Survey of Models for Acquiring the Optical Properties of Translucent Materials


1Technical University of Denmark
{jerf, sorgre}@dtu.dk
2DFM A/S, Danish National Metrology Institute
{saJ, jsm, peh}@dfm.dk
3KU Leuven, Belgium
{antonio.correia, youri.meuret}@kuleuven.be
4RISE - Research Institutes of Sweden
li.yang@ri.se

Abstract
The outset of realistic rendering is a desire to reproduce the appearance of the real world. Rendering techniques therefore operate at a scale corresponding to the size of objects that we observe with our naked eyes. At the same time, rendering techniques must be able to deal with objects of nearly arbitrary shapes and materials. These requirements lead to techniques that oftentimes leave the task of setting the optical properties of the materials to the user. Matching the appearance of real objects by manual adjustment of optical properties is however nearly impossible. We can render objects with a plausible appearance in this way but cannot compare the appearance of a manufactured item to that of its digital twin. This is especially true in the case of translucent objects, where we need more than a goniometric measurement of the optical properties. In this survey, we provide an overview of forward and inverse models for acquiring the optical properties of translucent materials. We map out the efforts in graphics research in this area and describe techniques available in related fields. Our objective is to provide a better understanding of the tools currently available for appearance specification when it comes to digital representations of real translucent objects.

CCS Concepts
• Computing methodologies → Reflectance modeling; Appearance and texture representations;
1. Introduction

The appearance of an object is a product of its geometrical shape and the optical properties of its constituent materials. We consider materials different when they are clearly separated by an interface. We can then segment a given object into different parts each with a set of optical properties that are functions (possibly spatially varying) of the wavelength of the light and also of the microgeometry of the material. To get a proper understanding of material appearance, we must therefore consider optical properties at different scales. This is illustrated in Figure 1. In this work, we provide an overview of models for acquiring optical properties and identify the different scales they can operate at. We focus on translucent materials and include both surface and subsurface scattering.

Conventional rendering techniques solve the rendering equation, which consists of a sum of emitted and reflected radiance [Nic65, Kaj86]. This equation involves the transfer of radiant flux between surfaces, and light-material interaction is described by a bidirectional scattering distribution function (BSDF). This works well for opaque and transparent objects. For light scattering volumes (participating media), however, we need to solve the radiative transfer equation (RTE) [Cha50, KV84]. The RTE models absorption, scattering, and emission of light as we move along a path traversing the volume. Interestingly, if we consider a light scattering volume with a refractive interface, we can use the rendering equation to specify boundary conditions for the radiative transfer equation [Arv93]. In turn, this enables us to render a medium that exhibits both surface and subsurface scattering using path tracing techniques [PKK00, RSK08], and path tracing is currently the method of choice in production rendering [FHP18].

Translucency means that light not only scatters at the surface but also propagates from one point of incidence through the interior of the object and emerges at other positions. Since a translucent object exhibits both surface and subsurface scattering, it is challenging to fully describe its optical properties. While surveys are available on models for light propagation in scattering media [CPP05, NGH18], the acquisition of optical properties (BSDF, absorption, scattering, emission) is rarely discussed. This leaves a hole in cases where our objective is to render the appearance of real world objects. Regardless of the virtues of the rendering technique and the light propagation model, a rendering can only be as accurate as the input optical properties. This survey is focused specifically on acquisition of optical properties of translucent objects. We discuss models that are useful for computing or deriving scattering functions and include inverse techniques useful for estimating apparent or intrinsic optical properties of translucent materials. We find this to be an important endeavour if we are to increase the relevance of graphics in the context of product appearance specification and comparison of photographed appearance with renderings of a digital twin (e.g. for quality assessment).

The fact that optical properties exist at different scales (Figure 1) is important for identifying the differences between models for acquiring optical properties. We can use light scattering simulation to go from optical properties at a microscopic scale to observable scattering at a more macroscopic scale. Conversely, we can use inverse models to go from observation to sample properties, where different models may be better suited to estimate different sample properties (surface or volume). Section 2 provides an overview of graphics research on forward and inverse models. This has multiple uses: tracking the development of the field, finding previous work on passage from one scale to another, identifying uncharted research territory. Sections 3–4 then provide an explanation of the different scales and how they connect. This serves to introduce terminology and more theoretical aspects. In the remaining sections, we cover techniques for acquisition of optical properties at different scales. We include discussion of techniques available in other research fields to inspire future development of graphics techniques.

Two books on material appearance modelling [DRS07, DLG13] seem to be the work most closely related to ours. These books are however broader in scope and make no attempt at relating graphics research to models often employed in other areas.

2. Model Classification

We find that a chronologically sorted list of references with markers for different scales and approximations provides a useful overview. Our version of such a list is in Table 1. We divide the microscopic scale into nano/micro and micro/milli. The nano/micro scale is for models that consider explicit microgeometry, like a patch of geometry observed in a microscope image or the shape of a light scattering fibre or particle. The micro/milli scale is for models considering microfacet normal distributions or particle size distributions. We divide the macroscopic scale into two categories: one for the bidirectional scattering-surface reflectance distribution function (BSSRDF) and one for the more approximate bidirectional reflectance/transmittance function (BRDF/BTDF). The point of making this distinction is that many models do not account for the fact that light may be incident at one surface point but emerge at another. The BSSRDF models take this into account while the BRDF/BTDF models do not (Figure 2). The BSDF mentioned previously is a function including both reflectance and transmittance (both BRDF and BTDF).

The marker system described in Table 1 is used to distinguish between the different types of research on acquiring optical properties. We have a marker for theoretical work and three for experimental work, and we distinguish between models operating with colour vectors and models operating with wavelength dependency. We use other markers too. In cases of brushed surfaces or fibrous materials, the light scattering depends on the orientation of the object. This is called anisotropy. Models that can account for this are referred to as anisotropic, while models that are invariant to object orientation are called isotropic. Some models assume a homogeneous material. The material is then the same in all surface points and the same in all interior points, whereas objects with spatially varying optical properties are referred to as heterogeneous. Finally, subsurface scattering is sometimes considered diffuse, which means that it does not depend on the directions of incidence and emergence of the light. We include a marker showing whether a model considers subsurface scattering to be diffuse or dependent on the directions of incident and/or emergent light.

We use the seminal work of Torrance and Sparrow [TS66, TS67] as a starting point. One reason for this choice is that they include magnesium oxide ceramic, which is a translucent material. The
Table 1: Overview of graphics models for acquiring the optical properties of translucent materials. We use the listed markers for classification of the research. Sections 2–3 explain the scales in the four columns with markers. An online version of the table is available at https://people.compute.dtu.dk/jerf/cg-opt-props.html.

<table>
<thead>
<tr>
<th>Paper</th>
<th>nano/micro</th>
<th>micro/milli</th>
<th>BSSRDF</th>
<th>BRDF/BTDF</th>
</tr>
</thead>
<tbody>
<tr>
<td>[TS66]</td>
<td>x2λi\→</td>
<td>tci\→</td>
<td>tc*i\→</td>
<td>←x3ci\→</td>
</tr>
<tr>
<td>[TS67]</td>
<td>tci\→</td>
<td>tci\→</td>
<td>tc*i\→</td>
<td>←tx2ci\→</td>
</tr>
<tr>
<td>[Bli77]</td>
<td>tci\→</td>
<td>tci\→</td>
<td>tc*i\→</td>
<td>←x2c*i\→</td>
</tr>
<tr>
<td>[CT81]</td>
<td>tci\→</td>
<td>tci\→</td>
<td>tc*i\→</td>
<td>←x3c*i\→</td>
</tr>
<tr>
<td>[Bl82]</td>
<td>tci\→</td>
<td>tci\→</td>
<td>tc*i\→</td>
<td>←x2c*i\→</td>
</tr>
<tr>
<td>[Kj84]</td>
<td>tci\→</td>
<td>tci\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
<tr>
<td>[CM87]</td>
<td>tci\→</td>
<td>tci\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
<tr>
<td>[KK89]</td>
<td>tci\→</td>
<td>tci\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
<tr>
<td>[HTG91]</td>
<td>tci\→</td>
<td>tci\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
<tr>
<td>[WAT92]</td>
<td>tci\→</td>
<td>tci\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
<tr>
<td>[War92]</td>
<td>tci\→</td>
<td>tci\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
<tr>
<td>[HM92]</td>
<td>tci\→</td>
<td>tci\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
<tr>
<td>[Hk93]</td>
<td>tci\→</td>
<td>tci\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
<tr>
<td>[GMN94]</td>
<td>tci\→</td>
<td>tci\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
<tr>
<td>[Ca96]</td>
<td>tci\→</td>
<td>tci\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
<tr>
<td>[NIDN97]</td>
<td>tci\→</td>
<td>tci\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
<tr>
<td>[Br97]</td>
<td>tci\→</td>
<td>tci\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
<tr>
<td>[JW97]</td>
<td>tci\→</td>
<td>tci\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
<tr>
<td>[DVGNK99]</td>
<td>x2ca\→</td>
<td>tci\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
<tr>
<td>[Mwl99]</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
<tr>
<td>[Sta99]</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
<tr>
<td>[DEJ99]</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
<tr>
<td>[Ph00]</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
<tr>
<td>[AP00]</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
<tr>
<td>[DHT*00]</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
<tr>
<td>[Sta01]</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
<tr>
<td>[Ek01]</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
<tr>
<td>[Lk01]</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
<tr>
<td>[Ly01]</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
<tr>
<td>[JML01]</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
<tr>
<td>[MPB03]</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
<tr>
<td>[Mic03]</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
<tr>
<td>[GL7*04]</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
<tr>
<td>[Ed04]</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
<tr>
<td>[Kk05]</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
<tr>
<td>[Nd05]</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
<tr>
<td>[Ttw*05]</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
<tr>
<td>[Dj05]</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
<tr>
<td>[Dj06]</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
<tr>
<td>[Pvb*06]</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
<tr>
<td>[Ng06]</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>tc*i\→</td>
<td>←tx2c*i\→</td>
</tr>
</tbody>
</table>

© 2020 The Author(s)
Computer Graphics Forum © 2020 The Eurographics Association and John Wiley & Sons Ltd.
Torrance-Sparrow BRDF model [TS67] is based on a distribution of surface microfacets consisting of specular v-grooves. This is an example of a forward model going from a normal distribution (nano/micro scale) to a macroscopic BRDF. In early work such as that of Torrance and Sparrow, subsurface scattering is approximated by a diffuse reflectance term in the BRDF [Lea79]. Scattering by particles was in graphics first considered by Blinn [Bli82], and subsurface scattering was first considered by Hanrahan and Krueger [HK93]. These efforts resulted in more advanced BSDFs.

Some models provide optical properties to be used in simulation of volumetric light transport. An early example is the ray tracing of volume densities by Kajiya and Von Herzen [KV84]. However, many such models were specifically developed for rendering of optically thin volumes without a refractive interface (participating media: clouds, smoke, atmosphere, etc.). In our table, these models have markers at the micro/milli scale without an arrow pointing to a more macroscopic scale.

While Hanrahan and Krueger [HK93] described simulation of subsurface scattering, they still ended up with a local BSDF model. The first graphics model to include the subsurface scattering effects resulting from different points of incidence and emergence (BSSRDF effects) is seemingly that of Dorsey et al. [DEI99]. These authors acquire optical properties of stone by using procedural models to generate colours and densities in volumetric texture slabs. The slabs are in turn modified by simulation of weathering effects.

An analytic BSSRDF model by Farrell et al. [FPW92] was introduced to graphics by Jensen et al. [JMLH01]. This model is often referred to as the standard dipole. Analytical models ease the development of inverse techniques. Given a flat, homogeneous sample of a translucent material, the work of Jensen et al. [JMLH01] includes an inverse technique for photographic estimation of its apparent optical properties. This was extended to work for heterogeneous materials by Peers et al. [PVBM06]. More recently, a solution for the diffusion equation including the directionality of the incident illumination [MSG05] was used for deriving an analytic directional dipole BSSRDF model [FHK14]. This also led to an inverse technique [ZIK17]. Inverse techniques have also been based on single scattering [NGD06] or full Monte Carlo ray tracing of the scattering volume [GZB13]. We place the markers for experimental techniques based on photographs at the BRDF/BTDF level. On the other hand, we place markers for techniques based on micro CT scans or profilometry at the nano/micro scale.

Many BRDF models are available in the literature, a survey is available from Guarnera et al. [GGG16]. We cannot include all of them here and keep our focus on models with relevance for translucent materials and a relation to more microscopic scales. Examples of models starting at the nano/micro scale are those for computing a BRDF based on a patch of microgeometry [Kaj85, CMS87, WAT92, Sta99]. An example of a corresponding experimental technique based on a patch of microgeometry is the work of Dong et al. [DWMG15]. They acquire normal distributions from microscope depth images (profilometry). Similarly, models for computing intrinsic optical properties based on the scattering of a plane wave of light by a spherical particle (Lorenz-Mie theory) also start from the nano/micro scale [Cal96, JW97, FCJ07]. When models include wave effects like diffraction, they become wavelength dependent and are marked by a $\lambda$.

