Understanding the Masking-Shadowing Function in Microfacet-Based BRDFs

Eric Heitz

Understanding the Masking-Shadowing Function in Microfacet-Based BRDFs

Eric Heitz

RESEARCH REPORT
N° 8468
February 2014
Project-Team Maverick

ISSN 0249-6399   ISRN INRIA/RR--8468--FR+ENG

Understanding the Masking-Shadowing Function in Microfacet-Based BRDFs

Eric Heitz

Project-Team Maverick

Research Report n° 8468 — February 2014 — 47 pages

Abstract: This document is dedicated to answering common questions regarding the masking-shadowing functions in microfacet-based BRDFs. We use the fact that the masking function (or geometric attenuation factor) is constrained by the visible projected area of the microsurface onto the view direction to derive the exact masking function. We introduce the distribution of visible normals from the microsurface, whose normalization factor is the masking function, and we show how the common form of microfacet-based BRDFs emerges from this distribution. The consequence of this is that only exact masking functions ensures correct normalization of microfacet-based BRDFs. Our derivation emphasizes that under the assumptions of common microfacet-based BRDF models, the exact masking function is the generalized form of Smith's masking function. We also discuss the properties of Smith's function used for shadowing and the consequences for the normalization of the BRDF. We review the historical V-cavity model and we show that the underlying surface profile is closer to a normal map than a displacement map. This intuition explains why this non-realistic model is responsible for wrong specular highlights at grazing view angles. Finally, we argue that the insights gained from these observations motivate new research directions in the field of microfacet theory. For instance, we show that masking functions are stretching invariant and we show how this property can be used to derive the masking function for anisotropic microsurface in a straightforward way. We discuss future work such as the incorporation of multiple scattering on the microsurface into BRDF models.

Key-words: microfacet theory, physically based rendering

RESEARCH CENTRE
GRENOBLE – RHÔNE-ALPES

Inovallée
655 avenue de l'Europe Montbonnot
38334 Saint Ismier Cedex

Dérivation et propriétés de la fonction d'ombrage dans les BRDFs à microfacettes

Résumé : Ce document a pour but de répondre à des questions récurrentes concernant la fonction d'ombrage dans les BRDFs à microfacettes. Nous utilisons le fait que la fonction d'ombrage soit contrainte par la surface projetée de la microsurface pour dériver sa forme exacte. Nous introduisons le concept de distribution de normales visibles, dont le coefficient de normalisation est la fonction d'ombrage, et nous montrons comment la BRDF émerge de cette distribution. Notre dérivation montre que, sous les hypothèses habituelles de la théorie des microfacettes, la fonction d'ombrage exacte est celle dérivée par Smith. Nous discutons les propriétés de la fonction de Smith ainsi que ses implications sur la normalisation de la BRDF.

Mots-clés : théorie des microfacettes, rendu physiquement réaliste

Contents

1 Introduction

Microfacet theory was originally developed in the field of optical physics to study scattering on statistical surfaces [BS63]. In the graphics community, we use it to derive physically based BRDFs [CT82, ON94, WMLT07], which are now widely used in both real-time and production rendering. Nowadays, microfacet theory is a fundamental background in computer graphics. For instance, each year at SIGGRAPH, the course on physically based rendering starts with an introduction to microfacet theory [MHH+12, MHM+13]. The goal of a typical microfacet course is to provide the main intuitions derived from the underlying physics as well as other considerations such as flexibility for artistic direction and computational efficiency. The combination of different components of microfacet-based BRDFs offers a wide range of possibilities and the best choice is not always obvious.

What This Document Is About The purpose of this document is to provide new insights and answer longstanding questions concerning the choice of the masking-shadowing function for microfacet-based BRDFs. These questions are answered in the summary-sections 2.4, 4.4, and 6.3. We advise the reader who wants to be spared technical details and go straight to the point to jump directly to these two sections. The rest of the document is dedicated to the reader willing to develop his intuition and understanding of microfacet theory.

What This Document Is Not About We do not introduce new BRDF models; we only discuss commonly used models. We don't advise the reader to use a model rather than another; we aim at providing knowledge on these models to help understand where they come from, what they are doing, and what we can expect from them. We don't recall their implementation or usage with specific rendering techniques since they are already used in the CG community; we focus on understanding their physical properties.

Ideas and Organization The ideas presented in this document were strongly inspired by two previous works:

- Ashikhmin et al. observe that the *visible projected area* is a quantity that is conserved from the macrosurface to the microsurface [APS00]. They use this knowledge to derive the general equation for an exact masking term, which ensures correct normalization and energy conservation. Their masking term is presented in its integral form and they do not derive a closed form. Instead, they precompute it numerically and store it in a look-up table.

- Ross et al. propose a study of the reflectance of the sea [RDP05]. They model the sea with a Gaussian rough surface (Beckmann distribution) and compute a normalized BRDF incorporating Smith's masking and shadowing functions [Smi67]. During the derivation, they observe that on Gaussian surfaces, the normalization coefficients of the BRDF and Smith's functions have similar expressions and simplify out. They note that this property is convenient for computational purposes, but they do not provide a physical reason as to why this is.

In Section 2, we make the connection between these two previous works. We provide a full analytical derivation of Ashikhmin et al.'s equation, whose solution turns out to be the generalized form of Smith's masking function, as observed by Ross et al. in the special case of Gaussian surfaces. This generalized form was already derived by Brown [Bro80] and recently introduced into the CG community by Walter et al. [WMLT07]. While our derivation does not provide a new result, it has the advantage of emphasizing the fact that the result is exact rather than approximate, and shows how masking is related to the concept of the visible projected area on arbitrary stochastic surfaces.

In Section 3, we demonstrate for the first time the *stretching invariance* property of the masking function. We show how it can be used to make a trivial derivation of the masking functions for several anisotropic normal distributions. This eases generalization to anisotropy of several previous results and spares heavy mathematical derivations.

In Section 4, we introduce the *distribution of visible normals* and we show how common BRDF models can be derived from this distribution. We recall that the reason why these models require shadowing is that they only model the first scattering event occuring on the microsurface. Common microfacet-based BRDFs do not model multiple scattering and are not normalized for this reason, i.e. they do not integrate to exactly 1. Starting from this observation, we propose a normalization test that we call the *Weak White Furnace Test*, that can be used to verify that common microfacet-based BRDFs are well designed, even if they only model the first scattering event.

In Section 5, we discuss the properties of Smith's function used for shadowing and we recall several masking-shadowing models that handle different types of correlation.

In Section 6, we review the properties of V-cavity surfaces and the classical associated masking function, which was historically the first one used in CG.

Finally, in Section 7, we discuss some of the limitations of the current microfacet framework and we propose possibilities for promising future work based on the insights gained in this investigation.

2 Derivation of the Masking Function

In this section we recall how the projected area of the microsurface can be used to derive the exact masking function [APS00]. We start by defining the concept of projected area (2.1). Applying it in the framework of microfacet theory (2.2) gives a new equation that we use to derive the masking function (2.3).

