\
0, & \text{otherwise}.
\end{cases}
\]

We can consider the PSD of such a windowed realization scaled by the inverse of the area of the window, which measures the “average” PSD over the defined window,

\[
\frac{1}{W^2} S\{\Pi_W f\},
\]

(8.6)

The above quantity is known as a periodogram of the process. The power spectrum of a WSS process \(f\) is then defined to be the expected value of such periodograms with the averaging window size tending to infinity,

\[
S_{ff} = \lim_{W \to \infty} \mathbb{E} \left[ \frac{1}{W^2} S\{\Pi_W f\} \right],
\]

(8.7)

The Wiener-Khinchin theorem states that the autocorrelation and the power spectrum of a WSS process form a Fourier transform pair,

\[
\mathcal{F}\{R_{ff}\} = S_{ff}.
\]

(8.8)

Also note that the expected value of the ACF of windowed realizations equals the autocorrelation of the process scaled by a bilinear hat function \(\Lambda_W\) dependent on
the window size,

\[
\Lambda_W R_{ff} = \mathbb{E} \left[ \frac{1}{W^2} R \{ \Pi_W f \} \right]
\]

\[
\Lambda_W (d) = \max \left( \frac{W}{2} - |d_x|, 0 \right) \max \left( \frac{W}{2} - |d_y|, 0 \right).
\]  \hfill (8.9)

Thus, as we alluded to right after defining the ACF in Eq. 8.1, for estimating the expected correlation between points, \( R \{ f \} \) does not properly weight the integral \( \int f(\mathbf{x}) f(\mathbf{x} - \mathbf{d}) d\mathbf{x} \). The proper weight is in fact \( 1/\Lambda_W (\mathbf{d}) \).

We can now describe our process of averaging the MPSDs obtained from multiple AFM scanning windows using the language of stochastic processes as follows: The surface modulation function is a WSS process \( m*(x) \) defined on the entire plane \( \mathbb{R}^2 \). Each time we perform an AFM scan on a \( W \) by \( W \) area, we obtain a windowed realization, \( m = \Pi_W m* \) of the process. For a fixed realization \( m \), its projected PSD \( \text{prj} S \{ m \} \) equals the expected one-dimensional PSD of its horizontal slices, \( \{ m(\cdot, Y) \} \), with \( Y \) distributed uniformly in \( [-W/2, W/2] \),

\[
\text{prj} S \{ m \} = \int S \{ m(\cdot, y) \} dy = W \mathbb{E}_{Y \sim U( -\frac{W}{2}, \frac{W}{2} )} [ S \{ m_Y \} ], \text{ } \hfill (8.10)
\]

where \( U(a,b) \) is the uniform distribution on the interval \( [a,b] \), and \( m_Y \) denotes the univariate function \( m(\cdot, Y) \). The first equality above is proved in Section 4.2 (Eq. 4.3), and the second equality follows from the definition of the expected value. Thus, for a fixed realization \( m \), by acquiring multiple horizontal slices of \( m \), we can estimate \( \text{prj} S \{ m \} \).

When we average the projected MPSD computed from multiple AFM scans taken at different lateral positions, we are estimating the expected value of \( \text{prj} S \{ m \} \) from multiple realizations of \( m \), which corresponds to taking the expected value of Eq. 8.10,

\[
\mathbb{E}[\text{prj} S \{ m \}] = W \mathbb{E}_m \mathbb{E}_Y [ S \{ m_Y \} ]. \text{ } \hfill (8.11)
\]
By noting that \( \text{prj} \) is linear and that \( \mathbb{E}_m \mathbb{E}_Y \) is just \( \mathbb{E} \), we obtain the follows,

\[
\text{prj} \mathbb{E}[S\{m\}] = W \mathbb{E}[S\{m_Y\}] \tag{8.12}
\]

where the expectation on the left side is over \( m \), and the expectation on the right side is over \( m \) and \( Y \). The above equation fully describes how \( \text{prj} \mathbb{E}[S\{m\}] \) is estimated:

We randomly pick windowed realizations \( \{m = \Pi_W m_\ast\} \); For each \( m \), we acquire its horizontal slices, \( \{m_Y\} \), corresponding to multiple samples of \( Y \); Finally we compute the one-dimensional PSD of these slices, \( \{S\{m_Y\}\} \), and average them to obtain an estimate on \( \mathbb{E}[S\{m_Y\}] \) and hence \( \text{prj} \mathbb{E}[S\{m\}] \). From \( \text{prj} \mathbb{E}[S\{m\}] \), we then solve for \( \mathbb{E}[S\{m\}] \), the expected MPSD.

It is a fair question to ask if we should seek unbiased estimates on the autocorrelation and the power spectrum of \( m_\ast \), e.g., by scaling the expected MACF \( \mathbb{E}[R\{m\}] = \mathbb{E}[R\{\Pi_W m_\ast\}] \) pointwise by \( 1/\Lambda_W \) (Eq. 8.9). However, the Kirchhoff integral should only be taken over a finite region over which the incident illumination is coherent. Hence, it is really the expected periodogram \( \mathbb{E}[S\{m\}] = \mathbb{E}[S\{\Pi_W m_\ast\}] \) for some finite \( W \) that should be used to compute surface reflectance (Section 3.2.4), not the power spectrum of the process \( m_\ast \), \( S_{m_\ast m_\ast} \). In any case, when \( W \) is sufficiently large, \( \mathbb{E}[S\{m\}] \) and \( S_{m_\ast m_\ast} \) will be very similar, and \( 1/\Lambda_W \) will be roughly constant over the nonzero support of \( \mathbb{E}[R\{m\}] \).