The reference descriptions in this section serve as a help for interpreting the information provided in Table 1. When exploring the table, we recommend use of a pdf reader with a shortcut for going back after following a link (e.g. the alt and left arrow shortcut in Acrobat Reader). The list of references we provide is non-exhaustive. We have selected references that we find useful in terms of getting an overview of graphics techniques for determining the optical properties of translucent materials. We encourage the reader to use the marker system for categorising papers appearing in related fields such as optics and computer vision. For an overview of more specific areas, we refer the reader to other surveys. A significant portion of the relevant research is concerned with the appearance of human faces and fibrous materials (hair, fabrics). We include some work on skin appearance [HK93, MWL09, DHT00, Sta01, JMLH01, BKK05, DJ05, DJ06, WMP06, GHP08, DWD08, IGAJ15], while more complete surveys are available on facial appearance capture [KRP15] and human tissue appearance [NM19]. We also include some work on the optical properties of hair and fibres [KK89, MJ03, ZW07, MM08, JMM09, ZRL09, KZ11, ZMB11, IM12, SBDD13, KSZ15, ACG17, YSJ17]. Again, a more complete survey is available on fabric appearance reproduction [CLM19].

With this overview of graphics models, we dive into the details of optical properties and how to determine them by means of forward and inverse models. Optical properties describe light-matter interaction, that is, scattering and absorption of light. In the next section, we look into the relations between light scattering at different scales. The employed nomenclature is listed in Table 2.

3. Multi-Scale Modeling of the Scattering of Light

Scattering functions describe light-matter interactions at the most macroscopic physical level. Given a number of arguments, a scattering function is a factor of proportionality between incident and scattered light. To derive such a function, we need to go to a smaller scale. At the smallest scale, everything is quantum particles. Photons have a probability of being found in one place or another and a probability of being absorbed or emitted by an atomic system. These probabilities are the same as the amplitude of properly normalized electromagnetic waves [Fey98]. Considering the mean effects of a continuum of photons, the wave amplitudes transition to the classical field vectors. Conceptually, a scattering function then describes the different outcomes of photon-electron interactions. In X-ray physics, quantum effects are sometimes important [Men18]. In the visible part of the spectrum, we usually need to know about photons only to estimate emission spectra or refractive indices. To derive scattering functions, we need a passage from microscopic scales to macroscopic scale.

The work of Feynman [FLS64, FLS65, Fey98] is an excellent source on the relations between quantum and wave theories. Mishchenko et al. [MTL06] provide a passage from Maxwell’s equations to the vector radiative transfer equation (where vectors represent polarization). However, a passage to radiative transfer theory involves a number of ambiguities [Mis13], which we will...
Instead of counting photons, which are extremely numerous in the visible part of the spectrum, we may consider electromagnetic waves. The number of photons per second is proportional to the energy flux of the electromagnetic field, in which the wave frequency corresponds to the photon energy. The material properties entering into macroscopic versions of Maxwell’s equations are permittivity, permeability, and conductivity. A nice way to summarize these material properties is by means of the (complex) index of refraction \( n \), extinction coefficient \( \sigma_t \), and absorption coefficient \( \sigma_a \). Given this material property, which depends on the wavelength \( \lambda \), we can derive how electromagnetic waves scatter at interfaces between media of different indices of refraction. In this regime of wave theory, all geometric features around the size of a wavelength should be defined. We can use this scale for computing the scattering at a position \( x \) in a more macroscopic model. The microgeometry used for computing such a scattering function then represents a differential area \( dA \) or volume \( dV \) around a position \( x \).

The scattering functions that we can compute by considering electromagnetic wave propagation are extremely useful. Consider a plane wave incident on a surface between two half-space media of different indices of refraction \( n_i \) and \( n_f \). By requiring continuity across the interface of the components of the electric field vector \( \mathbf{E} \) and the magnetic vector \( \mathbf{H} \) that are tangent to the surface, we can derive the laws of reflection and refraction and Fresnel’s equations [Bel67]. This provides us with the bidirectional scattering distribution function (BSDF) of a perfectly smooth interface. We can also consider a plane wave in a medium of refractive index \( n_{\text{med}} \) on a spherical particle of refractive index \( n_p \). This gives rise to a spherical wave that provides a directional distribution of the scattered light (a phase function) [BH83]. These functions are essential, but we need a yet more macroscopic scale to describe the scattering function of a translucent material.

At the next scale, we consider an ensemble of surface micro-facets or an ensemble of small particles in a volume. Each micro-facet has a BSDF \( f_{mn} \) [TS67, Bli77, WMLT07] and each particle a phase function from the microscale \( p_m \) [Bli82, Cal96, FCJ07]. Based on an ensemble, defined by a distribution of microfacet normals or a size distribution of small particles, we can derive or compute the bulk scattering properties of a material at a more macroscopic scale. The scale is then significantly larger than the wavelength of visible light, and it then makes sense to use geometrical optics and thus ray tracing as the model of light propagation [Wat92, MPH+15]. To keep track of the flow of energy, we turn to radiometry [Nic63] and radiative transfer theory [Cha50]. At this scale, we can use path tracing [Rus88, PH00, FCIJ12, PJJH16] to compute the appearance or the global scattering effect of a translucent object.

Path tracing operates locally. We probabilistically consider scattering and absorption events as we trace a path through an object or between object surfaces. Macroscopic scattering functions represent the global scattering effect. Given the surface \( X \) of a translucent object and its optical properties, a macroscopic scattering function represents the net effect of all in-surface and sub-surface scattering. The theoretical link between the local and the global formulation is provided by Preissendorfer [Pre65]. Venable and Hsia [VH74] presented a technical report on the measurement of such a complicated scattering function, while Nicodemus et al. [NRH77] provided a seminal report defining various simplified versions of the scattering function, namely the well-known bidirectional distribution functions.

### 4. Local to Global Models and Optical Properties

Consider a point \( x \) along a ray of light traveling in the direction \( \mathbf{u} \). We use boldface to denote position vectors and arrow overline to denote a direction vector of unit length. In the local formulation, we can consider the change in radiance \( \mathbf{L} \) as we take an infinitesimal step along the ray. This change is written mathematically as an

<table>
<thead>
<tr>
<th>BSDF</th>
<th>Bidirectional Scattering Distribution Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSSRDF</td>
<td>Bidirectional Scattering-Surface Reflectance Distribution Function</td>
</tr>
<tr>
<td>BRDF</td>
<td>Bidirectional Reflectance Distribution Function</td>
</tr>
<tr>
<td>BTDF</td>
<td>Bidirectional Transmittance Distribution Function</td>
</tr>
<tr>
<td>( f_{mn} )</td>
<td>microfacet BSDF</td>
</tr>
<tr>
<td>( f_s )</td>
<td>BSDF</td>
</tr>
<tr>
<td>( S )</td>
<td>BSSRDF</td>
</tr>
<tr>
<td>( f_r )</td>
<td>BRDF</td>
</tr>
<tr>
<td>( f_t )</td>
<td>BTDF</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>wavelength</td>
</tr>
<tr>
<td>( n_{xx} )</td>
<td>(complex) index of refraction of medium ( xx )</td>
</tr>
<tr>
<td>( \Phi )</td>
<td>radiant flux</td>
</tr>
<tr>
<td>( L )</td>
<td>radiance</td>
</tr>
<tr>
<td>( L_\chi )</td>
<td>spectral radiance</td>
</tr>
<tr>
<td>( E )</td>
<td>irradiance</td>
</tr>
<tr>
<td>( \sigma_g )</td>
<td>scattering coefficient</td>
</tr>
<tr>
<td>( \sigma_a )</td>
<td>absorption coefficient</td>
</tr>
<tr>
<td>( \sigma_t )</td>
<td>extinction coefficient (( \sigma_t = \sigma_s + \sigma_a ))</td>
</tr>
<tr>
<td>( g )</td>
<td>asymmetry parameter</td>
</tr>
<tr>
<td>( p )</td>
<td>phase function</td>
</tr>
<tr>
<td>( p_m )</td>
<td>particle phase function</td>
</tr>
<tr>
<td>( C_s )</td>
<td>particle scattering cross section</td>
</tr>
<tr>
<td>( C_a )</td>
<td>particle absorption cross section</td>
</tr>
<tr>
<td>( \tau )</td>
<td>optical depth</td>
</tr>
<tr>
<td>( R_g )</td>
<td>Root-Mean-Square (RMS) roughness</td>
</tr>
<tr>
<td>( R_{g,q} )</td>
<td>RMS slope</td>
</tr>
<tr>
<td>( \bar{R}_{g,\text{rel}} )</td>
<td>band limited RMS roughness</td>
</tr>
<tr>
<td>( R_{g,\text{tot}} )</td>
<td>total RMS roughness</td>
</tr>
<tr>
<td>( \mathcal{F} )</td>
<td>Fourier transform</td>
</tr>
<tr>
<td>BTF</td>
<td>Bidirectional Texture Function</td>
</tr>
<tr>
<td>RTE</td>
<td>Radiative Transfer Equation</td>
</tr>
<tr>
<td>KM</td>
<td>Kubelka-Munk</td>
</tr>
<tr>
<td>RCWA</td>
<td>Rigorous Coupled-Wave Analysis</td>
</tr>
<tr>
<td>FEM</td>
<td>Finite Element Method</td>
</tr>
<tr>
<td>FDTD</td>
<td>Finite Difference Time Domain</td>
</tr>
<tr>
<td>RR</td>
<td>Rayleigh-Rice</td>
</tr>
<tr>
<td>ACF</td>
<td>Auto-Correlation Function</td>
</tr>
<tr>
<td>GHS</td>
<td>Generalised Harvey Shack</td>
</tr>
<tr>
<td>PSD</td>
<td>Power Spectral Density</td>
</tr>
</tbody>
</table>
integro-differential equation, where the derivative is with respect to the distance \( s \) travelled along the ray. Writing this in terms of a directional derivative, we have the radiative transfer equation in its differential form [Cha50]:

\[
(\bar{\omega} \cdot \nabla) L(x, \bar{\omega}) = -\left( \sigma_s(x) + \sigma_t(x) \right) L(x, \bar{\omega}) + \ell_{\text{r}}(x, \bar{\omega}) + \sigma_t(x) \int_{4\pi} p(x, \bar{\omega}', \bar{\omega}) L(x, \bar{\omega}') \, d\omega',
\]

where \( \ell_{\text{r}} \) is radiant emitted in the direction of interest per unit distance travelled through the medium. The scattering coefficient \( \sigma_s \) denotes the amount of scattering per unit distance travelled through the medium. The absorption coefficient \( \sigma_t \) is similar but denotes the amount of absorption. In the case of anisotropic materials, these coefficients depend on the direction of light propagation \( \bar{\omega} \) [JAM+10].

The first term of the RTE then describes the loss of radiant due to absorption and out-scattering. The second term describes emission of radiant per unit distance travelled through the medium and the third term describes in-scattering of radiant per unit distance. The integral is over all directions \( \bar{\omega}' \), that is, all \( 4\pi \) solid angles in the unit sphere, and \( p \) is the phase function. The phase function is thus used here to find the part of the radiance incident from all directions that is scattered into the direction of interest \( \bar{\omega} \). The phase function can be isotropic or anisotropic. For real materials, the latter is usually the case, and this is referred to as scattering anisotropy (not to be confused with the material anisotropy discussed above).

Radiance is defined in terms of radiant flux \( \Phi \) (radiant energy per unit time) at an element of surface area \( dA \). The energy flows in a direction coned described by an element of solid angle \( d\omega \), and the radiance is the part of the energy flow that projects to the area \( dA \). Mathematically, radiance is defined by [Nic63]

\[
L = \frac{\partial^2 \Phi}{\partial A_\perp \, d\omega} = \frac{\partial^2 \Phi}{\cos \theta \, dA \, d\omega},
\]

where \( dA_\perp \) is projected area and \( \theta \) is the angle between the surface normal \( \bar{n} \) of \( dA \) and the direction \( \bar{\omega} \) of \( d\omega \). This definition of radiance works well when light scatters between sample elements. However, if light scatters in a volume, there is no surface normal to describe the projected area. As the element of area, we instead use the total scattering cross section of the particles that scatter light in an element of volume \( dV \).

The scattering coefficient \( \sigma_s \) is the total scattering cross section per unit volume. For a group of particles scattering light independently, we have [vdH57]

\[
\sigma_s = \int_0^\infty C_s(r) \, dN(r),
\]

where \( r \) is the size (radius) while \( C_s \) is the scattering cross section of a light scattering particle, and \( dN \) is the number density. This means that we can define the total scattering cross section by [SH02]

\[
dA_s = \sigma_s \, dV
\]

and this is used in place of the element of projected area in the definition of radiance (\( dA_\perp \) becomes \( dA_s \)) in order to describe scattered radiance in a volume. The scattering cross section of a particle as well as its phase function \( p_{\text{sc}} \) and absorption cross section \( C_a \) can be calculated using Maxwell’s equations, where Poynting’s vector provides the direction and magnitude of the energy in the field. We thereby have a link in these optical properties to light scattering at a more microscopic scale.