2.1 Measuring Radiance on a Surface

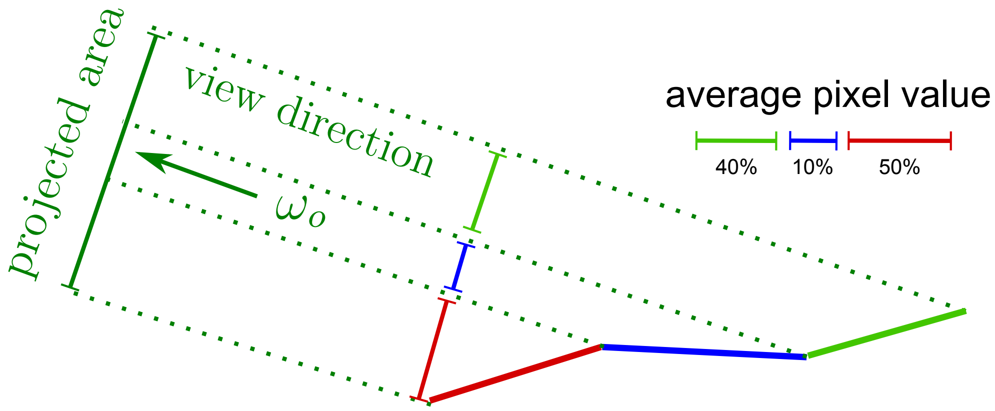


Figure 1: Filtering integral.

Radiance is the energy density arriving from a solid angle. When multiple surfaces project onto the same pixel, the total radiance $L(\omega_o)$ measured for the pixel is the sum of the radiances of each surface $L(\omega_o, x)$ weighted by their projected area on the pixel footprint (as shown in Figure 1):

$$L(\omega_o) = \frac{\int \text{projected area}(x)\, L(\omega_o, x)\, dx}{\int \text{projected area}(x)\, dx}. \tag{1}$$

In computer graphics, BRDFs model the radiance measured in the view direction, i.e. the average energy arriving over the pixel footprint. The projected area of each surface in the view direction is a view-dependent weighting factor and the sum of the projected areas, $\int \text{projected area}(x)\, dx$, is an important normalization coefficient. Dividing by this sum ensures that the density of energy is preserved.

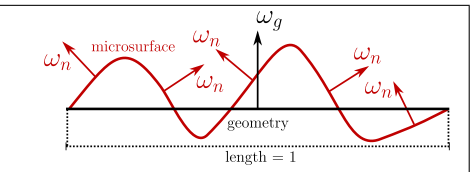
In the following section we will see that, in accordance with microfacet theory, the microfacets are also weighted by their projected area, and that the masking function (or geometric attenuation factor) is the normalization coefficient required for energy preservation.

2.2 Microfacet Projections

Table 1 illustrates different kinds of projection in microfacet theory:

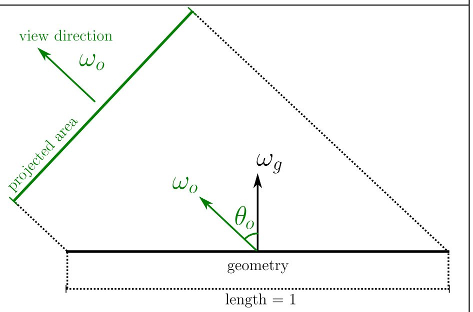
(a) The projection of the microfacets onto the geometry is a unit patch:

$$\int_\Omega \langle \omega_n, \omega_g \rangle \, D(\omega_n) \, d\omega_n = 1$$



(b) Projected area of the unit patch onto the view direction:

projected area $= \omega_o \cdot \omega_g = \cos\theta_o$



(c) The visible projected area of the microsurface onto the view direction:

projected area $= \displaystyle\int_\Omega G_1(\omega_o, \omega_n) \, \langle \omega_o, \omega_n \rangle \, D(\omega_n) \, d\omega_n$

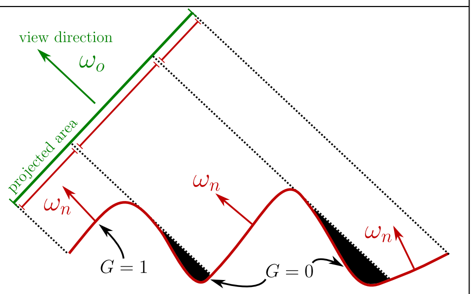


Table 1: Projections in microfacet theory.

$\omega_g = (0, 0, 1)$	geometric normal
$\omega_o = (x_o, y_o, z_o)$	view direction
θ_o	view angle ($\cos\theta_o = z_o$)
$\omega_n = (x_n, y_n, z_n)$	microfacet normal
θ_n	normal angle ($\cos\theta_n = z_n$)
$D(\omega_n)$	normal distribution function
$G_1(\omega_o, \omega_n)$	masking function
$\omega \cdot \omega$	dot product
$\lvert\omega \cdot \omega\rvert$	absolute value of the dot product
$\langle \omega, \omega \rangle$	clamped dot product
$\chi^+(a)$	Heaviside function: 1 if $a > 0$ and 0 if $a \leq 0$

Table 2: Notation.

(a) Projection onto the Geometry The area of the microfacets projected onto the geometry is the area of the unit patch supporting the microfacets (Table 1(a)), whose area is 1 by convention. Hence, while the normal distribution D is not necessarily normalized on the sphere ($\int_\Omega D(\omega_n)\, d\omega_n \neq 1$), its projection onto the geometry is normalized:

$$\int_\Omega \langle \omega_n, \omega_g \rangle\, D(\omega_n)\, d\omega_n = 1. \tag{2}$$

(b) Projected Area of the Unit Patch The unit patch on which the microfacets are defined is a unit planar element and its projected area onto the view direction (Table 1(b)) is the cosine of the view angle θ_o:

$$\boxed{\text{projected area} = \omega_o \cdot \omega_g = \cos\theta_o} \tag{3}$$

(c) Visible Projected Area of the Microsurface The projected area of the unit patch onto the view direction is also the *visible* projected area of the microsurface. It is the sum of the visible projected area of each microfacet. A microfacet, with normal ω_n, has a projected area that is proportional to its presence over the microsurface (given by distribution $D(\omega_n)$) and to the geometric projection factor $\langle \omega_n, \omega_o \rangle$. Note that here we use the clamped dot product $\langle -, - \rangle$ because backface-culled microfacets are not visible. Also, microfacets occluded by the microsurface do not contribute to the projected area and must be removed from the sum. This is achieved by multiplying by a masking function $G_1(\omega_o, \omega_n)$ that ranges in $[0, 1]$ and gives the ratio of microfacets with normal ω_n that are visible in the view direction ω_o. Occluded microfacets with normal ω_n are thus discarded (Table 1(c)). Finally, the projected area of the microsurface onto the view direction is given by:

$$\boxed{\text{projected area} = \int_\Omega G_1(\omega_o, \omega_n)\, \langle \omega_o, \omega_n \rangle\, D(\omega_n)\, d\omega_n}. \tag{4}$$

2.3 Masking Function G_1

In this section we use the equations related to the projected area to introduce a constraint on the masking function. Combined with the usual assumption that masking and normal orientation are independent, this constraint leads to the integral form of the exact masking function G_1 presented by Ashikhmin et al. [APS00].

A Constraint on the Masking Function Table 1 emphasizes a fundamental property of microfacet theory: the visible projected area of the microfacets from Equation (4) is exactly the projected area of the geometric patch given in Equation (3). This equivalence imposes a constraint on the masking function, which is formalized by the following equation:

$$\cos\theta_o = \int_\Omega G_1(\omega_o, \omega_n)\, \langle \omega_o, \omega_n \rangle\, D(\omega_n)\, d\omega_n, \tag{5}$$

However, this is still not sufficient to entirely determine G_1, since for a fixed view direction ω_o, the masking function is a continuous 2D function ($G_1(\omega_o, \omega_n)$ is defined for each normal).