### 8.3 Estimating Expected MACF Using Interpolation

In this section, we consider computing the expected MACF \( \mathbb{E}[R\{m\}] \) instead of the expected MPSD \( \mathbb{E}[S\{m\}] \). First define the slice operator, \( \text{slc} \), which takes a bivariate function and returns its horizontal slice through the origin,

\[
(\text{slc} f)(x) = f(x, 0). \tag{8.13}
\]
The Fourier central slice theorem states that \( \text{prj} \circ \mathcal{F} = \mathcal{F} \circ \text{slc} \), which can be expressed using the following commutative diagram,

\[
\begin{array}{c}
\text{prj} \mathbb{E}[\mathcal{S}\{m\}] & \xleftarrow{\text{prj}} & \mathbb{E}[\mathcal{S}\{m\}] \\
\mathcal{F} & \quad & \mathcal{F} \\
\text{slc} \mathbb{E}[\mathcal{R}\{m\}] & \xleftarrow{\text{slc}} & \mathbb{E}[\mathcal{R}\{m\}].
\end{array}
\] (8.14)

In addition, Eq. 8.12 shows that \( \text{prj} \mathbb{E}[\mathcal{S}\{m\}] = \mathcal{W} \mathbb{E}[\mathcal{S}\{m_Y\}] \). Hence we can replace entries on the left column of the above diagram by \( \mathcal{W} \mathbb{E}[\mathcal{S}\{m_Y\}] \) and its inverse Fourier transform, \( \mathcal{W} \mathbb{E}[\mathcal{R}\{m_Y\}] \), to obtain the following commutative diagram,

\[
\begin{array}{c}
\mathcal{W} \mathbb{E}[\mathcal{S}\{m_Y\}] & \xleftarrow{\text{prj}} & \mathbb{E}[\mathcal{S}\{m\}] \\
\mathcal{F} & \quad & \mathcal{F} \\
\mathcal{W} \mathbb{E}[\mathcal{R}\{m_Y\}] & \xleftarrow{\text{slc}} & \mathbb{E}[\mathcal{R}\{m\}].
\end{array}
\] (8.15)

We remark the following:

- The top row of the diagram relates expected MPSDs, defined over the frequency domain, whereas the bottom row relates expected MACFs, defined over the spatial domain.

- The one-dimensional expected MPSD and MACF on the left column of the diagram can be estimated from the one-dimensional surface profiles alone, while the two-dimensional expected MPSD and MACF on the right column determine surface reflectance.

- As the Fourier transform is information preserving, we can follow the vertical arrows in both directions. In contrast, the horizontal arrows, corresponding to \( \text{prj} \) and \( \text{slc} \), discard information.

- Instead of showing that \( \text{prj} \mathbb{E}[\mathcal{S}\{m\}] = \mathcal{W} \mathbb{E}[\mathcal{S}\{m_Y\}] \) as we have done, we can
arrive at the above commutative diagram by equivalently showing that

\[ \text{slc} \mathbb{E}[\mathcal{R}\{m\}] = W \mathbb{E}[\mathcal{R}\{m_Y\}]. \]  

(8.16)

To see that the above holds, note that \( \mathbb{E}[\mathcal{R}\{m\}](d) \) measures the expected correlation between \( m_*(x) \) and \( m_*(x - d) \), weighted by \( \Lambda_W \). If we simply let \( d = [d, 0]^T \), we obtain \( \text{slc} \mathbb{E}[\mathcal{R}\{m\}] \). But under this restriction, the expected correlation between \( m_*(x) \) and \( m_*(x - d) \) is simply the (one-dimensional) autocorrelation of a random slice of \( m_* \), which after suitable weighting, is given by \( \mathbb{E}[\mathcal{R}\{m_Y\}] \).

Our previous efforts in Chapters 6 and 7 at computing \( \mathbb{E}[\mathcal{S}\{m\}] \) can now be viewed as inverting the frequency domain operator, \( \text{prj} \), on the top row of the diagram. We next consider inverting the spatial domain operator, \( \text{slc} \), by working from the bottom row of the diagram. In this case, the appropriate thing to do seems a lot more obvious as shown next.

8.3.1 Circumferential Linear Interpolation

Just as with \( \text{prj}_\beta \), we generalize \( \text{slc} \) by defining

\[ \text{slc}_\beta f = \text{slc}(f \circ \text{rot}_-\beta), \]  

(8.17)

so \( \text{slc}_\beta f \) is the slice of \( f \) corresponding to the X axis rotated counterclockwise by the angle \( \beta \). Suppose our input measurements contain height profiles along a direction defined by the angle \( \beta \). We can calculate the corresponding expected one-dimensional MACF along the direction \( \beta \), denoted by \( \mathbb{E}[\mathcal{R}\{m_\beta,Y\}] \), where \( m_\beta,Y \) is a random horizontal slice of a windowed realization of \( (m_* \circ \text{rot}_\beta) \).