In the case of a translucent object, we are not so interested in the local scattering events along a ray of light traversing the medium. We would rather like to determine the light emerging at the surface of the object. The emergent radiant flux is the observable quantity. We thus introduce an object boundary \( X \), which would commonly be defined by a triangle mesh. At a surface area element \( dA \), the amount of radiant flux incident per unit area is called irradiance \( (E = \Phi/\,dA) \). We can then specify a function \( f_j \) that describes the factor of proportionality between an element of outgoing radiance \( dL_o \) and an element of irradiance \( dE \) [BDW81]:

\[
f_j(x, \bar{\omega}_i, \bar{\omega}_o) = \frac{dL_o(x, \bar{\omega}_o)}{dE(x, \bar{\omega}_o)},
\]

where \( dE \) is a differential element of irradiance incident from a solid angle that is the limit is only one direction \( \bar{\omega}_o \). Note that \( f_j \) is a mesoscopic BSDF that (like \( \sigma_s, \sigma_a \)) can be calculated for a given surface microgeometry by resorting to scattering at a more microscopic scale. The boundary conditions for a translucent object with surface points \( x_o \in X \) are then

\[
L_o(x_o, \bar{\omega}_o) = \int_{4\pi} f_j(x_o, \bar{\omega}_i, \bar{\omega}_o) \, dE(x_o, \bar{\omega}_i)
\]

\[
= \int_{4\pi} f_j(x_o, \bar{\omega}_i, \bar{\omega}_o) \int_{4\pi} p(x_o, \bar{\omega}_i', \bar{\omega}_i) L_i(x_o, \bar{\omega}_i') \, d\omega_i' \, e^{-\tau(x',s)} \, d\omega_i.
\]

We have now reached a level of theoretical development where we can start deriving the macroscopic scattering functions known as bidirectional distribution functions. This is done by solving the RTE with given boundary conditions. Taking a non-emissive medium \( (\ell_{\text{r}} = 0) \) and reformulating the differential derivative in Eq. (1) as a derivative with respect to distance \( s \) traveled along the ray \( x(s) = x_o + s \bar{\omega} \), we find the RTE in integral form [Cha50]

\[
L(x, \bar{\omega}) = L(0, \bar{\omega}) e^{-\tau(0,s)} + \int_0^s \sigma_s(s') \int_{4\pi} p(s', \bar{\omega}', \bar{\omega}) L(s', \bar{\omega}') \, d\omega' \, e^{-\tau(s',s)} \, ds' \]

where \( \tau \) is optical depth:

\[
\tau(s',s) = \int_{s'}^s \sigma_a(t) \, dt = \int_{s'}^s (\sigma_a(t) + \sigma_s(t)) \, dt,
\]

and \( \sigma_a = \sigma_s + \sigma_a \) is referred to as the extinction coefficient.

The first term in Eq. (7) is called the beam-transmitted radiance [Pre65] or the reduced intensity term [Ish78]. This is the radiance that passes through the interior of the medium without being scattered or absorbed. The second term is the path radiance [Pre65] or the reduced intensity term [Ish78]. For radiance entering the medium \( L(0, \bar{\omega}) = L_o(x_o, \bar{\omega}) \), let us make an operator \( S_0^{(j)} \) that returns the reduced intensity term, and an integral operator \( S_1 \) returning the diffuse intensity term for a given ray. We can now repeatedly apply \( S_1 \) to get the effect of multiple scattering events and construct an operator that represents the net effect of the subsurface scattering in the medium [Pre65]

\[
S_0 + \sum_{j=1}^\infty S_1^j = \sum_{j=0}^\infty S_1^j.
\]
If we similarly construct an integral operator \( F_j \) that given incident radiance \( L_i \) uses the boundary conditions in Eq. (6) to obtain the radiance \( L_o \) scattered from a surface element, we can write

\[
S = F_j \sum_{j=0}^\infty S_j/F_j.
\]

The operator then applies to radiance incident in positions \( x_i \) everywhere on the surface \( X \) of an object and returns reflected radiance, so that \( L_r = L_i S \). Following Preisendorfer [Pre65], we can then define a function based on this scattering operator

\[
S(X; x_i, \tilde{\omega}; x_o, \tilde{\omega}_o) = \lim_{\chi_i \to x_i} L_i S(x_i, \tilde{\omega}_i) / L_i(x_i, \Omega_i) \Omega_i(x_i, \Omega_i) = dL_o(x_o, \tilde{\omega}_o) / dL_i(x_i, \tilde{\omega}_i) . \tag{11}
\]

This is the bidirectional scattering-surface reflectance distribution function (BSSRDF) defined by Nicodemus et al. [NRH+77] using the latter expression. By means of the scattering operator, Eq. (11) provides a link between a scattering function at the most macroscopic level and a level where we can employ path tracing.

We can also work with path tracing at the most macroscopic level. We would then render the following (extended rendering) equation instead of solving the RTE with boundary conditions [Pre65, JMLH01, FHK14]:

\[
L_o(x_o, \tilde{\omega}_o) = L_i(x_i, \tilde{\omega}_i) + \int_{2\pi} \int S(X; x_i, \tilde{\omega}_i; x_o, \tilde{\omega}_o) L_i(x_i, \tilde{\omega}_i) \cos \theta_i d\Omega dA . \tag{12}
\]

Here, \( L_o \) is radiance emitted at the surface, which is a consequence of the volume emission term \( \ell_x \) in Eq. (1). This term was not included in the BSSRDF as we assumed a non-emissive medium.

Chandrasekhar [Cha58] showed that a scattering function like \( S \) is reciprocal, meaning that

\[
S(X; x_i, \tilde{\omega}_i; x_o, \tilde{\omega}_o) = S(X; x_o, \tilde{\omega}_o; x_i, \tilde{\omega}_i) . \tag{13}
\]

This is also clear from the scattering operator as phase functions \( p \) and BSDFs \( f_p \) are likewise reciprocal. Venable and Hsia [VH74] provide measurement equations for the \( S \)-function including the spectral dependency that we can include with all radiometric quantities and optical properties if we want to be explicit about it. Nicodemus et al. [NRH+77] further simplified the BSSRDF by integrating it over a uniformly irradiated area around the point of the emergent radiance. This led to the bidirectional reflectance distribution function (BRDF):

\[
f_r(x, \tilde{\omega}_i, \tilde{\omega}_o) = \int_A S(X; x, \tilde{\omega}_i; x, \tilde{\omega}_o) dA_1 / F(x, \tilde{\omega}_i) \tag{14}
\]

Note the difference in scale between the BSDF in Eq. (5) and this BRDF approximation of the BSSRDF. The former is concerned with surface scattering only. The latter includes any local subsurface scattering that an observed object might exhibit. If we use the approximate BRDF in Eq. (12), so that \( f_r \) replaces \( \int_A S \) \( dA \), we have the conventional rendering equation [Nic65, Kaq86].

**Figure 2:** Illustration of collective scattering distribution functions. The BSSRDF includes both surface and subsurface scattering and models that light can be incident and emergent at different surface positions. The macroscopic BSDF is an approximation of the BSSRDF that assumes scattering to be local with light incident and emergent in the same macroscopic surface position.

5. Collective Scattering Distribution Functions

The BSSRDF, \( S \) in Eq. (11), is a very general scattering function that describes the relationship between any configuration of incoming and outgoing light and the net scattering effect of an object. Figure 2 illustrates how the BSSRDF allows for both in-surface and subsurface scattering throughout a translucent object. The BSSRDF is a function of the object geometry and, given a parametrization of the object surface, four spatial and four angular variables.

As the BRDF is local and considers light to be incident and emergent in the same macroscopic surface position with directions in the same hemisphere, we need a different function for specifying bidirectional transmittance. The bidirectional transmittance distribution function (BTDF) is like the BRDF, but describes the transmittance properties of a thin scattering film. The BTDF has light emerging with a direction in the opposite hemisphere from a point on the opposite side of the film that is not significantly separated from the point of incidence, see Figure 2. Like the BRDF, the BTDF is a function of four angular variables.

With macroscopic BRDF and BTDF functions, we can also define a macroscopic BSDF by combining these two [BDW81]. This is conceptually different from the mesoscopic BSDF defined in Eq. (5), which specifically describes the scattering at an interface. The macroscopic BSDF is illustrated in Figure 2 and describes the scattering of light in materials where volume scattering is not significant. If we consider the differential element of surface area with relevant scattering \( dA \) to be the same both for incident, reflected, and transmitted light, we can measure the BSDF using

\[
f_r(\tilde{\omega}_i, \tilde{\omega}_o) = \frac{dL_o}{dE} = \frac{d^2\Phi_o}{\cos \theta_o dA d\Omega_o} = \frac{d\Phi_o}{dA} \approx \frac{\Phi_o/\Omega_o}{\Phi_i/\cos \theta_i} , \tag{15}
\]

where \( \Phi \) is radiant flux incident from approximately one direction \( \tilde{\omega}_o \) and \( \Phi_o \) is the outgoing flux scattered through solid angle \( \Omega_o \) around \( \tilde{\omega}_o \) at scattering angle \( \theta_o \) relative to the surface normal. The \( \cos \theta_i \) factor corrects for the projection of the illuminated area when viewed from the scattering direction, this is sometimes omit-
With known reflectance properties. Additionally, knowledge of the where the sample is either absent or replaced by a reference sample to characterize the irradiance is to use a reference measurement dimensions \[E2311\]. The setup is illustrated in Figure 3. One way intensity versus scattering direction \(\vec{w}\) where a light source illuminates the sample under study (producing measurements are typically performed with a goniometric approach between the measured and the ‘true’ functions \[Sto12\]. BSDF measurements may also occur out of plane.

When discussing measurements, we specify scattering functions in terms of radiant flux. The reason for this is that not all radiometric quantities are directly measurable. A photodetector converts photons into current and thus responds to the radiant energy entering the volume of the detector through its surface. The detector responsivity is thus a function of position, direction, wavelength, time, and polarization of the incident radiation \[VH74, NKH76\]. To describe an optical radiation measurement mathematically, we should define a measurement equation. This is an equation relating detector signal to responsivity-weighted incident radiance integrated over exposure time and wavelength interval as well as surface area and solid angle of the detector \[VH74, KN78\]. We can convert the signal into radiometric quantities such as radiant flux (through division by calibration measurement and exposure time). Specification of a measurement equation is important for forward and inverse models, as we can use it to convert simulation of reflected or scattered radiance into the quantities that would be used for measurement of the collective scattering distribution functions.

6. Experimental Methodologies

As indicated in the previous section, experimental instrumentation used for measuring a collective scattering distribution function has finite limits on resolution and accuracy. This leads to differences between the measured and the ‘true’ functions \[Sto12\]. BSDF measurements are typically performed with a goniometric approach where a light source illuminates the sample under study (producing irradiance), and a moving detector records the outgoing light intensity versus scattering direction \(\hat{o}_s\) by scanning in one or two dimensions \[E2311\]. The setup is illustrated in Figure 3. One way to characterize the irradiance is to use a reference measurement where the sample is either absent or replaced by a reference sample with known reflectance properties. Additionally, knowledge of the detector aperture size and distance from the sample is necessary for calculating the BSDF (as we need \(\Omega_s\)). The direction of incidence \(\hat{o}_i\) can be varied by either moving the light source or rotating the sample. Lasers are often used as light sources, but other sources can also be applied, especially when employing multiple wavelengths. Naturally, the wavelengths employed in visual appearance measurements are confined to the visible part of the spectrum. Outside this region, technical difficulties arise when going below 200 nm or above 15,000 nm. Further, when using a coherent light source combined with small acceptance apertures, speckle can occur, causing erratic variations in the scattered light \[E2311\]. This can be helped by averaging over larger areas or over several positions. The BRDF is obtained by measuring the BSDF in a reflective geometry, and the BTDF is obtained by measuring in a transmissive geometry.

In contrast to the BSDF, no standard sampling method is available for the 8-dimensional BSSRDF owing to its high dimensionality, and no primary equipment exists for BSSRDF measurements. A common way to estimate the BSSRDF is using a camera-based approach as proposed by Jensen et al. \[JMLH01\]. Here, a tightly focused beam illuminates the sample with a very narrow spatial distribution \(\sigma_i\) in Eq. (11)) while a camera records emergent light from a fixed angle but at multiple positions \(\sigma_o\) in Eq. (11)) simultaneously. A reference measurement with an ideal diffuser was used to determine the irradiance. Jensen et al. \[JMLH01\] divide their BSSRDF model into a directionally dependent single scattering term and a directionally independent multiple scattering term. The latter is accessible by their experimental method as it depends only on the distance between \(\sigma_i\) and \(\sigma_o\). Gkioulekas et al. \[GZB+13\] presented a camera-based measurement of a more complete set of optical properties \((\sigma_i, \sigma_o, p)\) where the incoming and outgoing light directions \((\hat{o}_i, \hat{o}_o)\) were varied within a plane using two motorized rotation stages. Inoshita et al. \[ITM+13\] presented a method for sampling the full BSSRDF. They employed a polyhedral mirror system to illuminate and observe an object from multiple directions.

If we consider large objects of arbitrary shape, and possibly heterogeneous materials, geometry and appearance become intricately intertwined. This makes it particularly difficult to estimate the apparent optical properties of such objects. There are two ways to go about arbitrary shapes: measure both the appearance and geometry simultaneously and separate the optical properties from the geometrical influence, or ignore all the complexity of inhomogeneities and geometry and represent the reflectances using bidirectional texture functions (BTFs) \[DVGN99, DHT+00, LYS01, TWL+05\]. BTFs include the effects of microgeometry and subsurface scattering without actually modelling the behaviour. It boils down to a bidirectionally dependent texture on a macroscopic 3D model, such as a triangle mesh. This is great for the purpose of re-rendering the original object. However, it hides most details of the material and optical properties, making parameterization or appearance transfer very difficult.