Normal/Masking Independence A classic assumption of microfacet theory is that the probability $G_1(\omega_o, \omega_n)$ given by the masking function is independent of the normal orientation ω_n for normals that are not backface-culled ($\langle \omega_o, \omega_n \rangle > 0$). The intuition behind this assumption is that the normal ω_n is a *local* property of the microfacet, while the potential occlusion responsible for masking occurs elsewhere on the surface and is thus a *non-local* property of the microfacet. This is why it is reasonable to suppose that they are not related. It is also important to note that this independence is only true for nonbackface-culled normals. Indeed, backface-culled normals ($\langle \omega_n, \omega_o \rangle < 0$) are always masked by themselves. In this case, the probability of masking is local. The normal/masking independence allows us to separate the integral on the part of the hemisphere associated with nonbackface-culled normals:[1]

$$\begin{aligned}\int_\Omega G_1(\omega_o, \omega_n) \langle \omega_o, \omega_n \rangle D(\omega_n)\, d\omega_n &= \frac{\int_\Omega \chi^+(\omega_o \cdot \omega_n)\, G_1(\omega_o, \omega_n)\, d\omega_n}{\int_\Omega \chi^+(\omega_o \cdot \omega_n)\, d\omega_n} \int_\Omega \langle \omega_o, \omega_n \rangle D(\omega_n)\, d\omega_n \\ &= \bar{G}_1^+(\omega_o) \int_\Omega \langle \omega_o, \omega_n \rangle D(\omega_n)\, d\omega_n, \end{aligned} \tag{6}$$

where:

$$\bar{G}_1^+(\omega_o) = \frac{\int_\Omega \chi^+(\omega_o \cdot \omega_n)\, G_1(\omega_o, \omega_n)\, d\omega_n}{\int_\Omega \chi^+(\omega_o \cdot \omega_n)\, d\omega_n} \tag{7}$$

without the second argument is the average masking function for the nonbackface-culled normals from view direction ω_o.

Integral Form of the Masking Function By combining Equations (5) and (6), we get:

$$\cos\theta_o = \bar{G}_1^+(\omega_o) \int_\Omega \langle \omega_n, \omega_o \rangle D(\omega_n)\, d\omega_n,$$

and the masking function for the visible normals from view direction ω_o is then given by:

$$\bar{G}_1^+(\omega_o) = \frac{\cos\theta_o}{\int_\Omega \langle \omega_o, \omega_n \rangle D(\omega_n)\, d\omega_n}. \tag{8}$$

Since we derived the average masking term for nonbackface-culled normals and we know that $G(\omega_o, \omega_n) = 0$ for backface-culled normals, we can define the masking term for all normals by multiplying with a Heaviside function:

$$\boxed{G_1(\omega_o, \omega_n) = \chi^+(\omega_o \cdot \omega_n) \frac{\cos\theta_o}{\int_\Omega \langle \omega_o, \omega_n \rangle D(\omega_n)\, d\omega_n}} \tag{9}$$

This is the integral form of the exact masking function under normal/masking independence presented by Ashikhmin et al. [APS00]. They use this integral expression to precompute the masking function and store it in a look-up table, for rendering.

[1] We use the fact that when two functions $a(x)$ and $b(x)$ are not correlated on domain X, they can be separately integrated: $\int_X a(x)\, b(x)\, dx = \int_X a(x)\, dx \int_X b(x)\, dx / \int_X dx$. We use this property for $x = \omega_n$, $a(x) = G_1(\omega_o, \omega_n)$, $b(x) = \langle \omega_o, \omega_n \rangle D(\omega_n)$ and X is the domain of the hemisphere where normals are not backface-culled ($\chi^+(\omega_o \cdot \omega_n) > 0$) and its area is $\int_X dx = \int_\Omega \chi^+(\omega_o \cdot \omega_n)\, d\omega_n$.

Smith's Masking Function By changing the integration domain from normal to slope space (we provide the detailed derivation in Appendix A) we show that Equation (9) can be rewritten:

$$\boxed{G_1(\omega_o, \omega_n) = \frac{\chi^+(\omega_o \cdot \omega_n)}{1 + \Lambda(\omega_o)}} \tag{10}$$

where $\frac{1}{1+\Lambda(\omega_o)}$ is the generalized form of Smith's masking function [Bro80, WMLT07], for which closed-form solutions are available for many stochastic surfaces, as shown in Section 3. Note that the only approximation we have made so far is the assumption of normal/masking independence, which is already made by microfacet-based BRDF models. Therefore, under this common assumption, Smith's masking function is exact.

Smith's Averaged Masking Functions Smith derived the masking function averaged over different quantities of the microsurface, such as the heights and the normals [Smi67]. The masking function $G_1(\omega_o, \omega_n)$ presented in Equation (10) is the form averaged over the heights of the microsurface and is the one that must be used in the BRDF. Indeed, since the heights are independent from what matters in the BRDF, we would just average over them. However, the normals are not all processed in the same way: backface-culled normals are not considered. In a BRDF model, only what is visible to the external viewer matters, because only radiance that can be measured by this viewer matters. If something exists on the surface but is not visible, then it won't be included in the BRDF. In a BRDF problem we are actually interested in the question: "*What proportion of nonbackface-culled normals are masked?*"

So why did Smith derive a normal-averaged form of his masking function if it is not useful for BRDFs? He actually wanted to answer the question: "*What proportion of normals are masked?*" The answer to this second question is important when studying properties intrinsic to the surface in other physics problems, but not with BRDFs.

This leads us to an important observation: since we only integrate radiance over what is visible, we have to ensure that we normalize over this domain rather than the entire microsurface. This is the topic of Section 4.

2.4 Summary

A frequently asked question concerning the masking function is: "*Among the different masking functions (or geometric attenuation factors), which one should I use? Are they all physically correct?*"
The typical answer to this question is: "*It is preferable to use Smith's masking function, because it depends on the normal distribution.*"

In this section, we showed that this answer is correct but the reason invoked is wrong, by developing the following ideas:

- The projected visible area of the microsurface equals the projected area of the macrosurface onto any projection direction.
- The masking function is constrained by this equality.
- Common microfacet-based BRDFs assume that the orientation of visible normals is independent of the probability of masking.

- Under this assumption, the masking function is completely determined; its exact form can be derived and is the generalized form of Smith's masking function.

The point here is that the reason why one has to choose Smith's masking function is not because it is a physically plausible approximation parametrized by the normal distribution. The real reason to choose it is that Smith's formula is the exact masking function under the assumption of the model (i.e., normal/masking independence). The fact that it is physically plausible and is parametrized by the normal distribution are not directly the reasons to choose it, but are some of the expected side effects of making the right choice.

Still, if we were to compare the analytical function with measured data, we would find that the predictions of the model are not exact. Indeed, Smith compared his formula to real-world measurements and discovered that it was a good approximation, but an approximation nonetheless. However, the approximation does not reside in his derivation because, within the framework of his model, his formula is exact. Instead, it resides in the description of real-world surfaces with statistical models (e.g. Gaussian statistics), and in the assumption of normal/masking independence.

3 Stretching Invariance of the Masking Function

In this section we investigate the invariance property of the masking function and of the slope distribution when the configuration is stretched. We use this knowledge to derive the masking functions for shape-invariant anisotropic distributions.