Each one-dimensional expected MACF \( \mathbb{E}[\mathcal{R}\{m_\beta,Y\}] \) is then a slice along \( \beta \) of the two-dimensional expected MACF \( \mathbb{E}[\mathcal{R}\{m\}] \) that passes through the origin. To recover
the appropriate thing to do is to interpolate these slices circumferentially. In other words, we assume the prior that any radial slice of $E[R\{m\}]$ is the linear interpolation between the two closest known radial slices of $E[R\{m\}]$.

Towards this end, we define a *circumferential linear interpolation* operator as follows: Let $\theta_1, \ldots, \theta_K$ be a finite collection of $K$ angles ordered increasingly, i.e., $0 \leq \theta_1 < \cdots < \theta_K < 2\pi$; Let $f_1, \ldots, f_K$ be $K$ complex-valued functions defined on nonnegative reals, $f_k : [0, \infty) \to \mathbb{C}$, such that they agree at 0, i.e., $f_1(0) = \cdots = f_K(0)$; Let $\theta = (\theta_1, \ldots, \theta_K)$ and $f = (f_1, \ldots, f_K)$; Finally define the operator $clerp_\theta$,

$$(clerp_\theta f)(\text{rot}_\beta \mathbf{x}, 0) = f_k(x) + \frac{\beta - \theta_k}{\theta_{k+1} - \theta_k} [f_{k+1}(x) - f_k(x)] \quad \text{for} \quad \beta \in [\theta_k, \theta_{k+1}), x \geq 0,$$

where the index $k$ ranges over $1, \ldots, K$, and indexing by $K+1$ is interpreted such that $\theta_{K+1} = \theta_1 + 2\pi$ and $f_{K+1} = f_1$. Thus $clerp_\theta$ interpolates the univariate functions $\{f_k\}$ so that $slc_\theta clerp_\theta f(x) = f_k(x)$ for $x \geq 0$.

The two-dimensional expected MACF can then be estimated from the one-dimensional expected MACFs as follows: Let $\{r_k = E[R\{m_{\beta_k,Y}\}] : k = 1, \ldots, K\}$ be the one-dimensional expected MACFs estimated from one-dimensional surface profiles along the directions $\{\beta_k\}$; Without loss of generality, we can assume $0 \leq \beta_1 < \beta_2 < \cdots < \beta_k < \pi$, since if some $\beta_j \in [\pi, 2\pi)$, we can simply replace it with $(\beta_j - \pi)$ and flip the corresponding one-dimensional expected MACF $r_j$; From each $r_k : \mathbb{R} \to \mathbb{C}$, we obtain two radial slices, $r^+_k : [0, \infty) \to \mathbb{C}$ and $r^-_k : [0, \infty) \to \mathbb{C}$, that correspond to values of $r_k$ along the positive axis and the negative axis, as $clerp$ works on functions defined on nonnegative reals; Finally estimate the two-dimensional
expected MACF $E[R \{m\}]$ via circumferential linear interpolation as

$$E[R \{m\}] = \text{clerp}_\beta \mathbf{r}$$

$$\mathbf{r} = (r_1^+, \ldots, r_K^+, r_1^-, \ldots, r_K^-)$$

$$\beta = (\beta_1^+, \ldots, \beta_K^+, \beta_1^-, \ldots, \beta_K^-)$$

$$r_k^+(x) = Wr_k(x) \text{ for } x \geq 0 \quad (8.19)$$

$$r_k^-(x) = Wr_k(-x) \text{ for } x \geq 0$$

$$\beta_k^+ = \beta_k$$

$$\beta_k^- = \beta_k + \pi.$$

More sophisticated interpolation schemes can also be employed to obtain smoother azimuthal variations for $E[R \{m\}]$.

### 8.3.2 Interpolating Property of MPSD Corresponding to Interpolated MACF

The two-dimensional MACF $E[R \{m\}]$ estimated using circumferential interpolation obviously satisfies that $\text{slc}_{\beta_k} E[R \{m\}] = r_k$. Due to the duality of slc and prj, its inverse Fourier transform, the expected MPSD $E[S \{m\}]$ also satisfies the following interesting interpolating property: Let $\{s_k = E[S \{m_{\beta_k} Y\}]\}$ be the estimated MPSD projections estimated from one-dimensional profiles along directions $\{\beta_k\}$; Then the projection of $E[S \{m\}]$ along some direction $\theta$ is also a slice along $\theta$ of the
circumferential linear interpolation of the estimated MPSD projection \( \{s_k\} \), i.e.,

\[
\text{prj}_\theta \mathbb{E}[S\{m\}] = \text{slc}_\theta \text{clerp}_\beta s
\]

\[
s = (s_1^+, \ldots, s_K^+, s_1^-, \ldots, s_K^-)
\]

\[
s_k^+(x) = Ws_k(x) \quad \text{for} \quad x \geq 0
\]

\[
s_k^-(x) = Ws_k(-x) \quad \text{for} \quad x \geq 0
\]

\[
s_k = \mathbb{E}[S\{m_{\beta_k,Y}\}] = \mathcal{F}^{-1}\{\mathbb{E}[R\{m_{\beta_k,y}\}]\}, \tag{8.20}
\]

where \( \beta \) is as defined in Eq. 8.19. The above can be derived by applying \( \text{slc}_\theta \) to both sides of Eq. 8.19 and noting that \( \text{slc}_\theta \mathbb{E}[R\{m\}] = \text{slc}_\theta \mathbb{E}[\mathcal{F}^{-1}\{S\{m\}\}] = \mathcal{F}^{-1} \text{prj}_\theta \mathbb{E}[S\{m\}] \) and \( \mathcal{F}\{\text{slc}_\theta \text{clerp}_\beta s\} = \text{slc}_\theta \text{clerp}_\beta r \).