In order to estimate the apparent optical properties of objects, we need to measure and account for their geometries. The simultaneous measurement of appearance and geometry is often performed straightforwardly by extending the camera-based methods mentioned above. Using cameras for appearance and optical 3D scanners for geometry has been demonstrated to estimate BRDFs of homogeneous \[MWL+99\] and inhomogeneous \[LKG+01\] ma-
terials and even extended to infer BSSRDFs [GLL*04, WMP*06, GHP*08]. A few studies [HLZ10, TFG*13] measured geometry and appearance from the same underlying data (images), thus removing the complications from combining the data afterwards. Additionally, for convex geometry, Munoz et al. [MES*11] inferred homogeneous BSSRDFs using only single images.

It is important to note that the accuracy of geometrical measurements may be, and often is, poor when the local curvature is comparable to the depth of the subsurface light transport. In fact, subsurface scattering often results in blurring of the light and, in turn, errors to the depth of the subsurface light transport.

Furthermore, Eq. (6) is used in order to model optical scanning objects, and a recent study highlights similar errors in state-of-the-art commercial 3D scanners [GGF*19]. It is therefore critical to consider optical scanning methods designed specifically for the scanning of translucent materials [IKL*10]. The preferred approach is currently to separate surface and subsurface scattering and try to eliminate or reduce the geometric errors. Various schemes have been proposed such as polarization filtering [CLFS07], modulated phase-shifting [CSL08], high frequency (micro) phase-shifting [GN12], and light transport analysis [OMK14]. An alternative is to coat the object with a fine layer of optically diffuse dust [GLL*04]. Unlike the other noncontact methods, coating requires full contact and directly modifies the object surface. The appearance of a translucent object with an arbitrary shape is still difficult to acquire, especially when the optical penetration depth is on the same order as the geometrical estimation errors.

7. Radiometric Models

When considering optical simulation models for how light interacts with scattering surfaces and scattering media, there are two main categories. One category is based on a complete description of how electromagnetic fields interact with materials. An appropriate name for such models is thus field models. These models provide the most realistic, physics-based description of how light is scattered by considering Maxwell’s equations. All coherent wave phenomena are then accounted for, which permits accurate modeling of effects such as interference and diffraction.

Another approach is to describe how radiant energy is redistributed by dissipative (absorption) and non-dissipative (scattering) effects inside a material. These radiometric models are employed in diverse fields such as optical molecular imaging, infrared and visible light photography in space and atmosphere, heat transfer, astrophysics, lighting design, and atmospheric science [THH13].

By focusing solely on the power distribution and ignoring phase information, radiometric models lack the capability of directly modeling effects that require a field representation, such as diffraction and interference. It is however important to realize that such field effects can at least partially be implemented in a radiometric model via the optical properties of the considered scattering materials and scattering surfaces, as explained in Section 3.

In essence, all radiometric models rely on the radiative transfer equation (1) in order to describe how radiant flux is redistributed inside a domain due to absorption, elastic scattering and emission. Furthermore, Eq. (6) is used in order to model how radiant flux is redistributed at a surface between two different materials, including smooth surfaces with specular reflection and transmission. This implies that in order to model an optical system reliably, radiometric models need three input properties for all considered media: the absorption coefficient, scatter coefficient and phase function ($\sigma_a$, $\sigma_s$, $p$), and the BSDF ($f_s$) for all considered surfaces. When there is a significant wavelength dependence in these optical properties, the spectral radiance should be considered, which extends the definition of radiance to include wavelength information:

$$L_\nu = \frac{dL}{dA}.$$  (16)

Solving the radiative transfer equation for a certain material and a certain illumination source permits obtaining the spectral radiance at the boundaries of the material, from which any related radiometric or photometric quantity can then be derived, such as the radiant intensity or the illuminance. In addition, due to Preisendorfer’s link between the RTE and the BSSRDF (see Section 4), values of the collective scattering distribution functions (BSSRDF, BRDF, BTDF) can be simulated for different wavelengths.

Many different methods are available for solving the radiative transfer equation [vdH80, VHWF13], such as discrete ordinate or $S_n$ methods [FYH*18] and spherical harmonics or $P_n$ methods [CSY*15, DBL10]. Of interest are methods such as the Kubelka-Munk model, widely adopted in papermaking and colour industries [AII80], methods based on the adding-doubling method, used in biomedical research [LCH*18, Prn95] and lighting [LMD*14, CCL*16], and Monte Carlo methods [WJZ95, FB09, CHCM17, NGHJ18] which are ubiquitous in many fields.

Importantly, there is some overlap between the numerical methods used to solve either radiometric models or field models. Finite element methods (described in Section 8) can be used to solve both the RTE and Maxwell’s equations and ray-tracing techniques can be extended to include field effects [SBW03]. Thus, it is important to stress that the main distinction between field models and radiometric models is conceptual: field models describe the interactions between electromagnetic waves and the material using Maxwell’s equations, while radiometric models describe how radiometric quantities are redistributed. Both models are essentially direct or forward methods as they permit simulating from intrinsic material properties how light is scattered. While we can use either approach at any scale, one should note that the smaller the geometrical features in question, the more important the wave effects included with the field models.

7.1. Kubelka-Munk Model

The Kubelka-Munk (KM) model [KM31] is a useful way of acquiring the diffuse reflectance of a material [HM92, DH96]. It is a 1D radiative transfer model assuming incidence on one side of a plane-parallel turbid medium, so that we have only two light fluxes propagating either upwards or downwards. When applied to homogeneous media, the theory relates the rates of change of light fluxes in two opposing directions at a position $z$ to the local degrees of absorption and forward and backward scattering. These in turn are stipulated to be proportional to the local flux intensities themselves,
\( \Phi^- (z) \) and \( \Phi^+ (z) \), namely

\[
\frac{d\Phi^-}{dz} = - (\eta_r + \eta_a) \Phi^- + \eta_r \Phi^+, \tag{17}
\]

\[
\frac{d\Phi^+}{dz} = - (\eta_r + \eta_a) \Phi^+ + \eta_r \Phi^-, \tag{18}
\]

where the proportionality constants are phenomenological coefficients of absorption and scattering, \( \eta_a \) and \( \eta_r \).

These phenomenological coefficients, \( \eta_a \) and \( \eta_r \), are linear functions of the intrinsic coefficients of absorption and scattering of the medium, \( \sigma_a \) and \( \sigma_r \), i.e. \[ \text{Nob85}\]

\[
\eta_a = \alpha \sigma_a, \quad \eta_r = \frac{\alpha}{2} \sigma_r, \tag{19}
\]

where

\[
\alpha = \int_0^\frac{\pi}{2} \frac{1}{\Phi_0} \frac{d\Phi^-}{d\Theta} \cos \Theta d\Theta. \tag{20}
\]

The factor \( \alpha \in [1, 2] \) is equal to unity for collimated incident light normal to a non-diffusing medium and equal to 2 when the light distribution is perfectly diffuse as in the case of the original KM model. For other types of light distributions, \( \alpha \) takes a value between these two extremes, depending on their respective diffuseness grades \[ \text{YH08}\].

For a diffuse medium layer of thickness \( D \) immersed in air, its reflectance and transmittance values can be expressed as \[ \text{YH08}\]

\[
R = K r_0 + (1 - r_0)(1 - r_1) \left[ \frac{R_{\infty} - r_1}{1 - R_{\infty}} \right]^2 e^{-2\beta D} - R_{\infty}(1 - R_{\infty} r_1) \left( R_{\infty} - r_1 \right)^2 e^{-2\beta D} - (1 - R_{\infty} r_1)^2 \right), \tag{21}
\]

and

\[
T = (1 - r_0)(1 - r_1) \frac{(1 - R_{\infty}^2) e^{-\beta D}}{(1 - r_1) R_{\infty}^2 - (R_{\infty} - r_1)^2 e^{-2\beta D}}, \tag{22}
\]

where \( \beta = \sqrt{\eta_r^2 + 2\eta_a \eta_r} \). The quantity \( R_{\infty} \) is the intrinsic reflection of the medium, that is,

\[
R_{\infty} = \frac{1 + \eta_a}{\eta_r} - \frac{\beta}{\eta_r}. \tag{23}
\]

In addition to the original two flux approximation, the KM model has been further extended to four fluxes, namely two for the upward and downward collimated radiant fluxes, and another two corresponding diffuse fluxes. It was reported that the four-flux model compares well with numerical solutions of the radiative transfer equation and with highly accurate Monte Carlo simulations \[ \text{VN97}\].

In combination with a multipole BSSRDF model for thin layers \[ \text{Wan98}\], the Kubelka-Munk theory can be used in frequency space to combine several different layers into one BSSRDF model for a multi-layered material \[ \text{DJ05, DJ06, DWd08}\]. This model is used extensively in modelling of the appearance of human skin \[ \text{NMM19}\], which is an indicator that the Kubelka-Munk theory is still highly relevant.

7.2. Adding-Doubling Method

The adding-doubling method solves the radiative transfer equation for time-independent, one-dimensional, azimuthally-averaged problems and homogeneous optical properties \[ \text{vdH68, Pra95}\]. This means that the light scattering distribution in adding-doubling can be described using the inclination angle \( \theta \). For brevity in equations, we use \( \mu = \cos \theta \). Figure 4 illustrates the directional discretization employed by adding-doubling. Thus, adding-doubling is well suited for quickly estimating the radiant intensity distribution or the BSDF of a material slab with known volume scattering properties.

In adding-doubling, the problem domain is first reduced to a layer with infinitesimal thickness, so that only a single scattering event can occur when a beam of light interacts with the material. This enables calculation of a reflection distribution \( R(\mu'; \mu) \) and a transmission distribution \( T(\mu'; \mu) \) describing the radiance normalized to an incident diffuse flux from the direction of incidence \( \partial o \) which is reflected or transmitted towards the outgoing direction \( \partial o' \).

With vectorized single-event layer distributions, adding-doubling proceeds with the doubling method. A given layer with surfaces 0 and 1 is doubled by positioning a layer 2 with the same properties on top. New transmission and reflection distribution functions for the combined layer are calculated using

\[
T^{ij} = T_{ij} (I - R_{i+1} R_{j+1})^{-1} T^{i+1}, \tag{24}
\]

\[
R^{ij} = T_{ij} (I - R_{i+1} R_{j+1})^{-1} R_{i+1} T^{j+1} + R^{j+1}.
\]

These are 2D matrix expressions in which \( I \) is the identity matrix, and \( i = 0 \) and \( j = 2 \) or conversely. The doubling process is iterated until the thickness of the combined layers is the same as the sample being modelled.

If necessary, e.g. in the case of a liquid sample inside a glass container, different layers can be added at the boundaries using the adding method. This uses expressions similar to Eqs. (24) to update the scattering distributions but accounts for the different refractive index of the added boundary layers. The resulting scattering functions can be used to predict how a distribution of incident light \( I(\mu) \) is scattered backwards or forward. The radiant intensity of light...
scattered forward is, for example,
\[ I(\mu) = \int_0^1 T(\mu', \mu) I(\mu') 2\mu' d\mu' \]

Other important radiometric quantities, such as the total reflection or transmission, can be calculated from the radiat intensity.

Importantly, the adding-doubling method imposes little restrictions on the properties of the scattering materials that can be used with it [Pra95]. It supports anisotropic phase functions and imposes virtually no limit on the size parameter (particle size to wavelength ratio). The key advantage of adding-doubling is its efficiency, making it well suited for solving the inverse problem of extracting the (spectral) volume scattering properties from total transmission and reflection, or radiant intensity measurements of samples, assuming these are homogeneous and planar [PvGW93, LMD]. Additionally, adding-doubling permits calculation of internal radiometric quantities, e.g. fluence. However, the method is not particularly well suited for this task and alternative methods based on discrete ordinates can be used instead.

The adding-doubling method is used in graphics as a forward model for computing the BRDF or the BSSRDF of multi-layered materials that include scattering within the layers [EKM01, JdJM14, Bel18, ZJ18]. However, to the best of our knowledge, the inverse technique has not been used in the graphics community. Inverse adding-doubling is dismissed by Gkioulekas et al. [GZB13] as being too limited in terms of angular scattering information and thus restricted to inversion based on low-parameter phase function models. Even so, inverse adding-doubling is the reference method for acquiring the wavelength-dependent scattering properties of translucent materials \((\sigma(\lambda), \sigma_d(\lambda), p(\mu, \lambda))\) in many other fields.

### 7.3. Monte Carlo Methods

Monte Carlo ray tracing is one of the most widely used methods for solving the radiative transfer equation (RTE). In rendering, the purpose is commonly to obtain realistic depictions of scenes with participating media [NGHJ18]. In other fields, such as lighting and biomedicine, the focus is usually on estimating a wide range of external and internal radiometric quantities or on modelling specific biomedical imaging devices [ZL13, PP17]. It is important to note that a Monte Carlo method is really applying Monte Carlo integration to the measurement equation for the detector that we are modelling (most often pixels of a camera). Any radiance term appearing in the measurement equation is then obtained through Monte Carlo integration of the integral version of the RTE including boundary conditions (6–8). Alternatively, if a BSSRDF is available, we would Monte Carlo integrate Eq. (12) instead.

Unlike the other radiometric models, Monte Carlo ray tracing follows a probabilistic view of the light scattering process [Kaj86, Rus88, WIZ95]. The redistribution of energy through scattering and absorption is individualized into eye rays or light rays or photon packets which interact with the material independently from one another. In this way, we can describe a light path (or eye path) before and after interacting with the material using solely geometric principles. The flexibility afforded by this approach is significant because, unlike the other radiometric models, any sample geometry and boundary condition can be readily used with, or closely approximated by, Monte Carlo ray tracing. With Monte Carlo, the three-dimensional radiative transfer problem can be solved for heterogeneous materials under arbitrary illumination conditions. Furthermore, Monte Carlo ray tracing provides a very straightforward way of modelling materials that exhibit inelastic scattering such as bioluminescent samples or fluorescent phosphors.