3.1 Masking Probability Invariance

Figure 2 shows the effect of stretching a 1D configuration with masking on a microsurface with a given view direction. Stretching the configuration is like stretching the picture, i.e. one dimension is multiplied by a constant factor. This operation does not change the topology of the configuration: after stretching, occluded rays are still occluded and non-occluded rays are still non-occluded. This is a key property: the masking probability is invariant to configuration stretching, where all the slopes involved in the configuration are scaled at the same time. This includes the slopes of the microsurface and the slope associated to the view direction. They are all scaled by the inverse of the stretching factor. The slope distribution width is thus also stretched by the inverse stretching factor.

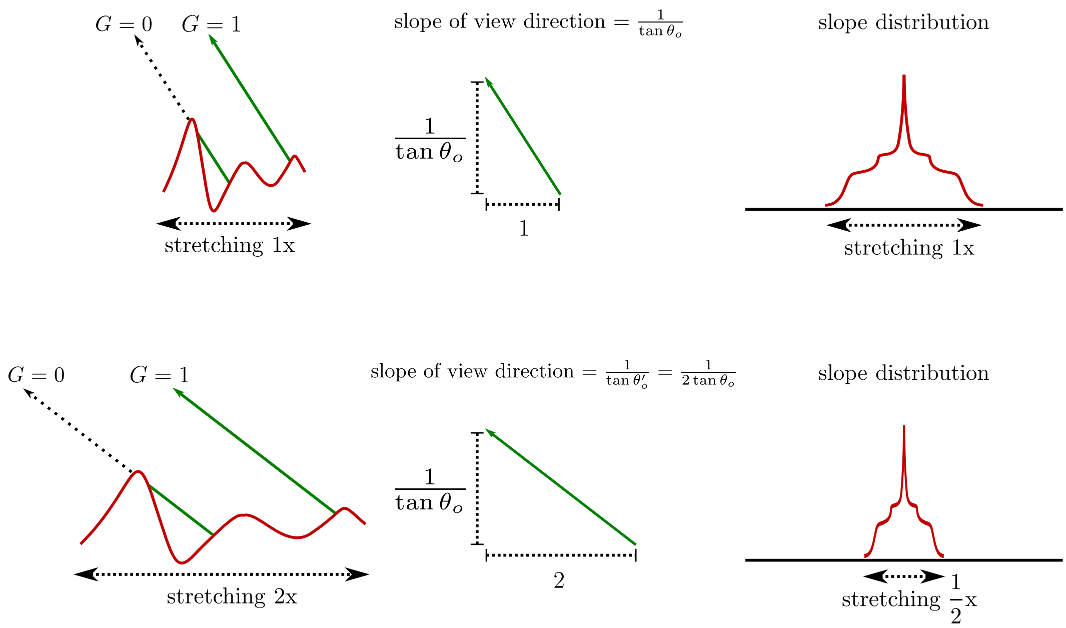


Figure 2: Stretching a 1D configuration by a factor 2 does not change masking probability G, but all the slopes of the configuration are downscaled by a factor $\frac{1}{2}$. This includes the slopes of the microsurface as well as the slope associated with the view direction.

3.2 Shape Invariance of the Distribution

Several isotropic parametric slope distributions depend on a roughness parameter α, where changing α is equivalent to stretching the distribution without changing its shape. This is the case when the slope distribution depends only on the ratio $\frac{\tan\theta_n}{\alpha}$ between the slope amplitude $\tan\theta_n$

of a normal of angle θ_n and roughness parameter α. We will now call this property *shape invariance* of the slope distribution, because distributions that exhibit this property have always the same shape and are only stretched by the roughness parameter. As shown in Figure 2, with shape-invariant slope distributions, stretching the configuration is equivalent to scaling the roughness parameters and the slope of the view vector by the same factor. It implies that the masking function depends only on the ratio $a = \frac{1}{\alpha \tan\theta_o}$, where $\frac{1}{\tan\theta_o}$ is the slope of the view direction[2]. Beckmann and GGX distributions are shape invariant and this is why their associated functions Λ depend only on a.

Beckmann Distribution

$$D(\omega_n) = \frac{\chi^+(\omega_n \cdot \omega_g)}{\pi\alpha^2\cos^4\theta_n} \exp\left(-\frac{\tan^2\theta_n}{\alpha^2}\right),$$
$$\Lambda(\omega_o) = \frac{\operatorname{erf}(a) - 1}{2} + \frac{1}{2\,a\sqrt{\pi}} \exp\left(-a^2\right).$$

where $a = \frac{1}{\alpha\tan\theta_o}$. Walter et al. [WMLT07] propose an accurate rational approximation for $G_1(\omega_o) = \frac{1}{1+\Lambda(\omega_o)}$, which we can use to approximate $\Lambda(\omega_o)$ (via $\Lambda(\omega_o) = \frac{1-G_1(\omega_o)}{G_1(\omega_o)}$):

$$\Lambda(\omega_o) \approx \begin{cases} \frac{1-1.259a+0.396a^2}{3.535a+2.181a^2} & \text{if } a < 1.6 \\ 0 & \text{otherwise.} \end{cases}$$

GGX Distribution

$$D(\omega_n) = \frac{\chi^+(\omega_n \cdot \omega_g)}{\pi\alpha^2\cos^4\theta_n\left(1 + \frac{\tan^2\theta_n}{\alpha^2}\right)^2},$$
$$\Lambda(\omega_o) = \frac{-1 + \sqrt{1 + \frac{1}{a^2}}}{2}.$$

where $a = \frac{1}{\alpha\tan\theta_o}$.

Shape variant distributions Note that not all distributions are shape invariant. For instance, the Phong distribution is not because it does not depend on ratio $\frac{\tan\theta_n}{\alpha}$. As the roughness changes, the shape of the Phong distribution changes.

3.3 Deriving Masking Functions for Shape-Invariant Anisotropic Distributions

The same shape-invariant distributions can be anisotropic if the shape is stretched with direction dependent factors. The slopes are weighted separately in each direction

$$a = \tan\theta_n \sqrt{\frac{\cos^2\phi_n}{\alpha_x^2} + \frac{\sin^2\phi_n}{\alpha_y^2}},$$

where $\tan\theta_n\cos\phi_n$ and $\tan\theta_n\sin\phi_n$ are the slopes and α_x and α_y are the stretching coefficients of the distribution in the x- and y-axis respectively.

[2] For a given direction with angle θ, the slope of the direction is $\frac{1}{\tan\theta_o}$ and should not be mistaken with the slope $\tan\theta$ of a microfacet orthogonal to this direction.

Figure 3 shows how isotropic shape-invariant distributions can be transformed into anisotropic distributions by stretching the surface. Reciprocally, any configuration with an anisotropic distribution can be transformed back to a configuration with an isotropic distribution.