### 8.3.3 Computational Considerations

Some additional computational considerations are worth mentioning. If only a slice of the two-dimensional expected MPSD \( \mathbb{E}[S\{m\}] \) needs to be evaluated, then the following commutative diagram, obtained by augmenting Eq. 8.15 using the Fourier slice theorem again, will come handy.

\[
\begin{array}{ccc}
W\mathbb{E}[S\{m_Y\}] & \xrightarrow{\text{prj}} & \mathbb{E}[S\{m\}] & \xrightarrow{\text{slc}_\beta} & \mathbb{E}[S\{m\}] \\
\mathcal{F} & & \mathcal{F} & & \mathcal{F}
\end{array}
\]

\[
\begin{array}{ccc}
W\mathbb{E}[R\{m_Y\}] & \xleftarrow{\text{scl}} & \mathbb{E}[R\{m\}] & \xrightarrow{\text{prj}_\beta} & \mathbb{E}[R\{m\}].
\end{array}
\tag{8.21}
\]

Starting from the two-dimensional expected MACF (bottom middle entry in the diagram) estimated using interpolation, we can compute the sought slice (top right entry in the diagram) using either \( \text{scl}_\beta \circ \mathcal{F} \), i.e., by first following the up arrow and then the right arrow in the diagram, or \( \mathcal{F} \circ \text{prj}_\beta \), i.e., by first following the right arrow and then the up arrow in the diagram. The Fourier transform in the first route is a two-dimensional one, requiring \( O(n^2 \log n) \) time using FFT where \( n \) is the number of equispaced discrete samples used to represent a one-dimensional function such as the
estimated \( \mathbb{E}[\mathcal{S}\{m_Y\}] \). The Fourier transform in the second route is a one-dimensional one, requiring \( O(n \log n) \) time using FFT, and the projection operator \( \text{prj}_\beta \) requires \( O(n^2) \) time. Thus, the first route, \( \text{slc}_\beta \circ \mathcal{F} \), runs in \( O(n^2 \log n) \) time whereas the second route, \( \mathcal{F} \circ \text{prj}_\beta \), runs in \( O(n^2) \) time and is hence faster.

Of course if the entire two-dimensional expected MPSD \( \mathbb{E}[\mathcal{S}\{m]\}] \) is needed, e.g., for evaluating surface reflectance value for multiple azimuthal directions, then performing the two-dimensional Fourier transform \textit{once} can be faster than performing \( \mathcal{F} \circ \text{prj}_\beta \) repeatedly for each azimuthal direction \( \beta \).

### 8.4 Denoising MPSD

Using the simple circumferential interpolation algorithm described in the previous section, we unfortunately can only obtain the reflectance prediction shown in Fig. 8.1, where the peak region exhibits much oscillation, although the non-peak region is characterized with reasonable accuracy. The noisy oscillation around the peak region naturally reminds us of the noise present in the estimated MPSD projection \( \text{prj}(\mathbb{E}[\mathcal{S}\{m]\}]) \) (recall from Sections 4.3 and 7.3.1) due to incomplete sampling and measurement errors. One obvious idea worth trying is to smooth out the noise in these projections by low-pass filtering. In this section, we propose a principled heuristics for choosing the low-pass bandlimit for denoising the MPSD.

![Figure 8.1: Reflectance computed directly from the estimated one-dimensional MACF using circumferential linear interpolation.](image-url)
8.4.1 Truncating MACF

Based on the analysis in the previous section, we know that smoothing the estimated \( \text{prj} \mathbb{E}[\mathcal{S}\{m\}] \propto \mathbb{E}[\mathcal{S}\{m_Y\}] \) by applying a bandlimit \( l \) is equivalent to truncating the one-dimensional expected MACF \( \mathbb{E}[\mathcal{R}\{m_Y\}] \) at \( l \). The one-dimensional expected MACF equals the expected correlation between \( m_*(x, Y) \) and \( m_*(x - d, Y) \) weighted by \( \Lambda_W(d, 0) \), so truncating it is equivalent to adopting the prior assumption that \( m_*(x, Y) \) and \( m_*(x - d, Y) \) are uncorrelated when \( |d| > l \). The best bandlimit however is likely different for each different surface, so we seek a principled way of choosing it.

Fig. 8.2 plots both the one-dimensional expected MPSDs and the magnitude of the one-dimensional expected MACFs estimated from the measured surface profiles. The one-dimensional expected MPSDs are especially noisy around the peak region. The magnitude of the expected MACF peaks at zero and attenuates rapidly as distance from the origin increases. It however does not reach zero, but instead fluctuates with a tiny nonzero magnitude even as distance from the origin increases further. Given its rapid attenuation, we hypothesize that the ground-truth value of \( \mathbb{E}[\mathcal{R}\{m_Y\}](d) \) is zero for large \( d \), and the nonzero fluctuating tail exhibited in \( |\mathbb{E}[\mathcal{R}\{m_Y\}](d)| \) is caused by incidental correlation due mostly to measurement noise and sparsity.