In Monte Carlo ray tracing, interactions between light rays and material volumes are modelled stochastically. For instance, the direction in which a light ray is scattered after interacting with a volume can be determined by sampling the phase function using a pseudo-random, a low discrepancy, or a quasi-random number [KTB11]. Similarly, we can decide whether to scatter or absorb a light ray, or whether to reflect or refract the ray at the surface, using the Russian roulette sampling technique [AK90].

After a sufficiently large number of paths have been traced through a translucent object and terminated, they are aggregated and their combined information is useful for simulating any radiometric quantities of interest. One option is to evaluate BRDF [PH00] or BSSRDF [DLR09] for various optical properties. For lighting, important quantities such as the luminance or the radiant intensity distribution (beam pattern) are typically calculated using Monte Carlo ray tracing [CHCM17], while in biomedicine the same methods are used for calculating internal radiometric quantities such as the fluence distribution [FB09, WIZ95].

The flexibility of Monte Carlo ray tracing comes at the expense of computational efficiency or simulation time. The accuracy of a simulation depends on the number of light paths (and/or eye paths) that we trace, and generally millions of different light rays need to be traced to decrease stochastic noise and obtain a reasonably accurate simulation. Depending on the radiometric quantity being simulated and the necessary accuracy, it may be necessary to trace a greatly varying number of rays, which impacts the simulation time. For instance, accurate estimates of the transmission and the reflection of a scattering sample are substantially faster to obtain than an accurate simulation of the spectral radiance distribution.

The lower computational efficiency of Monte Carlo ray tracing in simulating a scattering distribution may make it a poor match to solve the inverse problem, which typically requires performing several simulations to obtain an accurate solution. In most cases, the flexibility afforded by Monte Carlo ray tracing overcomes this limitation. In addition, the recent development in graphics processing hardware has significantly improved the performance of Monte Carlo ray tracing and made it a competitive approach to solving the inverse problem [MZL16, NDVZ19, LHJ19].

### 8. Field Models

The collective scattering distribution functions (BSSRDF, BRDF, and BTDF) are good candidates for real time monitoring of the surface finishing in industrial processes since they describe the appearance on a macroscopic scale, and the measurements are typically vibration insensitive. We will in this section describe how appearance can be linked to direct physical measurands of the object using a physical model of the interaction between light and matter. This model may be a solution to Maxwell’s equations, the Helmholtz equations or approximations hereof.
A rigorous solution to Maxwell’s equations, found by methods like Lorenz-Mie theory, finite elements, finite difference time domain, and the rigorous coupled-wave analysis, gives the full scattering solution including polarization and coherence properties. As illustrated in Figure 5, the field interference between coherent light waves can have a dramatic effect on the intensity distribution of the scattered light when the critical dimension of the object is comparable to the wavelength. Clearly, these methods should be used for structure and elements with sub-wavelength features. These are becoming increasingly important within many fields such as advanced process control in the semiconductor device manufacturing industry and size determination of nanoparticles within the particle sizing industry. However, real world problems are very often too complex to be solved using rigorous solutions. The scalar solution to the Helmholtz equation is therefore the workhorse for establishing the link between appearance and the direct physical parameters that can be used to characterize an object.

Collective scattering distribution functions are directly linked to optical properties through an inverse relation. Inverse problems are in general ill-posed, but are solvable if we have enough a priori information. We will show that in the smooth Root-Mean-Square (RMS) roughness limit, we do not need any a priori knowledge to handle surface scattering, but as the RMS roughness increases, the need for a priori information increases. The last part of this section describes how the volume scattering and substrate scattering may be incorporated into the description. The incorporation of substrate scattering is the simpler of the two since the substrate can be investigated as a surface by sample turning. Non-destructive direct physical measurements of volume scatterers are not accessible, and we will therefore handle the volume scatterers as a statistical distribution of scatter sizes that may be represented as interface layer(s) within the volume when using scalar diffraction.

8.1. Lorenz-Mie Theory

The scattering of a plane wave of light by a sphere was fully described by Lorenz [Lor90]. Later, Mie [Mie08] found the same solution with an outset in Maxwell’s equations and included the case of an absorbing sphere. This rigorous solution for scattering of light by a sphere is remarkably useful when it comes to computation of the far field scattering properties (\(\sigma_u, \sigma_s, p\)) of a medium containing a random distribution of scattering particles [Cal96, JW97, FCJ07, DFKB16, FK19].

The Lorenz-Mie solution for Maxwell’s equations is a series expansion, which is fairly inexpensive to evaluate numerically. The expansion coefficients are referred to as Lorenz-Mie coefficients \((a_n, b_n)\). These coefficients depend on size parameters that are simple functions of particle radius, wavelength, and refractive indices of host medium and particle. The scattering cross section in Eq. (3) is then

\[
C_s = \frac{\lambda^2}{2\pi n_{med}^2} \sum_{n=1}^{\infty} (2n+1) \left( |a_n|^2 + |b_n|^2 \right),
\]

where \(n_{med}\) is the refractive index of the host medium. Similarly, we can find the absorption cross section \(C_a = C_t - C_s\) by computing the extinction cross section \(C_t\) using

\[
C_t = \frac{\lambda^2}{2\pi} \sum_{n=1}^{\infty} (2n+1) \text{Re} \left( \frac{a_n + b_n}{n_{med}} \right),
\]

where \(\text{Re}\) takes the real part of a complex number. Assuming independent scattering by the particles in a medium, which is a reasonable assumption for a volume fraction of particulate inclusions below 0.1, and given a particle number density (a size distribution), we can derive \(\sigma_t, \sigma_s\) from these cross sections using Eq. (3).

We can also use the Lorenz-Mie coefficients for computing the phase function of a spherical particle. As an example, this works well for droplets up to a radius of 0.4 mm after which deformations due to gravity start resulting in visible deviations in the distribution of the scattered light [SML*12]. Another option is to directly calculate the asymmetry parameter, which is the mean cosine of the scattering angle defined by

\[
g = \frac{1}{4\pi} \int p(\omega') (\hat{\omega}' \cdot \hat{\omega}) \, d\omega'.
\]

For the phase function of a spherical particle, we have [vdH57]

\[
g = \frac{\sum_{n=1}^{\infty} \left( \frac{n(n+2)}{n+1} \text{Re} (a_n a^*_n + b_n b^*_n) + \frac{2n+1}{n+1} \text{Re} (a_n b_n^*) \right)}{2 \sum_{n=1}^{\infty} (2n+1) \left( |a_n|^2 + |b_n|^2 \right)},
\]

where an asterisk * denotes the complex conjugate.

Clearly, Lorenz-Mie theory provides a fairly straightforward way of computing the scattering properties of materials composed of low concentrations of randomly distributed particles that are reasonably spherical. However, the Lorenz-Mie theory provides no information regarding the surface scattering of a translucent material. To deal with this, and in cases where the scatterers are more tightly packed or spatially-correlated [JAG18], we need more general numerical field models. In the following, we discuss field models providing solutions for more arbitrary geometry. This means that we can include interference of waves scattered by different geometric features and multiple scattering between the features. However, these models require precise specification of the micro- or nanoscopic features and are not as easily connected to more macroscopic apparent particle size distributions.

8.2. Finite Element Methods

The finite element method (FEM) is a technique for numerically solving partial differential equations (PDEs) [Red06]. It is flexible and used among many physical disciplines such as optics [DM93].
mechanical engineering [Hug00], and fluid dynamics [ZTN14]. In
graphics, the finite element method is commonly referred to as the
radiosity method [TM93]. Starting with Rushmeier and Tor-
rance [RT87], it has often been used for solving the radiative trans-
fer equation (1). Sheng et al. [SSWN13] used this kind of so-
olution for computing the BSSRDF of a discretized translucent object.
However, no work in graphics seems to use FEM for computing
optical properties by solving Maxwell’s equations.

Three fundamental steps are employed in FEM:
1) Meshing: discretizing the continuous sample by dividing the
structure into N smaller sub-domains or elements defined by a
mesh.
2) Approximating a solution for each of the N elements ui, as a
local value, c_i multiplied by a polynomial ψ_i, describing how the
solution varies in the parameter space across the element. These are
referred to as element equations.
3) Assembling the global system matrix as a linear combination
into the original PDEs and applying boundary conditions, and
finally solving for the global solution.

The FEM method controls the numerical accuracy through mesh
refinement (N) and the polynomial degree of ψ_i on the individual
patches of the mesh. The system matrix size increases with N and
the order of the polynomials ψ_i. This becomes (as always) a trade-
off between accuracy and a high computation time.

The finite element method allows for very complex geometries.
Some software packages allow for geometries to be defined by CAD files (COMSOL), and some are specifically designed for optical
simulations (JCMWave).

8.3. Finite Difference Time Domain
The finite difference time domain (FDTD) method calculates the
electric and magnetic fields at a given point in space, and advances
them in small time steps, ∆t, and spatial steps, ∆x, ∆y, ∆z, accord-
ing to Maxwells equations [TH05]. By using a Yee grid scheme
[Yee66], where the fields are stored inside the resulting unit cells,
Maxwell’s divergence equations are naturally satisfied, and we are left with
\[
\frac{\partial E}{\partial t} = -\frac{1}{\mu} \nabla \times H, \quad \frac{\partial H}{\partial t} = \frac{1}{\varepsilon} \nabla \times E. \quad (29)
\]
For simplicity, we have here assumed a sample with permeability \(\mu\) and
permittivity \(\varepsilon\) but no conductivity. Since a change in \(H\) implies
a change in \(E\), the \(E\) field is found from \(H\), and \(H\) is found from
\(E\) in each time step, ∆t. Thus, the numerical calculations of the
individual fields are performed in steps of \(\Delta t\). One can then evaluate
\(H(t + \Delta t)\) from \(E(t)\), \(E(t + \Delta t)\) from \(H(t + \Delta t)\), \(H(t + 2\Delta t)\)
from \(E(t + \Delta t)\) and so forth. This is done for grid points defined by
unit cells of volume: \(V = \Delta x \cdot \Delta y \cdot \Delta z\). A dense grid gives a higher
accuracy at the cost of computation time. For light, ∆t is further
restricted by: \(\Delta t \ll \frac{\Delta x}{c}\), to satisfy the CFL condition [CFL67].

Since the calculations are performed in the time domain for all
points, it is trivial to graphically represent the system at a given time
step, which makes it intuitive to study an impulse response of the
electromagnetic system. The method is most efficient when the di-

densions of the examined structures are similar to the wavelength
of the light used [Sch17]. This technique can handle multiple wave-
lengths simultaneously, but requires a very large grid if one wants
to calculate diffraction efficiencies in the far field, which in turn
increases the computation time.

FDTD was introduced to the graphics community by Musbach et al. [MMRO13]. They used it for computing a tabulated BRDF. In
particular, they modelled the blue iridescent appearance of the mor-
pho butterfly. This species of butterfly has been used as a case study
by others as well (Sun [Sun06], for example). Another use of FDTD in
graphics is to compute a nanostructure that exhibits a particular
reflectance colour and fabricating a surface with this structure us-
ing additive manufacturing [AHB18]. The main objection to FDTD
seems to be that it would be extremely time consuming to compute a
collective scattering distribution function if the microgeometry of
interest has an extent of a cubic millimetre or more [WVJH17].

8.4. Aperiodic Rigorous Coupled-Wave Analysis
The concept of RCWA relies on solving Maxwell’s equations insi-
de a medium that is uniform in the \(z\)-direction (normal to the sur-
face) so that the light propagating in the medium can be regarded as
plane waves; and subsequently performing Fourier expansion of
Maxwell’s equation along the \(x\) and \(y\) directions. Analytical
treatment of the wave equation in the \(z\)-direction enables RCWA to
model uniform structures of arbitrary heights. General structures,
that are not uniform in the \(z\)-direction, are subdivided into layers
that are uniform in the \(z\)-direction, and the boundary condition may
be solved using the enhanced transmittance matrix approach as ex-
plained in great detail in a recent review paper [MH16]. The RCWA
was originally developed for periodic structures [MGP95] and
contained a wrong treatment of the Fourier expansion for transverse
magnetic (TM) waves. However, the TM error was later corrected with
the introduction of the normal vector method [SRK07].

Any computer program works, in practice, in a finite-size vir-
tual space. The propagation of electromagnetic waves is confined to
this region and can be further restricted to a specific region of
interest by introducing a perfectly matched layer (PML). A PML is an artificial layer that enables accurate calculation of the electro-
magnetic field inside the region of interest by isolating this region from
neighbouring areas. The combination of RCWA with a PML is
called aperiodic RCWA (aRCWA). Today, aRCWA is a mature
technology for evaluation of the electromagnetic field that offers
great flexibility and short computation times. Among the popular
computational methods, FEM, FDTD, and aRCWA, in electromag-
netic analysis, aRCWA is the youngest and most unknown. How-
ever, the method is very well adapted for nanophotonics and capa-
bale of providing solutions to problems that cannot be solved with the
other methods [HMLN17]. The output from the rigorous methods
are the scattered power over the incident power (diffraction effi-
ciency), which may be transformed into a BSDF by the use of
Eq. (15). To the best of our knowledge, this approach has not yet
been applied in graphics.