We use this property to derive the masking functions of anisotropic distributions. We start from a configuration with a shape-invariant anisotropic distribution with parameters α_x and α_y and a view vector $\omega_o = (x_o, y_o, z_o)$. By stretching the x-axis direction by a factor $\frac{\alpha_x}{\alpha_y}$ the surface roughness becomes:

$$\begin{aligned} \alpha'_x &= \alpha_x \frac{\alpha_y}{\alpha_x} \\ &= \alpha_y, \\ \alpha'_y &= \alpha_y. \end{aligned}$$

The stretched surface is isotropic with roughness α_y and the view vector and its slope after stretching are:

$$\begin{aligned} \omega'_o &= (\frac{\alpha_x}{\alpha_y} x_o, y_o, z_o) \\ &= (\frac{\alpha_x}{\alpha_y} \cos\phi_o \sin\theta_o, \sin\phi_o \sin\theta_o, \cos\theta_o), \\ \frac{1}{\tan\theta'_o} &= \frac{z_o}{\sqrt{\frac{\alpha_x^2}{\alpha_y^2} x_o^2 + y_o^2}} \\ &= \frac{1}{\sqrt{\frac{\alpha_x^2}{\alpha_y^2} \cos^2\phi_o + \sin^2\phi_o} \tan\theta_o}. \end{aligned}$$

The masking function of an isotropic distribution depends only on the ratio $a = \frac{1}{\alpha \tan\theta_o}$ and since $\alpha = \alpha_y$ the ratio of the stretched surface is:

$$\begin{aligned} a' &= \frac{1}{\alpha_y \tan\theta'_o} \\ &= \frac{1}{\alpha_y \sqrt{\cos^2\phi_o \frac{\alpha_x^2}{\alpha_y^2} + \sin^2\phi_o} \tan\theta_o} \\ &= \frac{1}{\sqrt{\cos^2\phi_o \alpha_x^2 + \sin^2\phi_o \alpha_y^2} \tan\theta_o} \\ &= \frac{1}{\alpha_o \tan\theta_o}, \end{aligned}$$

where

$$\alpha_o = \sqrt{\cos^2\phi_o \alpha_x^2 + \sin^2\phi_o \alpha_y^2} \tag{11}$$

is the *roughness projected in the view direction.* This shows that masking functions associated to anisotropic shape-invariant slope distributions are the masking function of the isotropic distributions parametrized by the roughness of the anisotropic surface projected in the view direction. We use this property to derive the masking functions for the anisotropic Beckmann and GGX distributions.

Anisotropic Beckmann Distribution

$$D(\omega_n) = \frac{\chi^+(\omega_n \cdot \omega_g)}{\pi\alpha_x\alpha_y\cos^4\theta_n}\exp\left(-\tan^2\theta_n(\frac{\cos^2\phi_n}{\alpha_x^2}+\frac{\sin^2\phi_n}{\alpha_y^2})\right),$$
$$\Lambda(\omega_o) = \frac{\operatorname{erf}(a)-1}{2}+\frac{1}{2\,a\sqrt{\pi}}\exp\left(-a^2\right).$$

where $a = \frac{1}{\alpha_o \tan\theta_o}$ and α_o is defined in Equation 11. The approximation of Λ for the isotropic Beckmann distribution can be used as well.

Anisotropic GGX Distribution

$$D(\omega_n) = \frac{\chi^+(\omega_n \cdot \omega_g)}{\pi\alpha_x\alpha_y\cos^4\theta_n\left(1+\tan^2\theta_n(\frac{\cos^2\phi_n}{\alpha_x^2}+\frac{\sin^2\phi_n}{\alpha_y^2})\right)^2},$$
$$\Lambda(\omega_o) = \frac{-1+\sqrt{1+\frac{1}{a^2}}}{2}.$$

where $a = \frac{1}{\alpha_o \tan\theta_o}$ and α_o is defined in Equation 11.

3.4 More Generalization

Arbitrary Shape-Invariant Distributions An important property of shape-invariant distributions is that all the information required for the masking function is contained in the same 1D function Λ for any roughness or anisotropy. Thus, if Λ is available, it can be used for an entire class of parametric distributions with varying roughness and anisotropy. One can easily design its own anisotropic normal distribution by choosing an arbitrary 1D function f and set:

$$D(\omega_n) = cf\left(\tan^2\theta_n(\frac{\cos^2\phi_n}{\alpha_x^2}+\frac{\sin^2\phi_n}{\alpha_y^2})\right).$$

where c would be the normalization coefficient of the distribution. The associated 1D function $\Lambda(\frac{1}{\alpha_o\tan\theta_o})$ can be numerically precomputed and tabulated or fitted with a rational polynomial, as Walter et al. did for the Beckmann distribution.

Non Axis-Aligned Stretching The stretching operation does not need to be axis-aligned. The general stretching in slope space can be redefined with a quadric. Let be Q a symmetric positive-definite matrix[3]:

$$Q^{-1} = \begin{bmatrix} \alpha_x^2 & r_{xy}\alpha_x\alpha_y \\ r_{xy}\alpha_x\alpha_y & \alpha_y^2 \end{bmatrix},$$

and r_{xy} is the correlation coefficient of the stretching in the x- and y-axis. The quadric Q defines a scalar product and a norm in the 2D euclidean space of the slopes:

$$\begin{aligned} ||\tilde{n}||^2 &= \sqrt{\langle\tilde{n},\tilde{n}\rangle} \\ &= \sqrt{\tilde{n}^T Q \tilde{n}}, \end{aligned}$$

[3]In the specific case of the Beckmann distribution, $\Sigma = \frac{1}{2}Q^{-1}$ is the covariance matrix of the Gaussian slope distribution.

where the 2D vector $\tilde{n} = \tan\theta_n(\cos\phi_n, \sin\phi_n)$ is the slope associated to a normal ω_n. The norm of the slope $||\tilde{n}||$ describes the stretching that occurs in slope space and is the argument of the distribution D. We gave the formulas for the norm and the projected roughness in the view direction in the case where $r_{xy} = 0$:

$$\begin{aligned}
||\tilde{n}||^2 &= \tan\theta_n(\cos\phi_n, \sin\phi_n)^T Q \tan\theta_n(\sin\phi_n, \cos\phi_n) \\
&= \tan^2\theta_n \left(\frac{\cos^2\phi_n}{\alpha_x^2} + \frac{\sin^2\phi_n}{\alpha_y^2} \right), \\
\alpha_o^2 &= \cos^2\phi_o \alpha_x^2 + \sin^2\phi_o \alpha_y^2.
\end{aligned}$$

In the general case with correlation where $r_{xy} \neq 0$ we have instead:

$$\begin{aligned}
||\tilde{n}|| &= \tan\theta_n(\cos\phi_n, \sin\phi_n)^T Q \tan\theta_n(\sin\phi_n, \cos\phi_n) \\
&= \tan^2\theta_n \left(\frac{\cos^2\phi_n \alpha_y^2 + \sin^2\phi_n \alpha_x^2 - 2\cos\phi_n \sin\phi_n r_{xy}\alpha_x\alpha_y}{\alpha_x^2\alpha_y^2 - r_{xy}^2\alpha_x^2\alpha_y^2} \right), \\
\alpha_o^2 &= \cos^2\phi_o \alpha_x^2 + \sin^2\phi_o \alpha_y^2 + 2\cos\phi_o \sin\phi_o r_{xy}\alpha_x\alpha_y.
\end{aligned}$$

For instance, in LEADR mapping, a correlated Beckmann distribution is used [DHI+13]. Note that setting the correlation coefficient $r_{xy} \in [-1, 1]$ to non-zero values affects the constant normalization factor of distribution D.