8.4.2 Choosing Bandlimits By Assuming Nonincreasing MACF

We truncate the one-dimensional expected MACF using the following simple heuristic: We assume its magnitude, \( |\mathbb{E}[\mathcal{R}\{m_Y\}](d)| \), to be nonincreasing as \( |d| \) increases, and truncate it at the first point, say \( |d| = l \), where the magnitude increases as \( |d| \) increases. \( l \) is referred to as the default bandlimit. The violation of the nonincreasing property near the \( l \) is attributed to measurement sparsity and noise, and values of
Figure 8.2: One-dimensional MPSD and MACF (corresponding to a projection and a slice of the two-dimensional MPSD and MACF) estimated from measured one-dimensional surface profiles. The plots are for $h_z = 1$. The MACF is complex valued and Hermitian, so its magnitude along the positive axis is plotted. The plotted MACFs are weighted by $1/\Lambda_W(d,0)$, so what is shown are unbiased estimates of $|R_{m,m}(d,0)|$, which exhibits increased variance when $d$ is close to the window size $W$. 
\( \mathbb{E}[\mathcal{R}\{m_Y\}](d) \) for \(|d| \geq l \) is deemed unreliable and hence set to zero.

The assumption that \(|\mathbb{E}[\mathcal{R}\{m_Y\}](d)|\) is nonincreasing has in fact a physical interpretation. By the definition of the surface modulation function \( m \) (and its idealized stochastic process representation \( m^* \)), the one-dimensional expected MACF can be written as

\[
\mathbb{E}[\mathcal{R}\{m_Y\}](d) = \mathbb{E}[e^{ih[z(x,Y)−z(x−d,Y)]]}]\Lambda_W(d, 0). \tag{8.22}
\]

The effect of scaling by \( \Lambda_W(d, 0) \) is minor since it is approximately constant for \( d \) much smaller than \( W \), and judging from Fig. 8.2 the appropriate bandlimit is much smaller than \( W \). The nondecreasing property arises from the following reasoning: Because the surface is randomly rough, with \( Y \) fixed, the distribution on \((z(x,Y)−z(x−d,Y))\) will be more and more dilated as \(|d|\) increases, causing \( \mathbb{E}[e^{ih[z(x,Y)−z(x−d,Y)]]}] \) to be closer to the origin as \(|d|\) increases. It then follows that \(|\mathbb{E}[\mathcal{R}\{m_Y\}](d)|\) is nonincreasing as \(|d|\) increases.

Fig. 8.3 shows the default bandlimits for the three surface samples computed according to the above heuristic, as well as the effects of low-pass filtering the estimated one-dimensional MPSDs with different bandlimits. Observe that, from inspecting the one-dimensional MPSDs alone, it is hard to decide which bandlimit is more appropriate. Applying the default bandlimit, or one twice as wide, or one even four times as wide, all results in similar and reasonably smooth MPSD projections. However, as we show later in Section 8.5.1, when circumferential linear interpolation is used to construct the two-dimensional MPSD, the default bandlimits achieve much more accurate prediction than the alternatives.

More sophisticated smoothing strategies, such as taping off the one-dimensional MACF gradually instead of abruptly, might be employed, though we have not found it necessary in practice. It is also unclear if a more sophisticated method will admit more principled theoretical underpinning.
Figure 8.3: The default bandlimits and the effects on the expected MPSD when the expected MACF is truncated at different bandlimits corresponding to 0.5, 1, 2, and 4 times the default.

8.5 Results and Discussion

In this section we evaluate the proposed MACF-interpolation-based algorithm for reflectance prediction, using the 10um, 17um, and mixture surface samples. Refer to Chapter 5 on how these samples are fabricated and collected their profile and reflectance measurements. In addition, we demonstrate how the optimization-based prediction algorithm previously proposed in Chapter 7 interact with MPSD projections bandlimiting procedure described in this chapter.

8.5.1 Prediction using Circumferential Linear Interpolation

Fig. 8.4 shows the reflectance computed by estimating the two-dimensional MACF via circumferential linear interpolation. Different bandlimits are applied to the one-dimensional MACF used for interpolation. Prediction accuracy is clearly sensitive to the bandlimit choice, but the simple heuristic described in Section 8.4.2 leads to near-
exact prediction for all three surfaces. Using a much narrower bandlimit over-dilates the reflectance curves as expected. On the other hand, it is both interesting and surprising to see that bandlimits merely 50% wider than default would immediately introduce large oscillation to the prediction, especially considering that the nonzero tails of the estimated one-dimensional MACF have tiny magnitude (Fig. 8.2) anyway, and the one-dimensional MPSDs smoothed using different bandlimits look all very similar (Fig. 8.3).

Figure 8.4: Reflectance computed by estimating MACFs via circumferential linear interpolation. The one-dimensional MACFs used as input are truncated at 0.5, 0.8, 1, 1.5, 2.0, and 4.0 times the default bandlimits.