8.5. Scalar Diffraction Models
The Kirchhoff integral of scalar diffraction theory was introduced to
graphics by Kajiya [Kaj85] and again by He et al. [HTSG91].
Kajiya’s BRDF model takes a surface microstructure as input. The BRDF of He et al. is based on normal distribution functions and includes subsurface scattering as a wavelength-dependent diffuse reflectance. Stam [Sto99] connected the Kirchhoff integral to Fourier transformation of a given microsurface height map. Dong et al. [DWMG15] followed up on Stam’s work and used it for profilometer-based BRDF acquisition. The scalar diffraction BRDF models to be discussed in the following (generalized Harvey-Shack, Rayleigh-Rice) were introduced to graphics by Löff et al. [LKYU12] and further investigated by others [HP17, WYHJ17, YHW+18], Löw et al. [LKYU12] derived a BRDF model independent of wavelength and fitted it to BRDFs measured by Matusik et al. [MPBM03]. Apart from this, the scalar diffraction models seem to be used in graphics only as forward models.

The rigorous electromagnetic solvers (FDTD, FEM, aRCCA) are forward solvers too, since they produce an electromagnetic field from a predefined structure. Sometimes, however, we are interested in the inverse problem: determining the structure given an electromagnetic field or intensity. This problem is solved by formulating an optimization problem, as described in Section 10, that finds the best matching solution through variation of the structure geometry. This approach needs a global optimization method or good initial a priori knowledge of the surface structure. Scalar diffraction theory (SDT) can provide this initial information [HN09] since SDT generally is so simple that it can be formulated as a true inverse problem, without traditional forward calculations. The scalar Helmholtz equation may be rewritten as the Lippmann-Schwinger equation, which together with regularization by Neumann series provides a very general method for solving the inverse problem [LKK13, CK13]. Some of the most widely used inverse scattering solutions are Rayleigh-Rice (RR) light scattering theory and microfacet theory. Both methods connect the BRDF to the power spectral density (PSD) [Sto12, GGG*16]. This can be done for a field formulation [LH13, JRHH17] or for a radiometric formulation by

$$f_r(\theta_i, \phi_i; \theta_r, \phi_r) = Q \frac{16\pi^2}{\lambda^2} \cos \theta_i \cos \theta_r \text{PSD}(\nu)$$

where $Q$ is a factor dependent on the polarization state of the light source, the optical properties, and the wavelength $\lambda$ [Sto12]. The spatial frequency $\nu$ is related to the in-plane scattering angle $\theta_r$, the azimuthal scattering angle $\theta_r$, and the in-plane angle of incidence $\theta_i$ by the following formula:

$$\nu = \sin \theta_r \cos \theta_i - \sin \theta_i.$$  

(31)

The PSD describes surface statistics in terms of spatial frequencies $\nu$. We can get the root-mean-square (RMS) roughness ($R_q$) from the PSD by integrating over the spatial frequencies, and the RMS slope ($R_{d\nu}$) by multiplying the PSD by $(2\pi \nu)^2$ before integrating over the spatial frequencies. The physical surface slope may be obtained by scanning the measurement area over the sample and monitoring the deflection of the specular position in the distribution function [BBG*08]. A resolution limited surface profile with the correct RMS roughness and RMS slope can be constructed from the deflection measurements, and this surface profile may be used as input to the rigorous and scalar electromagnetic solvers.

For larger surface roughness, another method called Generalised Harvey-Shack (GHS) has been developed by Krywonos et al. [KHC11]. GHS is a transfer function describing the interaction between light and a rough surface, provided that the surface PSD (or alternatively the auto-correlation function, ACF) is known, and that the surface has a Gaussian height distribution. Then

$$f_r(\theta_i, \phi_i; \theta_r, \phi_r) = Q F(H_i(\hat{x}; \mu_i, \rho_i))$$

$$H_i(\hat{x}; \mu_i, \rho_i) = \exp\left[-(2\pi\hat{R}_{\text{rel}}(\mu_i + \rho_i))^2 \left(1 - \frac{\text{ACF}(\nu)}{R_{\text{rel}}^2}\right)\right].$$

(32)

(33)

where $F$ is the Fourier transform, $\hat{x} = \frac{\nu}{\lambda}$, $\mu_i = \cos \theta_i$, $\rho_i = \cos \theta_r$, $R_{\text{rel}} = \frac{\hat{R}_{\text{rel}}}{\lambda}$ is the band limited RMS roughness found by integrating the PSD curve over the frequency interval $\frac{1}{\lambda} < \nu < \frac{1}{\lambda}$. Here, $d$ is the diameter of the measurement spot, $\frac{1}{\lambda}$ is the spatial frequency limit for propagation modes and $R_{\text{rel}}$ is the total RMS roughness found by integrating the experimental PSD curve over the entire positive frequency range $\frac{1}{\lambda} < \nu < \frac{1}{\lambda}$, where $\nu_{\text{min}} = \frac{1}{d}$ and $\nu_{\text{max}} = \frac{1}{\lambda}$ are the experimental cutoff frequencies of the instrument. Consequently,

$$R_q^2 = 4 \int_{\nu_{\text{min}}}^{\nu_{\text{max}}} \text{PSD}((2\pi)^2) \text{PSD}((2\pi)^2) \text{dv}.$$  

(34)

(35)

The two-dimensional PSD from a physical surface may be obtained from microscope images of a surface and turned into the one-dimensional PSD considered here by image averaging along the sample BRDF measurement direction [FFH+15]. It is advisable to use a PSD curve with a wider frequency range than the frequency range obtained by transforming the BRDF angles into frequency space [Sto12]. This can be achieved by making confocal microscope images of the same area with different numerical apertures and plotting them together as one PSD plot. The PSD plot is then fitted using the ABC model [CTL90] (the $K$-correlation model) over the entire frequency space. The ACF may be obtained as the inverse Fourier transform of the PSD curve and the roughness parameters by integrating the PSD curve in Eqs. (34) and (35).

It is very important to realize that the presented PSD formalism enables us to link the nano/micro length scale with the macroscopic mm/cm length scale. The PSD method thus provides a uniform, scale-independent method for scattering estimation. Here, we show the results for the ABC model with $C = 2$, which is typically a good approximation:

$$\text{PSD}_k(\nu) = \frac{A}{1 + (B\nu)^2}$$

$$\text{ACF}(\nu) = \frac{A}{B^2} \left(1 - \frac{\nu_{\text{max}}}{\nu_{\text{max}}}\right)$$

$$R_q^2 = 4A^2 \frac{1}{B^3} \left(1 - \frac{\nu_{\text{max}}}{\nu_{\text{max}}}\right)$$

$$R_{d\nu}^2 = \frac{16\pi^2 A}{B^3} \left(1 - \frac{\nu_{\text{max}}}{\nu_{\text{max}}}\right).$$

(36)

(37)

(38)

(39)

The use of the ABC model enables us to perform the inverse Fourier transformation and the integration analytically, which is a great help in fitting the GHS to an observed BRDF and thereby obtain-
The GHS algorithm is based on the angular spectrum method, which is presented in the following references [Sto12, Goo05, HKB06].

The periodic contribution in the PSD amplitudes arises from sinusoidal waves (sinusoidal gratings). The amplitudes of sinusoidal gratings have been calculated by many authors and may be found in the following references [Sto12, Goo05, HKB06].

Comparisons of the GHS model with more rigorous methods have found good overall agreement [CH13, SDC+11, HN09]. The GHS algorithm is based on the angular spectrum method, which has a computational complexity of $O(N^2 \log(N))$, whereas the rigorous solver has a best-case complexity of $O(N^2)$. A BRDF from the GHS theory thus serves as an excellent starting point for a more rigorous treatment, when higher accuracy is needed.

### 8.6. Inclusion of Subsurface Scatterers and Substrate Surface

We can use rigorous solvers for simulating light scattering by the entire sample provided that the computational domain is not too big and that the scattering by surface, substrate, and volume are known with sufficient precision. However, this is often not the case, so we have to consider other methods. One approach is to approximate the entire system by three independent subsystems consisting of a subsystem for surface scattering $S_r$, a subsystem for back surface scattering $S_{\text{vol}}$, and a subsystem for subsurface scattering $S_{\text{vol}}$ as shown in Figure 6. In order to calculate BSSRDF and BTDF expressions, we need knowledge of the three subsystems.

To describe surface and back surface scattering, we may use the same procedure. The substrate surface is investigated first using imaging methods like confocal microscopy. Next, the sample is inverted and an image is acquired, the negative of this image is the back surface facing the sample. The PSD is now calculated for the back surface image and inserted in either the GHS or the RR method described in Section 8.5. Subsurface scatterers play a role when the experimentally observed scattering field cannot be explained by the surface and the back surface scattering methods.

In cases where the contribution from subsurface scatterers is much smaller than the surface scattering, the subsurface may be handled as a perturbation to the surface scattering. The most straightforward way of including subsurface scattering is by adding an empirical formula for the subsurface scattering [HGP14]. Here, we will for the first time show how the GHS model can be used to model subsurface scatterers.

GHS calculation is much faster than rigorous calculations and therefore much more well-suited for the regression method for solving the inverse problem (discussed in Section 10). For the BSSRDF, we have

$$S(\theta_i) = f_r(\theta_i) + S_{\text{vol}}(\theta_i) + S_{\text{back}}(\theta_i) + f_i(\theta_i)$$

(41)

$$f_r(\theta_i) = \sum_j Q_{\text{ref},jk} F(H_i(\hat{x};\mu_j,\mu_k))$$

(42)

where we have neglected higher order contributions, $Q_{\text{ref}}$ is given by $Q$ in Eq. (32) and the subscript ref is for reflection. Similarly, we will use the subscript trans for transmission.

The transmittance through the surface dividing the incident medium 1 with refractive index $n_1$ and the volume medium 2 with refractive index $n_2$ is given by $Q_{\text{trans}} H_i(\hat{x};\mu_j,\mu_k)$, which may be written out as

$$Q_{\text{trans}} \exp \left[ -2\pi n_1 n_2 (n_1 n_2 - 1) \right] \left( 1 - \frac{\text{ACF}(\delta)}{R_{\text{out}}^2} \right).$$

This expression tells us the amount of light scattered into a specified scattering angle for a given incident angle. This process may be repeated until the angular 2D scattering function has been put together, see Figure 6. The scattered light is propagated along the scattered angle in the homogeneous medium, with real or complex refractive index, until it interacts with an interface layer with a known PSD, embedded in the volume. One incident wave on the surface may create several transmitted scattered waves. All the transmitted scattered waves are incident waves, with incident angles $(\theta_i, j = 1,\ldots)$, on the interface layer embedded in the volume and the reflected scattered waves from the embedded interface layer are propagating upwards through the surface layer where they scatter once more before entering the incident medium as the volume scattering term ($S_{\text{vol}}$). Calculations of BRDF using GHS therefore require a separate 2D numerical Fourier transform for each scattering angle. This is manageable if analytic PSD functions exist for all layers (surface, interface, and back surface), and we may write

$$S_{\text{vol}}(\theta_i) = \sum_{jk} Q_{\text{vol},jk} T_{jk} F(H_i(\hat{x};\mu_j,\mu_k))$$

(43)

The summation over incident angles $(\mu_j,\mu_k = \cos(\theta_i,j))$ is performed for all incident light on the embedded interface layer generated by an incident angle $(\theta_i,j)$ on the surface layer. The subscript $jk$ on $Q_{\text{vol}}$ indicates that the factor depends on angle of incidence. $T_{jk}$ is a damping factor for the $k$’th incident angle on the embedded interface.
Objective functions are usually different for different inverse problems. An objective function is often formulated so that it includes multiple sets of measurements and simulations, incorporates *a priori* information about the inverse problem, or uses different metrics for the similarity between the experimental data and the simulation data.

The problem then becomes finding specific scattering parameters $x$ (these are different for intensity and field models) such that

$$
\arg \min_x F(x).
$$

One should carefully consider which of the many methods currently available for numerical optimization are best suited for the problem at hand. Different types of least-squares, convex optimization, and metaheuristic methods have been successfully applied to solve the inverse problem for scattering materials. Solving the inverse scattering problem is complicated by the fact that the problem is often nonlinear and improperly posed, meaning that the measured data is not perfect, and an exact solution does not exist. The job then becomes to find approximate solutions to the stabilized problem. A general discussion of such issues in inverse scattering theory can found in [CK13]. The following subsections will provide an overview of different considerations and methods used for solving the inverse problem when using either radiometric or field models as the forward model.

### 9.1. Inverse Radiometric Models

Inverse methods that are based on a forward radiometric model are usually applied to estimate three different optical properties from measurements: the absorption coefficient $\sigma_a$, the scattering coefficient $\sigma_s$, and the phase function $p$. Their effect is illustrated in Figure 8. The geometry of the scattering sample and its refractive index over the wavelength range of interest are typically assumed to be known with a high degree of accuracy. This means that in most scenarios, an inverse radiometric model is applied to estimate the volume scattering properties of a translucent material and not its surface scattering properties. While the scattering and absorption coefficients are scalars (at least for an isotropic homogeneous material and a fixed wavelength), the phase function is a distribution that can take a wide variety of different shapes. For most problems, it is assumed that the phase function only depends on the angular difference between the incident and scattered direction, reducing it to $p(\theta)$. In some cases, the full phase function distribution is estimated with an inverse method, for instance by optimizing a discretized form of the distribution [GZB+13]. More commonly however, we use phase function models [LMD+14] that try to strike a balance between complete flexibility (arbitrary phase function) and simplicity (few model parameters).