Vertical Shearing and Non-Centered Distributions Figure 4 shows that the masking function is also invariant by vertical shearing. Applying a vertical shear on the configuration is equivalent to increasing all the slopes of the configuration by a constant value. As before, this includes the slopes of the microsurface and the slope associated to the view direction. We call the *mesosurface* the average plane generated by the microsurface and it is represented in blue in the figure. The slope of the mesosurface is the average slope of the microsurface $\bar{\tilde{n}} = (\bar{x}_{\tilde{n}}, \bar{y}_{\tilde{n}})$ and corresponds to where the distribution of slopes is centered around. Typically, the distribution of slopes is centered around 0. It means that the mesosurface is aligned with the macrosurface. However, this assumption is wrong when the macrogeometry is amplified with another high-frequency representation. The very purpose of bump maps, normal maps or displacement maps is to generate a mesonormal by perturbating the macronormal. For instance, in Olano and Baker's LEAN mapping [OB10], a multi-scale non-centered Gaussian slope distribution is used. In this case, the distribution of slopes is almost never centered around 0. If the rendering is physically based, one has to use a masking function extended to non-centered distributions to make sure that everything is still well-defined. Fortunately, the vertical shearing invariance shows that the masking function of a non-centered microsurface is the same as the masking function of a centered microsurface with offsetted slopes. This property was used in LEADR mapping [DHI+13] where microfacet theory is extended to non-centered distributions. To account for non-centering, one has to include the offset in the computation of the argument of distribution D and of the factor

a for the masking function:

$$
\begin{aligned}
||\tilde{n}|| &= (\tan\theta_n(\cos\phi_n, \sin\phi_n) - (x_{\tilde{n}}, y_{\tilde{n}}))^T Q (\tan\theta_n(\cos\phi_n, \sin\phi_n) - (x_{\tilde{n}}, y_{\tilde{n}})) \\
&= \frac{(\tan\theta_n \cos\phi_n - x_{\tilde{n}})^2 \alpha_y^2}{\alpha_x^2\alpha_y^2 - r_{xy}^2\alpha_x^2\alpha_y^2} + \frac{(\tan\theta_n \sin\phi_n - y_{\tilde{n}})^2 \alpha_x^2}{\alpha_x^2\alpha_y^2 - r_{xy}^2\alpha_x^2\alpha_y^2} \\
&- 2\frac{(\tan\theta_n \cos\phi_n - x_{\tilde{n}})(\tan\theta_n \sin\phi_n - y_{\tilde{n}}) r_{xy}\alpha_x\alpha_y}{\alpha_x^2\alpha_y^2 - r_{xy}^2\alpha_x^2\alpha_y^2}, \\
a &= \frac{\frac{1}{\tan\theta_o} - (\cos\phi_o \bar{x}_{\tilde{n}} + \sin\phi_o \bar{y}_{\tilde{n}})}{\alpha_o}, \\
\alpha_o^2 &= \cos^2\phi_o \alpha_x^2 + \sin^2\phi_o \alpha_y^2 + 2\cos\phi_o \sin\phi_o r_{xy}\alpha_x\alpha_y.
\end{aligned}
$$

Note that vertical shearing does not affect the projected roughness α_o^2 and the normalization factor of the distribution. This is intuitive because stretching changes the shape of the distribution, and thus the roughness, while shearing offsets the distribution without changing its shape.

Note also that this property is available only in the slope space and does not hold in the space of the normals: offsetting the slope distribution is not equivalent to rotating the average normal and keeping the same roughness. This is because the rotation operator in the space of the normals is very non-linear in the space of the slopes.

Another important thing for non-centered distributions is that the visible projected area has to be computed from the mesonormal. The factor $\cos\theta_o$ in the BRDF must be replaced by the projected area of the mesosurface, which is $\frac{\omega_{\bar{n}}\cdot\omega_o}{\omega_{\bar{n}}\cdot\omega_g}$, where $\omega_{\bar{n}}$ is the normal of the mesosurface. In the case where the mesosurface is the macrosurface we have $\omega_{\bar{n}} = \omega_g$ and we get back to $\frac{\omega_{\bar{n}}\cdot\omega_o}{\omega_{\bar{n}}\cdot\omega_g} = \frac{\cos\theta_o}{1}$, so this is consistent. More details are available in the LEADR mapping paper.

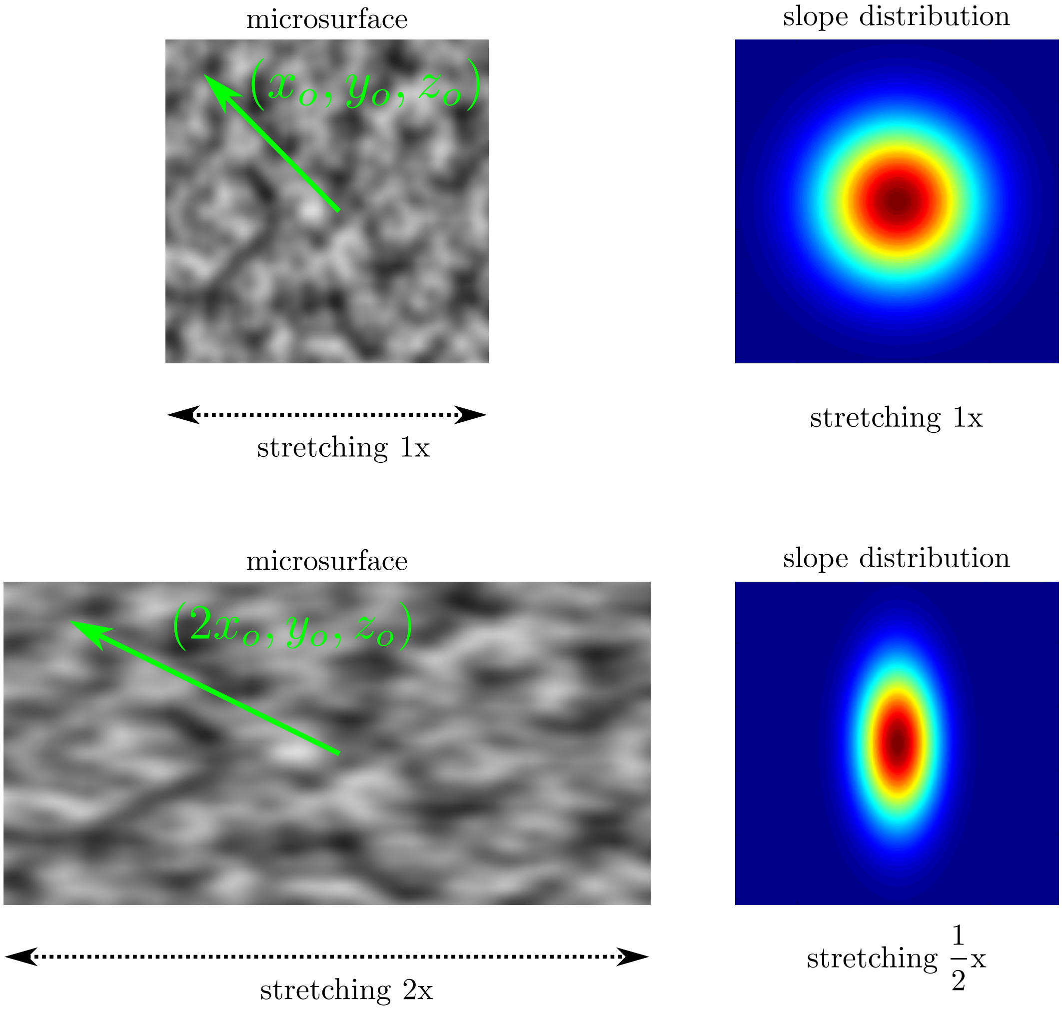


Figure 3: Stretching a 2D configuration by a factor 2. The coordinates of the view vector are stretched as well.