8.5.2 Effects of Bandlimits on Optimization-Based Prediction

Fig. 8.5 shows surface reflectance computed using the optimization-based algorithm presented in Chapter 7, using MPSD projection denoised using different bandlimits as input. The optimization-based algorithm essentially produces the same prediction irrespective of the amount of noise as long as the MPSD projections are not over-smoothed, i.e., when bandlimits narrower than the default are applied. This is in contrast to the sensitivity of the MACF-interpolation-based algorithm to the
bandlimits.

![Graphs showing reflectance computed using the optimization-based algorithm using MPSD projections low-pass filtered with different bandlimits.](image)

Figure 8.5: Reflectance computed using the optimization-based algorithm using MPSD projections low-pass filtered with different bandlimits.

Note however that the optimization-based algorithm makes quite a few strong prior assumptions of its own. In particular, smoothness regularization is the most important prior for allowing good prediction in the presence of noise in the estimated MPSD projections (Section 7.3.2), and serves an analogous purpose as the bandlimit. It is thus natural to ask if suitably bandlimited MPSD projections make smoothness regularization unnecessary. The answer is indeed yes, as shown in Fig. 8.6. When the MPSD projections are low-pass filtered with the default bandlimits, setting smoothness regularization weight $\kappa_s = 0$ still leads to near-exact prediction.

8.6 Summary

In this chapter, we tackled the problem of recovering the expected MPSD by working with its inverse Fourier transform, the MACF. The definition of the MACF immediately leads to a principled heuristic for choosing a bandlimit for low-pass filtering the noisy estimated MPSD projections. Moreover, the MACF-centric perspective
Figure 8.6: Reflectance computed using the optimization-based algorithm with different smoothness regularization weights and different MPSD projection bandlimits. The left and the middle columns correspond to applying the default bandlimits and 2 times the default. The right column corresponds to using the MPSD projections without bandlimiting.
allows us to design a conceptually and computationally simple interpolation-based algorithm for estimating the two-dimensional MACF from one-dimensional surface profile measurements. Our bandlimit heuristic and interpolation-based algorithm together produce near-exact prediction.

Despite of the simplicity of the interpolation-based algorithm, it too assumes some strong priors: namely (i) any radial slice of the expected MACF is assumed to equal the interpolation of known radial slices determined by the one-dimensional surface profiles; and (ii) the magnitude of the expected MACF is nonincreasing. Nonetheless, compared with the priors assumed in the optimization-based algorithm, these two are intrinsic to the search space and hence simpler and more obvious.
Chapter 9

Conclusion

This thesis demonstrated reflectance prediction from profilometer measurements with significantly better accuracy than shown before for non-precision-fabricated real-world two-dimensionally rough surfaces. The conventional wisdom seems to be that to predict the reflectance of a two-dimensionally rough surface, it is best to acquire two-dimensional measurements and then compute based on them. However, we identified why modern profilometers based on WLI and AFM are unlikely to produce two-dimensional topography maps with sufficient fidelity to allow accurate reflectance prediction. Instead, by relying only on one-dimensional surface profiles, we are able to acquire more accurate measurements to feed into the reflectance computation, which in turn enables the accurate prediction. Moreover, we do not loss much in the process, since one-dimensional measurements along different azimuthal directions can be used to specify the reflectance of anisotropic surfaces.

The core computational problem tackled in this thesis is that of recovering the two-dimensional (expected) MPSD given estimates on its projections. We proposed two algorithms for solving this ill-posed linear inverse problem, one based on optimization in the primal domain and the other on interpolation in the dual (Fourier) domain. The two can be viewed as belonging to a continuous spectrum, with optimization
being strictly more powerful and flexible but possibly less efficient. Interpolation has incorporated the prior that every dual-domain radial slice of the unknown signal is determined by the interpolation of known slices. Exactly one signal will satisfy such a strong prior. This dual-domain interpolating property is equivalent to the condition that all primal-domain projections of the unknown signal equal exactly the interpolation of known projections, as shown in Section 8.3.2. Hence an equivalent primal-domain optimization problem can be formulated by letting the primal-domain projections along all possible directions be specified according to the interpolation of known projections, and letting the search space be the set of all bivariate functions. Such a primal-domain optimization should in theory yield the same answer as dual-domain interpolation, though the latter is more efficient as it employs the Fourier slice theorem to directly construct the single solution. Of course, the actual primal-domain optimization algorithm proposed in this thesis used a different set of priors based on the shaped microfacet model.

Some possible future research directions include the follows:

- The anisotropic aspect of our prediction methods should be experimentally evaluated. For the optimization-based methods, application to anisotropic surfaces will likely require anisotropic shape basis functions to be employed. The interpolation-based methods, on the other hand, can be applied to anisotropic surfaces with little change.

- The proposed shaped microfacet model can be used to design new parametric BRDF models that support diffractive effects.

- This thesis shows that the MPSD (and equivalently the MACF) is the fundamental characterization of surface reflectance under the Kirchhoff approximation. New BRDF models can potentially be developed by modeling the MPSD and the MACF. In the context of inverse scattering, i.e., inferring topographical
statistics from optical reflectance measurements, our results imply that the MPSD and the MACF are perhaps more suitable candidates for estimation than traditional surface ACF and PSD.