For particles smaller than the wavelength of the light, the Rayleigh phase function is a good choice with no directional parameters. The definition is [Cha50]

$$
p_R(\theta) = \frac{3}{4\pi} (1 + \cos^2 \theta).
$$

The Rayleigh phase function even comes with simple importance sampling and a well-defined particle scattering cross section $C_t$ that
HG and GK phase function models are unable to reproduce. \[RM80\] Monte Carlo ray tracing. The Gegenbauer kernel phase function enables efficient use of it in \[HK93\] described importance sampling of model are two of the most widely used phase functions. The Henyey-Greenstein phase function is \([HG40]\)

\[PHG(\theta) = \frac{1}{4\pi} \frac{1 - g^2}{(1 + g^2 - 2g\cos(\theta))^{3/2}}. \tag{49}\]

Hanrahan and Krueger \([HK93]\) described importance sampling of the Henyey-Greenstein phase function to enable efficient use of it in a Monte Carlo ray tracing. The Gegenbauer kernel phase function is \([RM80]\)

\[p_{GK}(\theta) = \frac{\pi}{\pi} \frac{\left(1 - g^2\right)^2}{(1 + g^2 - 2g\cos(\theta))} \left(1 + k^2 - 2k\cos(\theta)\right)^{-\gamma-1}. \tag{50}\]

The asymmetry parameter \(g\) (defined by Eq. (27)) is a common parameter appearing in many phase function models. It describes the directional tendency of the phase function (see Figure 8): \(g = 1\) is perfect forward scattering, \(g = -1\) is perfect back scattering, and \(g = 0\) is symmetric scattering. The shape factor \(\gamma\) of the GK phase function permits capturing a wider diversity of scattering anisotropy, with \(\gamma = 0.5\) reverting the GK phase function to the Henyey-Greenstein model.

For larger particles, the Henyey-Greenstein (HG) phase function and the Gegenbauer kernel (GK) or Reynold and McCormick model are two of the most widely used phase functions. The Henyey-Greenstein phase function is \([HG40]\)

\[p_{HG}(\theta) = \frac{1}{4\pi} \frac{1 - g^2}{(1 + g^2 - 2g\cos(\theta))^{3/2}}. \tag{49}\]

Hanrahan and Krueger \([HK93]\) described importance sampling of the Henyey-Greenstein phase function to enable efficient use of it in a Monte Carlo ray tracing. The Gegenbauer kernel phase function is \([RM80]\)

\[p_{GK}(\theta) = \frac{\pi}{\pi} \frac{\left(1 - g^2\right)^2}{(1 + g^2 - 2g\cos(\theta))} \left(1 + k^2 - 2k\cos(\theta)\right)^{-\gamma-1}. \tag{50}\]

The asymmetry parameter \(g\) (defined by Eq. (27)) is a common parameter appearing in many phase function models. It describes the directional tendency of the phase function (see Figure 8): \(g = 1\) is perfect forward scattering, \(g = -1\) is perfect back scattering, and \(g = 0\) is symmetric scattering. The shape factor \(\gamma\) of the GK phase function permits capturing a wider diversity of scattering anisotropy, with \(\gamma = 0.5\) reverting the GK phase function to the Henyey-Greenstein model.

For larger particles, the Henyey-Greenstein (HG) phase function and the Gegenbauer kernel (GK) or Reynold and McCormick model are two of the most widely used phase functions. The Henyey-Greenstein phase function is \([HG40]\)

\[p_{HG}(\theta) = \frac{1}{4\pi} \frac{1 - g^2}{(1 + g^2 - 2g\cos(\theta))^{3/2}}. \tag{49}\]

Hanrahan and Krueger \([HK93]\) described importance sampling of the Henyey-Greenstein phase function to enable efficient use of it in a Monte Carlo ray tracing. The Gegenbauer kernel phase function is \([RM80]\)

\[p_{GK}(\theta) = \frac{\pi}{\pi} \frac{\left(1 - g^2\right)^2}{(1 + g^2 - 2g\cos(\theta))} \left(1 + k^2 - 2k\cos(\theta)\right)^{-\gamma-1}. \tag{50}\]

The asymmetry parameter \(g\) (defined by Eq. (27)) is a common parameter appearing in many phase function models. It describes the directional tendency of the phase function (see Figure 8): \(g = 1\) is perfect forward scattering, \(g = -1\) is perfect back scattering, and \(g = 0\) is symmetric scattering. The shape factor \(\gamma\) of the GK phase function permits capturing a wider diversity of scattering anisotropy, with \(\gamma = 0.5\) reverting the GK phase function to the Henyey-Greenstein model.

For larger particles, the Henyey-Greenstein (HG) phase function and the Gegenbauer kernel (GK) or Reynold and McCormick model are two of the most widely used phase functions. The Henyey-Greenstein phase function is \([HG40]\)

\[p_{HG}(\theta) = \frac{1}{4\pi} \frac{1 - g^2}{(1 + g^2 - 2g\cos(\theta))^{3/2}}. \tag{49}\]

Hanrahan and Krueger \([HK93]\) described importance sampling of the Henyey-Greenstein phase function to enable efficient use of it in a Monte Carlo ray tracing. The Gegenbauer kernel phase function is \([RM80]\)

\[p_{GK}(\theta) = \frac{\pi}{\pi} \frac{\left(1 - g^2\right)^2}{(1 + g^2 - 2g\cos(\theta))} \left(1 + k^2 - 2k\cos(\theta)\right)^{-\gamma-1}. \tag{50}\]

The asymmetry parameter \(g\) (defined by Eq. (27)) is a common parameter appearing in many phase function models. It describes the directional tendency of the phase function (see Figure 8): \(g = 1\) is perfect forward scattering, \(g = -1\) is perfect back scattering, and \(g = 0\) is symmetric scattering. The shape factor \(\gamma\) of the GK phase function permits capturing a wider diversity of scattering anisotropy, with \(\gamma = 0.5\) reverting the GK phase function to the Henyey-Greenstein model.

For larger particles, the Henyey-Greenstein (HG) phase function and the Gegenbauer kernel (GK) models \([WGXZ19]\]. With a phase function model, the inverse problem is significantly simplified from trying to estimate \(x = [\sigma_0, \sigma_s, p(\theta_0), p(\theta_1), \ldots, p(\theta_n)]\) to instead solving for \(x = [\sigma_0, \sigma_s, g, \gamma]\) (when using a GK phase function). This is a very significant reduction of the search space for the optimization problem.

Although the three main model properties are the same for most fields working on scattering characterization using radiometric models, there are many different inverse methods for estimating these model properties. An inverse model is typically designed by the following considerations: what is the measured data, which forward model is best suited for simulating the scattered light distribution of interest, what objective function represents the best match of simulations and measurements, and which numerical optimization method can most efficiently minimize the objective function.

In a typical experiment, we measure the combined effect of multiple scattering and absorption events. A way to overcome this issue is using samples with isolated single scattering effects \([NGD06]\]. Such samples, however, are often not available or difficult to produce \([GZB13]\]. When we consider multiple scattering, the type of measurement data is quite essential (Fig. 8). We can choose to measure the total (spectral) signal or the radiant intensity of the transmitted and/or reflected light. Alternatively, we can use a more complete measurement of the scattered spectral radiance. Studies have shown that the fitting accuracy is usually better if a more complete measurement of the scattered light is used \([LLA13,CHCM17]\]. So fitting to a measured BSSRDF, a measured spectral radiance distribution, or multiple camera images \([GZB13,HLC18]\) will provide more accurate results than fitting to radiant intensity or only total transmission/reflection. However, as found in similarity theory \([WPW89]\), two distinct sets of optical properties can produce indistinguishable optical performance. This is potentially an issue, but can also be turned to an advantage \([ZRB14]\).

The choice of the forward model is heavily influenced by the type of measurements being used. The Kubelka-Munk model can be a good match if measurements of the diffuse reflectance are available for a coated paint sample, while the adding-doubling method could be better suited for transmitted and reflected radiant intensity measurements of liquid samples. For measurements with more degrees of freedom, more flexible methods such as Monte Carlo ray tracing make more attractive forward models. An important aspect when choosing a forward model is of course its computational efficiency. Most optimization methods require several iterations to solve Eq. (46), which means that an inefficient/slow forward model significantly influences the time it takes to obtain an estimate of the volume scattering properties.

With a suitable forward model, the next step is to define an objective function that describes how well the simulations for the scattered light distribution approximate the measurements. Generally, a least-squares-like formulation is used for the objective function \([GZB13,MSY09]\), with an adequate choice of weighting factors to avoid overfitting. Objective functions based on calculating a norm between measurements and simulations are also found in practice, especially for inverse adding-doubling methods \([PGW93,LMD14]\). Other types of objective functions which
focus on minimizing the differences between the shape of the simulated and measured scattered distributions have also been described, using the normalized cross-correlation \([L\alpha^*13,C\alpha^*16]\) or the cosine distance \([C\alpha\mu^*17]\). Similarly, when using measurements that can include a significant amount of variance or are very dense, it can be useful to define the objective function so that it minimizes the difference of aggregate data such as a histogram of the measured and simulated distributions \([Y\alpha\chi^*16]\).

The final consideration in developing an inverse model is choosing a numerical optimization method. The choice of optimization method can significantly influence the accuracy of the inverse model. In cases where the search space of the inverse problem is small or can be adequately constrained, it is possible to exhaustively search the entire space \([M\alpha\nu^*09]\), but the search space is more often too large for this approach. Furthermore, the objective function chosen for the inverse problem can return the same value for different combinations of scattering model parameters, that is, the inverse problem can be ill-conditioned. Despite this, convex optimization methods \([N\alpha^*06]\), direct search methods \([W\alpha^*19]\) and metaheuristic methods \([G\alpha\nu^*10]\) have all been successfully used to solve the inverse problem.

The choice of optimization method can be deeply influenced by the type of forward model used, as the objective function can be stochastic (e.g. when using Monte Carlo ray tracing) or deterministic (e.g. using adding-doubling). Trust-region algorithms and other convex optimization methods and direct search methods have successfully been used to solve the inverse problem when using deterministic objective functions \([P\alpha\gamma^*93,L\alpha^*13,L\mu^*14,C\alpha^*16]\). For stochastic objective functions, evaluating first- and second-order derivatives may become infeasible, decreasing the efficiency of least-squares and convex optimization methods. This can be tackled by considering stochastic gradient descent approaches \([G\alpha\nu^*13,Z\alpha\nu^*14,K\alpha\nu^*15,V\alpha\nu^*18]\), for instance.

Optimization algorithms based on metaheuristic principles use an entirely different approach. Different metaheuristic methods such as genetic algorithms \([K\alpha^*20]\), simulated annealing \([Z\alpha\nu^*17,M\alpha^*11]\), or particle swarm methods \([M\alpha^*16]\) impose virtually no constraints on the objective function. Furthermore, they can be parameterized to search the parameter space globally, which can help to minimize the ill-conditioning of the inverse problem albeit at the cost of longer search times.

9.2. Inverse Field Models

When the surface under investigation is patterned with nanostructures, our model must consider the wave nature of light if we desire accurate information regarding the sample. This is done by using field models for the simulation of the optical signal. Conversely, this also means that we can use accurate simulations together with measured data to break the Abbé diffraction limit of conventional microscopes and achieve information about the physical morphology of the surface, provided that the material properties of the sample are known.

This approach is used in the field of scatterometry, the inverse problem is solved to find the physical parameters of a sample based on an optical fingerprint \([M\alpha^*16]\). The technique is utilized in the semiconductor industry, where it provides a fast and accurate measurement of nanostructures. With scatterometry, an overlay accuracy between two lithography steps with an error better than 0.5 nm (within three standard deviations) has been demonstrated with commercial instruments \([S\alpha^*12]\). Often, when scatterometry is used for in-line inspection, a library of simulated optical signals is generated beforehand. This method allows for the large computational burden to be managed by a one time execution. At the production line, the optical signal is measured, and a simple database look up is performed to find the best match between model and data. If sufficient computational power or time is available, the inverse modeling can also be performed by a direct optimization algorithm where a series of simulations is performed. For each simulation, the sample parameters are changed based on the results from the previous simulation. To model the optical response from a sample, several models can be used as described in Section 8. The different models provide varying degrees of flexibility, accuracy, and speed. It is therefore, again, essential to consider which model is best suited for a given problem.

10. Inverse Model Validation and Parameter Extraction

In order to validate optical models, they must be compared with physical measurements. First, when choosing a model, it is important to consider the underlying physics of the measurement. Secondly, the model should describe the data well. We can simulate a physical signal using a model \(f\) that estimates the signal as a function of a set of parameters describing the illuminated sample \(Y_s\) (optical constants of the material, distribution and form of material, surface parameters) and a set of parameters describing the lighting conditions \(Y_L\) (intensity of incoming light, angle of incidence, angle of observation, polarization, wavelength).

The simulated signal is usually compared with measured values \(M\) while incorporating their corresponding measurement uncertainties \(U\). This is done, in a way similar to Eq. (46), by computing the \(\chi^2\) quantity:

\[
\chi^2 = \frac{1}{N} \sum_{j=0}^{N} \left( \frac{M_j - f_j(Y_s, Y_L)}{U_j} \right)^2 ,
\]

(51)

where \(N\) is the number of measurements (the length of the vector \(M\)). A good model function \(f\) finds a minimum \(\chi^2\) when the model input parameters \(Y_s\) and \(Y_L\) describe the physical experiment.