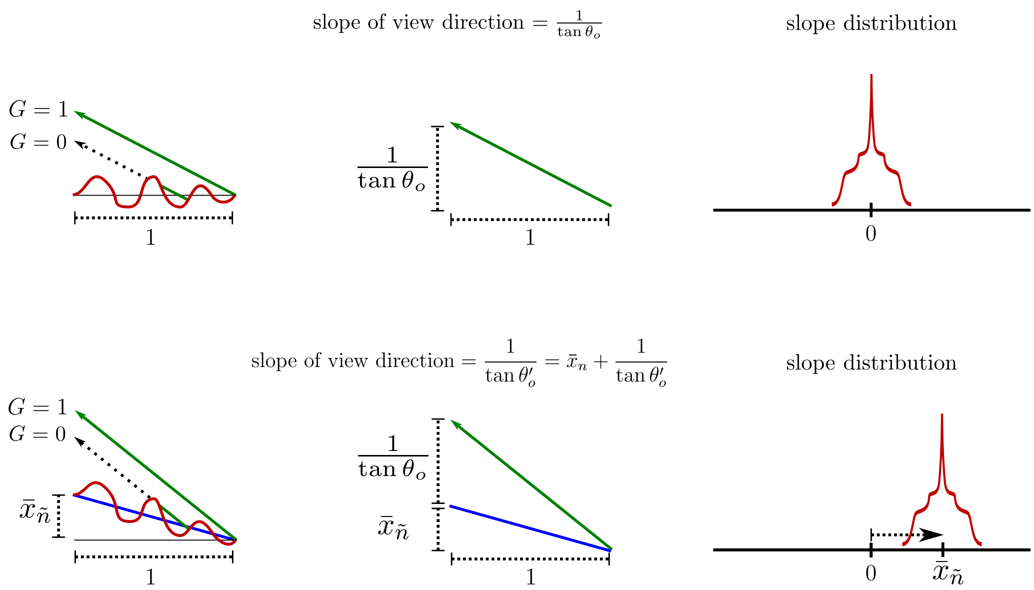


Figure 4: Vertical shearing of a 1D configuration. It does not change masking probability G, but all the slopes of the configuration are increased by a constant factor $x_{\tilde{n}}$. This includes the slopes of the microsurface as well as the slope associated with the view direction. The slope distribution is shifted by an offset $\bar{x}_{\tilde{n}}$ and is no longer centered around 0.

4 Microfacet-Based BRDFs

In this section, we define the distribution of visible normals (4.1) and we show how specular microfacet models are constructed upon this distribution (4.2). We show that Smith's masking function is the normalization coefficient of the distribution of visible normals and we discuss the link with energy conservation for BRDFs constructed from this distribution (4.3).

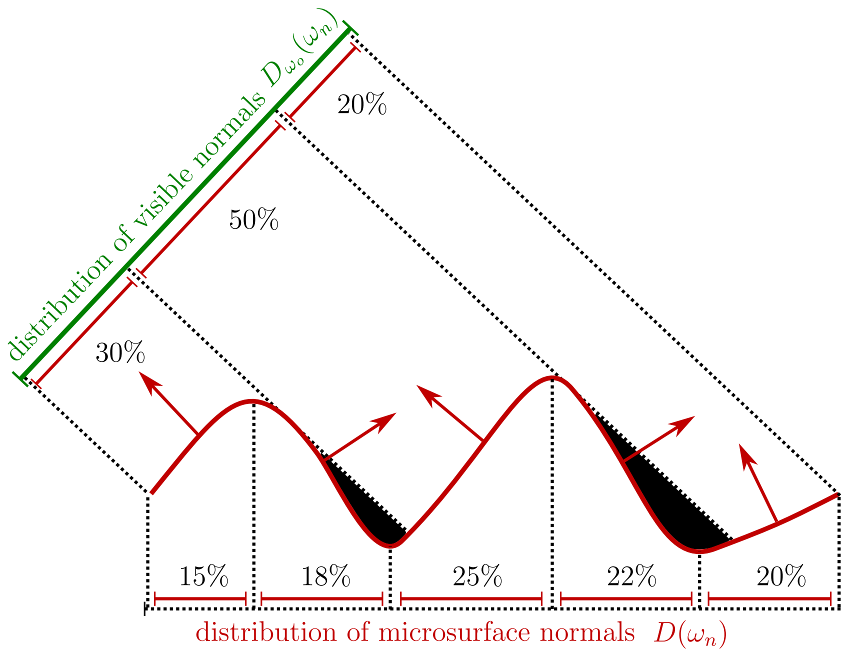


Figure 5: The distribution of microsurface normal $D(\omega_n)$ is an intrinsic surface property, while the distribution of visible normals $D_{\omega_o}(\omega_n)$ is view-dependent.

4.1 Distribution of Visible Normals

In this section, we will show that Equation (1) can be formulated in a microfacet paradigm as:

$$L(\omega_o) = \frac{1}{\cos\theta_o} \int_\Omega L(\omega_o, \omega_n)\, G_1(\omega_o, \omega_n)\, \langle \omega_o, \omega_n \rangle\, D(\omega_n)\, d\omega_n, \tag{12}$$

where $L(\omega_o, \omega_n)$ is the outgoing radiance from the microfacets with normal ω_n, and the factor $\frac{1}{\cos\theta_o}$ is here to normalize the integral by the projected area of the unit patch. We can see that the outgoing radiance from the surface is the sum of the outgoing radiance from each microfacet weighted by what we call *the distribution of visible normals* illustrated in Figure 5. It is the normal distribution weighted by the projected area (the clamped cosine) of each normal and by the masking function:

$$D_{\omega_o}(\omega_n) = \frac{G_1(\omega_o, \omega_n)\, \langle \omega_o, \omega_n \rangle\, D(\omega_n)}{\cos\theta_o}, \tag{13}$$

It is important that the distribution of visible normals $D_{\omega_o}(\omega_n)$ is normalized, because we use it as a weighting function to average radiances:

$$L(\omega_o) = \int_\Omega L(\omega_o, \omega_n)\, D_{\omega_o}(\omega_n)\, d\omega_n, \tag{14}$$

and, as explained in Section 2.1 and Figure 1, averaging radiances is only valid if the weighting function is normalized. This last equation is well defined because the integral in the denominator of Equation (1), which ensured correct normalization, is now represented in the masking function G_1. Indeed, we have seen in Equation (9) that Smith's masking function is defined as:

$$G_1(\omega_o, \omega_n) = \frac{\cos\theta_o\ \chi^+(\omega_o \cdot \omega_n)}{\int_\Omega \langle \omega_o, \omega_n' \rangle\, D(\omega_n')\, d\omega_n'},$$

so through substitution we can verify that the distribution of visible normals is normalized:

$$\begin{aligned}
\int_\Omega D_{\omega_o}(\omega_n)\, d\omega_n &= \int_\Omega \frac{G_1(\omega_o, \omega_n)\, \langle \omega_o, \omega_n \rangle\, D(\omega_n)}{\cos\theta_o}\, d\omega_n \\
&= \frac{\int_\Omega \langle \omega_o, \omega_n \rangle\, D(\omega_n)\, d\omega_n}{\int_\Omega \langle \omega_o, \omega_n \rangle\, D(\omega_n)\, d\omega_n} \\
&= 1,
\end{aligned}$$

and the average outgoing radiance from Equations (12) and (14) can thus be expressed in the same form as Equation (1), emphasizing the correct normalization:

$$L(\omega_o) = \frac{\int_\Omega L(\omega_o, \omega_n)\, \langle \omega_o, \omega_n \rangle\, D(\omega_n)\, d\omega_n}{\int_\Omega \langle \omega_o, \omega_n \rangle\, D(\omega_n)\, d\omega_n}.$$