- Due to the demonstrated accuracy, the proposed reflectance prediction methods can potentially be used to build a hybrid BRDF-measuring scheme in which optical measurement is only made on a sparse subset of the BRDF domain, for calibration and for handling situations where prediction will be less accurate, e.g., small grazing incident and reflection angles, while the remaining BRDF values can be computed from profilometer measurements, with arbitrary angular resolution and for any wavelength.
References


152


Editors of Webster’s New World College Dictionaries (2016). *Webster’s New World College Dictionary*. Webster’s New World, Boston, 5th edition.


Appendix A

Notations and Glossary

A.1 Operators

- Vector dot product
\times Vector cross product
\|v\| Euclidean norm of a vector v
|x| Absolute value of a real number x
|z| Complex modulus of a complex number z
\bar{z} Complex conjugate of a complex number z
\text{Re} z Real part of a complex number z
* Convolution
\circ Function composition
\partial/\partial x Partial derivative
\partial/\partial \mathbf{n} Normal derivative
\nabla \times Curl operator
\nabla \cdot Divergence operator
\nabla Gradient
\nabla^2 Laplacian
\mathcal{F}\{f\} Fourier transform of f
\mathcal{F}_x\{f\} Fourier transform of f with respect to x
\mathbb{E}[X] Expectation of a random variable X
\mathbb{E}_X[Y] Expectation, treating only X as a random variable
\text{rot}_\beta \mathbf{x} Rotate a two-dimensional vector \mathbf{x} counterclockwise by the angle \beta
\text{prj}_f Vertical projection, \( x \mapsto \int_y f(x, y) dy \)
\text{slc}_f Horizontal slicing, \( x \mapsto f(x, 0) \)
\text{prj}_\beta f Rotated projection, same as \text{prj}(f \circ \text{rot}_{-\beta})
\text{slc}_\beta f Rotated slicing, same as \text{slc}(f \circ \text{rot}_{-\beta})
\text{rvl}_f Revolve a univariate function f defined on \([0, \infty)\) into a bivariate function defined on \(\mathbb{R}^2\)
\mathcal{R}\{f\} ACF of a function f
$S\{f\}$ \hspace{1cm} PSD of a function $f$

$R_{ff}$ \hspace{1cm} Autocorrelation of a wide-sense stationary stochastic process $f$

$S_{ff}$ \hspace{1cm} Power spectrum of a wide-sense stationary stochastic process $f$

$\text{clerp}_\beta(f)$ \hspace{1cm} Circumferential linear interpolation

### A.2 Notation

- **$v$**: Vector quantity
- **$\hat{v}$**: Unit vector in the direction of $v$
- **$[x, y, z]^T$**: Column vector
- **$[x; y; z], [x; z]$**: Column vector, formed by the vertical concatenation of column vectors or scalars
- **$\mathbb{R}$**: The set of real numbers
- **$\mathbb{C}$**: The set of complex numbers
- **$0$**: The origin

### Physical Optics and Kirchhoff Approximation

- **$E$**: Electric field
- **$H$**: Magnetic field
- **$\mu$**: Permeability
- **$\epsilon$**: Permittivity
- **$u(r, t)$**: Time-dependent scalar wave function
- **$v$**: Speed of wave propagation
- **$f$**: Frequency of a monochromatic wave
- **$A(r)$**: Amplitude of a monochromatic wave at position $r$
- **$\phi(r)$**: Phase of a monochromatic wave at position $r$
- **$U(r)$**: Complex amplitude at position $r$ of a generic monochromatic wave; Also refers specifically to the wave field resulting from rough surface scattering
- **$k$**: Wave number
- **$\hat{k}$**: Propagation direction of a plane wave
- **$k$**: Wave vector of monochromatic plane wave
- **$\psi(r)$**: Complex amplitude of a monochromatic plane wave at position $r$
- **$\psi_1, \psi_2$**: Incident and reflected monochromatic plane waves
- **$U_1, U_2$**: Complex amplitude at $0$ of the incident and reflected monochromatic plane waves
- **$k_1, k_2$**: Wave vector of the incident and reflected monochromatic plane waves
- **$\hat{n}$**: Unit surface normal
- **$E_1, E_2$**: Complex amplitude of the incident and reflected electric fields
**Cartesian bases for representing the incident and reflected electric fields**

\( f_1, f_2 \)

**Perpendicular polarized component of the incident and reflected electric fields**

\( E_{\perp,1}, E_{\perp,2} \)

**Parallel polarized component of the incident and reflected electric fields**

\( E_{\parallel,1}, E_{\parallel,2} \)

**Perpendicular and parallel polarized Fresnel coefficients for incident angle \( \theta \)**

\( F_{\perp}(\theta), F_{\parallel}(\theta) \)

**Viewpoint**

\( v \)

**Distance between the viewpoint \( v \) and \( 0 \)**

\( R \)

**Viewing direction**

\( \hat{v} \)

**Surface patch used in the Kirchhoff approximation**

\( S \)

**A point on the surface patch**

\( s \)

**Reflected monochromatic plane wave at \( s \)**

\( \psi_{2,s} \)

**Unit surface normal at \( s \)**

\( \mathbf{n}(s) \)

**Average surface normal**

\( \mathbf{n}_s \)

**Auxiliary function for applying Green’s second identity**

\( G \)

**Fresnel coefficient derived from \( \mathbf{n}_s \), can represent either polarization choice**

\( F \)