As an example, we look at the in-plane BRDF from a reflective surface characterized by a surface roughness \(R_o\). Here, the measured and is commonly the intensity of the reflected light \(I\) as a function of the incoming angle \(\theta_i\) and the outgoing angle \(\theta_o\). Conventionally, \(\theta_i\) is kept constant, so the model can be evaluated using:

\[
\chi^2 = \frac{1}{N} \sum_{j=0}^{N} \left( \frac{I_j(\theta_i, \theta_o) - f_j(R_o, n_s, \theta_i, \theta_o)}{U_j(I(\theta_i, \theta_o))} \right)^2 .
\]

(52)

The uncertainty \(U(I(\theta_i, \theta_o))\) is usually evaluated using Poisson statistics, which results in \(U(I(\theta_i, \theta_o)) = \sqrt{I(\theta_i, \theta_o)}\). The model \(f(R_o, n_s, \theta_i, \theta_o)\) is a function of the surface roughness \(R_o\) and the refractive index \(n_s\) of the sample. Information about the surface is also commonly obtained by measuring the intensity as a function of wavelength and/or polarization \([R\alpha^*11,M\alpha^*18]\).
The initial guess for the sample parameters $Y_S$ is often found using other techniques such as optical profilometry or atomic force microscopy as well as a priori information regarding the sample. If measurements are available, one may calculate the associated PSD curve, which may be used as a guideline for selecting either the scalar diffraction method or more rigorous methods for the inverse modelling. The scalar diffraction will, in general, be a good method if the PSD curve does not contain large amplitudes from periodic surface structure contributions, and if the ABC model mentioned in Section 8.5 gives a good description of the PSD curve. The scalar diffraction method may also be used as a first attempt if the a priori information justifies its use. However, optimization without a priori knowledge or measurement should be avoided.

Using a standardized chi-square ($\chi^2$) method, one gains relevant statistics for the evaluation of a model. A good model should converge towards the best parameter set, which has a normalised chi-square value close to 1. It may be necessary to replace the scalar diffraction model with more rigorous modelling in the final iterations. By using the physical measurement uncertainties for the weights, one can assure that the model is not drastically biased by noisy measurements. Various ways of minimizing $\chi^2$ exist. We describe a couple of ways of using direct optimization in the following subsections. Other methods for optimization not discussed in detail here include Bayesian approaches, which allow for a priori likelihood distributions of parameters in the model [ZC15, HWS17], and neural networks, where an empirical model is trained from a large amount of data [RMRL02, GRT08, M0NH18].

### 10.1. Levenberg-Marquardt

The Levenberg-Marquardt algorithm is a combination of the gradient descend method and the Gauss-Newton method. Optimization can always be rephrased as a minimization problem for a cost function $h(Y_S)$. In the gradient descend method, the change in $Y_S$ is simply given by the gradient $-\partial h/\partial Y_S$, whereas in the Gauss-Newton method, the value of $Y_S$ is iteratively changed according to

$$Y_{S,i+1} = Y_{S,i} - (\nabla^2 h)^{-1} \nabla^1 h,$$

where $i$ is an iteration parameter. To avoid overshooting, the step size is scaled with respect to the second derivative $\partial^2 h/\partial^2 Y_S$, which is numerically estimated by the approximated Hessian $\nabla^2 h$, where $\nabla$ is the Jacobian of the cost function $h$. This makes intuitive sense when considering that a larger second derivative would correspond to a quickly changing function, and smaller steps are then better.

The Levenberg-Marquardt algorithm is a modification of the Gauss-Newton method called the Levenberg-Marquardt algorithm, which adds a damping term $\zeta$:

$$Y_{S,i+1} = Y_{S,i} - \left( J^T J + \zeta (J^T J) o I \right)^{-1} \nabla h,$$

where $\circ$ is the Hadamard product. A large initial value is typically used for $\zeta$. If a step $t$ causes $h$ to increase, $\zeta$ is increased, otherwise $\zeta$ is decreased to approach the Gauss-Newton method. For small diagonal elements of $J^T J$, larger steps are preferred for speed. Once a solution has been found, we can estimate the uncertainties of the evaluated parameters using

$$U(Y_S) = \sqrt{I o (J^T U^2(M) J)^{-1} I},$$

where $U^2(M)$ is a matrix containing all the squared measurement uncertainties $U^2$ in the diagonal.

Bagher et al. [BSH12] used the Levenberg-Marquardt algorithm to evaluate how well different BRDF models fitted the BRDFs measured by Matusik et al. [MPBM03].

### 10.2. General Least Square

In most models, the input parameters $Y_S$ and $Y_L$ are assumed to be known and exact. In the real world, we are however only able to measure an estimate of the “true” value. These estimated values have corresponding uncertainties, which can be included in the model evaluation by employing a least squares method in its general form [Nie02]. This method however suffers from slow convergence, and it is therefore mostly useful if the model function can be evaluated quickly or if one has a sufficiently good initial guess.

### 10.3. Bootstrap

A common assumption is that the experimental data follows a normal distribution. This is however not always the case. Fully determining the underlying distribution requires measurements of the entire distribution, which ranges from inconvenient to impossible. In a sense, the empirical distribution is the least prejudiced estimate possible of the underlying distribution. This is utilized in the bootstrap method for Efron [Efr82]. The bootstrap method is used to estimate parameters of a large population based on a measured subset of the population. With a set of observations $POP$ from a population, one can make a subset of surrogate observations $pop$ containing a random selection of elements from $POP$. By making several subsets, $pop$, and calculating their averages $\bar{pop}$, one can use these calculated means to describe the entire population using only the subset $POP$ with respect to mean, standard deviation, and confidence intervals.

### 10.4. Hybrid Metrology

As mentioned in Section 7, the model function $f$, should be chosen to contain as much information as possible. However, measurement setups are typically designed to give information about specific parameters (reflected intensity, depolarization, sample structure, sample material properties, etc.). It is therefore ideal to use several different instruments to obtain different properties and incorporate them into the evaluation of $f$ through parameters $Y_S$ and/or $Y_L$.

When the inverse problem does not have a unique best solution, the problem is considered ill-posed [Had02]. The most common way to deal with ill-posed problems is Tikhonov regularization [GH99]. This is also known as “ridge regression” in the field of statistics [HK70] and goes by the name of “weight decay” in modern machine learning [Bis95]. Tikhonov regularization is used to penalize overly complex models and to include additional measurement data in the cost function for model validation. As an example, one could have alternative ways of measuring a subset of the parameters describing the sample, $b \subseteq Y_S$, and (51) becomes

$$\chi^2 = \sum_{i=1}^{n} \left( \frac{b_{m,i} - b_{c,i}}{U_{b,i}} \right)^2 + \frac{1}{N} \sum_{j=0}^{N} \left( \frac{M_j - f_j(Y_S,Y_L)}{U_j} \right)^2,$$

where $M_j = \sum_{i=1}^{n} \left( \frac{b_{m,i} - b_{c,i}}{U_{b,i}} \right)^2 + \frac{1}{N} \sum_{j=0}^{N} \left( \frac{M_j - f_j(Y_S,Y_L)}{U_j} \right)^2$. 

© 2020 The Author(s). Computer Graphics Forum © 2020 The Eurographics Association and John Wiley & Sons Ltd.
where \( n \) is the number of supplementary measurements, \( b_{m,j} \) is the \( j \)th measured value, \( b_{c,j} \) is the \( j \)th calculated value, which is taken to be the value of the corresponding parameter in \( T_S \) used in \( f_j \), and \( U_{b,j} \) is the uncertainty associated with the measurement of the parameter \( b_{m,j} \).

11. Discussion and Conclusions

For several decades, the foci of computer graphics have been shape, animation, light transport, processing hardware, and interaction interfaces. This has led to graphics programs with an ability to render very believable imagery, and to do this not only for motion pictures but also for interactive applications. However, realistic rendering of specific real objects based on their intrinsic optical properties, so-called predictive rendering [UWP06], is still a significant challenge. Setting up a pipeline for acquiring the shape and optical properties of a non-opaque object in order to perform quantitative comparison of photographs with rendered images is not straightforward. The comprehensive pipeline needed to do this for glass objects [SDN+17] clearly underlines that validation of predictive rendering is a challenge. While shape acquisition is one important part of the challenge, acquisition of optical properties is equally important. The key issue is that a full global light transport simulation can only be as accurate as the optical properties provided as input for the rendering. Rusmeier [Rus95] referred to this as the input problem. We have provided a survey of techniques that address this input challenge.

Material appearance is most commonly described by texture images combined with parsimonious analytic reflectance models. When this is the case, the connection to physics is almost exclusively phenomenological (connection to macroscopic observations, but no obvious connection to first principles). The problem in such an approach is that it misses a link to the material microstructure, which usually determines the physical properties of the material. With the advent of additive manufacturing and digital twins of industrial products, a need is emerging for connections between phenomenological graphics techniques and the properties of real world materials. We provide a map of existing work in graphics that brings us closer to material appearance modelling based on the physical microscale or nanoscale properties of real materials. We find that such a map is an excellent outset for addressing the input challenge and thereby building predictive rendering models.

An interesting aspect of predictive rendering is that it enables rendering of synthetic data sets for machine learning. The microstructure of a material is usually a key indicator of the quality of the material. Thus, if we ensure a connection between physical microstructure and the optical properties used in a predictive rendering, we can build a system for vision-based quality control using machine learning based on synthetic data. We provided a review of different models for computing or inferring the collective scattering distribution functions (BSSRDF, BRDF, BTDF, BSDF) of an object. These models are divided into radiometric models and field models. Radiometric models are well suited for efficiently simulating the scattering distribution from samples with known volume scattering properties (forward model) and to extract these properties from measurements (inverse model).

Depending on the scattering properties and geometry of a material, different radiometric models are available to simulate the collective scattering distribution functions. When using these models for the inverse problem, their computational efficiency and stochastic or deterministic nature determines both their suitability and the type of optimization method that is best used. While there are currently many inverse radiometric models available to extract the volume scattering properties of translucent materials, there is limited work done on estimating the surface scattering properties or combined surface and bulk scattering properties of translucent material samples. For this, a hybrid approach that uses field models for surface properties and radiometric models for volume properties could merge the best features from both types in order to extract all relevant optical properties that are needed for realistic rendering of translucent objects.

The field models have been divided into two subcategories called scalar diffraction theory (RR, GHS) and rigorous modelling (FEM, FDTD, aRCW). The scalar diffraction theory is a solution to the Helmholtz equation, which is a 2D partial differential equation. The third dimension (the surface normal) is treated as a mathematical layer of no physical extent. The scalar methods we presented are wide field scattering methods not limited by the paraxial approximation often used in textbooks. The main limitation of these methods is how well the mathematical layer represents the third dimension in the physical 3D problem to be solved. Another limitation of the scalar model is its missing ability to treat the polarization component along the third dimension. These limitations prevent the scalar methods from handling physical problems with typical extent of more than a few wavelengths in the third dimension. The rigorous solutions, on the other hand, are full 3D solutions of Maxwell’s equations which theoretically have no upper limit, and are only limited by the quantum limit, which typically becomes important when the dimensions approach atomic sizes. However, in practice there is also an upper limit on the computational domain that can be handled. This is typically about 100 \( \mu \)m in each dimension.

A multitude of approaches have been invented to solve roughness scattering problems, all with their advantages and disadvantages. Often used methods are based on geometrical optics and single scattering theories, which historically have been popular due to their simplicity. However, there is a natural limit to the scattering problems that these theories can represent, since multiple scattering will occur when the surface roughness amplitude and frequency increases. The upper frequency limit is generally set by the Rayleigh criterion. Several models have been introduced lately in order to deal with multiple scattering in a geometrical optics setting [HdD16, LJJ+18, XH18]. An approach to overcome these issues, while keeping wave effects and a relatively simple formulation, is the GHS method. This is based on the principle of non-paraxial scalar diffraction theory, which significantly improves scattering predictions for large angle scattering compared to single scattering rough surface theories, and has a relatively low computational cost compared to simple scalar methods like RR (low accuracy, no energy conservation) and rigorous methods like FEM, FDTD and aRCW (high computational cost).

The mentioned methods are all forward methods that calculate the scattering output for an object with a given distribution of scat-
References


© 2020 The Author(s)
Computer Graphics Forum © 2020 The Eurographics Association and John Wiley & Sons Ltd.


[FB09] Fang Q., Boas D. A.: Monte Carlo simulation of photon migration in 3D turbid media accelerated by graphics processing units. Optics Express 17, 22 (2009), 20178–20190. 9, 11


injection molded nanostructures at the fabrication line. Metallösungen.

Surface Topography: Metrology and Properties 4, 2 (2016), 023003. 13, 18

Note 910-1. 8

National Bureau of Standards, March 1976, pp. 1–9. NBS Technical Note 910-1. 8


Nie65 NICODEMUS F. E.: Directional reflectance and emissivity of an opaque surface. Applied optics 4, 7 (1965), 767–775. 2, 7


PERIVASAMY V., PRAMANIK M.: Advances in Monte Carlo simulation for light propagation in tissue. *IEEE Reviews in Biomedical Engineering* 10 (2017), 122–135. 11