**Heightfield defining the surface geometry**

\( z \)

**Surface modulation function**

\( m \)

**Light direction, equals \( -\hat{k}_1 \)**

\( \hat{l} \)

**Half vector, equals \( (\hat{v} + \hat{l})/2 \)**

\( \mathbf{h} \)

**2D vector formed by the X and Y components of \( \mathbf{h} \)**

\( \mathbf{h}_{xy} \)

**Z component of \( \mathbf{h} \)**

\( h_z \)

**Incident and reflection angle**

\( \theta_i, \theta_r \)

---

**Reflectance Prediction from Surface Profiles**

\( z_j \)

A randomly displaced, one-dimensional horizontal slice of the surface heightfield \( z \)

\( d_j \)

Random displacement that appears in data captured by AFM

\( P \)

MPSD, power spectral density of surface modulation function

\( p_j \)

One-dimensional MPSD computed from \( z_j \)

\( p \)

Estimated MPSD projection

\( W \)

Dimension of the AFM scanning window

\( N \)

The number of one-dimensional height profiles acquired

\( \Phi(\hat{I}) \)

Radiant flux recorded by the gonioreflectometer for light direction \( \hat{I} \)

**Shaped Microfacet Model**

\( f \)

Identifies a facet
\( \mathbf{c} \) Lateral (XY) location of facet centroid
\( z \) Vertical (X) location of facet centroid
\( \Omega \) Shape function
\( \theta \) Slant angle
\( \alpha \) Azimuthal angle
\( (c_f, z_f, \Omega_f, \theta_f, \alpha_f) \) Parameters of a particular facet \( f \)
\( \mathbf{s}_f \) Slant vector of facet \( f \)
\( \mathbf{s} \) Function mapping from a pair of slant and azimuthal angles to the slant vector
\( h \) Scalar introduced to simplify notation, equals \( -2kh_z \)
\( \mathcal{S} \) Fixed probability space of surfaces consisting of \( M \) shaped facets
\( M \) Number of facets in each surface in \( \mathcal{S} \)
\( \mathbb{E}[P] \) Expected MPSD of surfaces in \( \mathcal{S} \)
\( W_{\theta, \alpha} \) Expected PSD of shape function conditioned on facet orientation \( (\theta, \alpha) \)
\{\( B_j \}\} Shape basis for representing \( W_{\theta, \alpha} \)
\{\( \Theta_k \}\} Slant basis
\{\( A_l \}\} Azimuthal basis
\{\( \Psi_{j,k,l} \}\} Dictionary elements for representing \( \mathbb{E}[P] \)
\( c_{j,k,l} \) Linear combination coefficient for dictionary element \( \Psi_{j,k,l} \) when representing \( \mathbb{E}[P] \)
\( \text{cyl} \) Cylinder function
\( \text{somb} \) Sombrero function

**Optimization-Based MPSD Estimation**

\{\( \Omega_j \}\} Disks of different sizes, whose PSDs form the shape basis
\( r_{\text{max}} \) Estimated upper bound on facet radius
\( N_{\Omega}, N_{\theta}, N_{\alpha} \) Sizes of the shape basis, slant basis, and azimuthal basis
\( \mathcal{E}_s \) Smoothness regularization functional
\( \kappa_s \) Smoothness regularization weight
\( \mathcal{E}_a \) Anisotropy regularization functional
\( \kappa_a \) Anisotropy regularization weight
\( p_{\beta}^{h_z} \) Estimated MPSD projection along azimuthal direction \( \beta \), dependent on \( h_z \)
\( \Psi_{j,k,l}^{h_z} \) MPSD dictionary element, dependent on \( h_z \)

**Interpolation-Based MACF Estimation**

\( \Pi_W \) Windowing function
\( \Lambda_W \) Bilinear hat function
\( m_s \) Surface modulation function viewed as an idealized wide-sense stationary stochastic process
Surface modulation function computed from a single AFM scan, corresponding to a windowed realization of $m_*$. 

$m_Y$ A random horizontal slice of $m$

$\{\beta_k\}$ Azimuthal directions along which one-dimensional surface profiles have been measured

$\{r_k\}$ MACF slices estimated from one-dimensional profiles along directions $\{\beta_k\}$

$\{s_k\}$ MPSD projections estimated from one-dimensional profiles along directions $\{\beta_k\}$

$l$ Default bandlimit for low-pass filtering MPSD

A.3 Glossary

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRDF</td>
<td>Bidirectional reflectance distribution function</td>
</tr>
<tr>
<td>AFM</td>
<td>Atomic force microscopy</td>
</tr>
<tr>
<td>WLI</td>
<td>White-light interferometry</td>
</tr>
<tr>
<td>CLSM</td>
<td>Confocal laser scanning microscopy</td>
</tr>
<tr>
<td>WSS</td>
<td>Wide-sense stationary</td>
</tr>
<tr>
<td>PSD</td>
<td>Power spectral density</td>
</tr>
<tr>
<td>ACF</td>
<td>Autocorrelation function</td>
</tr>
<tr>
<td>MPSD</td>
<td>Modulated power spectral density</td>
</tr>
<tr>
<td>MACF</td>
<td>Modulated autocorrelation function</td>
</tr>
</tbody>
</table>