4.2 Construction of the BRDF

We now construct the BRDF upon the distribution of visible normals. The radiance $L(\omega_o, \omega_n)$ of each microfacet can be expressed in terms of the micro-BRDF $\rho_\mu(\omega_o, \omega_i, \omega_n)$ associated with each microfacet, by integrating over the domain of the incident directions Ω_i (we keep Ω for the space of the normals):

$$L(\omega_o, \omega_n) = \int_{\Omega_i} dL(\omega_o, \omega_n) = \int_{\Omega_i} \rho_\mu(\omega_o, \omega_i, \omega_n)\, \langle \omega_i, \omega_n \rangle\, L(\omega_i)\, d\omega_i,$$

where the micro-BRDF is defined as the ratio of the differential outgoing radiance to the differential incoming irradiance:

$$dL(\omega_o, \omega_n) = \rho_\mu(\omega_o, \omega_i, \omega_n)\, \langle \omega_i, \omega_n \rangle\, L(\omega_i)\, d\omega_i.$$

We can use the definition of the micro-BRDF to differentiate Equation (14):

$$\begin{aligned}
dL(\omega_o) &= \int_\Omega dL(\omega_o, \omega_n)\, D_{\omega_o}(\omega_n)\, d\omega_n \\
&= L(\omega_i)\, d\omega_i \int_\Omega \rho_\mu(\omega_o, \omega_i, \omega_n)\, \langle \omega_i, \omega_n \rangle\, D_{\omega_o}(\omega_n)\, d\omega_n, \qquad (15)
\end{aligned}$$

where $L(\omega_i)\,d\omega_i$ can be moved outside the integral because it does not depend on ω_n. We will now instantiate this equation for the special case where the microfacets are perfect mirrors. The micro-BRDF for mirror-like microfacets is:

$$\begin{aligned}\rho_\mu(\omega_o,\omega_i,\omega_n) &= \left\|\frac{\partial\omega_h}{\partial\omega_i}\right\| \frac{F(\omega_o,\omega_h)\,\delta_{\omega_h}(\omega_n)}{|\omega_i\cdot\omega_h|} \\ &= \frac{F(\omega_o,\omega_h)\,\delta_{\omega_h}(\omega_n)}{4\,|\omega_i\cdot\omega_h|^2} \end{aligned} \tag{16}$$

where $\left\|\frac{\partial\omega_h}{\partial\omega_i}\right\| = \frac{1}{4\,|\omega_i\cdot\omega_h|}$ is the Jacobian of the reflection transformation [WMLT07], and F is the Fresnel term. In Equation (15), by substituting $\rho_\mu(\omega_o,\omega_i,\omega_n)$ from Equation (16) and $D_{\omega_o}(\omega_n)$ from Equation (13) we get:

$$dL(\omega_o) = L(\omega_i)\,d\omega_i \int_\Omega \frac{F(\omega_o,\omega_h)\,\delta_{\omega_h}(\omega_n)}{4\,|\omega_i\cdot\omega_h|^2}\,\langle\omega_i,\omega_n\rangle\,\frac{G_1(\omega_o,\omega_n)\,\langle\omega_o,\omega_n\rangle\,D(\omega_n)}{\cos\theta_o}\,d\omega_n$$

The delta function $\delta_{\omega_h}(\omega_n)$ allows us to replace the integral by the integrand evaluated at $\omega_n = \omega_h$, and the fact that $\omega_o\cdot\omega_h = \omega_i\cdot\omega_h$ reduces the expression to:

$$dL(\omega_o) = L(\omega_i)\,d\omega_i\,\frac{F(\omega_o,\omega_h)\,G_1(\omega_o,\omega_h)\,D(\omega_h)}{4\,\cos\theta_o}.$$

Since the macro-BRDF is defined by equation:

$$dL(\omega_o) = \rho(\omega_o,\omega_i)\,\cos\theta_i\,L(\omega_i)\,d\omega_i$$

we arrive at the following:

$$\begin{aligned}\rho(\omega_o,\omega_i) &= \frac{F(\omega_o,\omega_h)\,G_1(\omega_o,\omega_h)\,D(\omega_h)}{4\,\cos\theta_o\,\cos\theta_i} \\ &= \frac{F(\omega_o,\omega_h)\,G_1(\omega_o,\omega_h)\,D(\omega_h)}{4\,|\omega_g\cdot\omega_o|\,|\omega_g\cdot\omega_i|}\end{aligned} \tag{17}$$

An important observation is that this equation models only how rays are reflected just after the first bounce *before* leaving the surface (Table 3(b)). However, a BRDF model must describe instead how rays are distributed *after* leaving the surface. The distribution before and after leaving the surface is not the same because some reflected rays hit the microsurface again and are reflected in another direction before leaving (Table 3(d)). Since the BRDF model derived here only accounts for the first bounce on the surface, rays involving multiple bounces (shown in black in Table 3(c)) have to be removed from the model, which is achieved by introducing a *shadowing* function. We replace the masking function G_1 by a masking-shadowing function G_2, and in doing so arrive at the well-known equation of specular microfacet-based BRDFs [WMLT07]:

$$\boxed{\rho(\omega_o,\omega_i) = \frac{F(\omega_o,\omega_h)\,G_2(\omega_o,\omega_i,\omega_h)\,D(\omega_h)}{4\,|\omega_g\cdot\omega_o|\,|\omega_g\cdot\omega_i|}}. \tag{18}$$

In Section 5, we discuss the choice of the masking-shadowing function.

4.3 The BRDF Normalization Test

The White Furnace Test The bidirectional scattering distribution function (BSDF) s is the sum of the bidirectional reflectance distribution function (BRDF) ρ defined on the upper hemisphere and the bidirectional transmittance distribution function (BTDF) t defined on the lower hemisphere:

$$s(\omega_o, \omega_i) = \rho(\omega_o, \omega_i) + t(\omega_o, \omega_i).$$

If we had a perfect surface that never dissipates energy into heat, then the energy of the rays would be perfectly preserved. Thus, an important property that should be verified by microfacet-based scattering models is that, when the surface absorption is 0, the distribution of scattered rays is perfectly normalized:

$$\int_\Omega s(\omega_o, \omega_i)\, |\omega_g \cdot \omega_i|\, d\omega_i = 1.$$

If the Fresnel term is always 1, then rays are never transmitted (they never penetrate the surface), the BTDF evaluates to $t = 0$, and the scattering model is then entirely defined by the BRDF (i.e., $s = \rho$). In this case, the rays are all reflected without energy loss and their distribution is normalized. This is modeled by the *White Furnace Test* equation:

$$\int_\Omega \rho(\omega_o, \omega_i)\, |\omega_g \cdot \omega_i|\, d\omega_i = 1.$$

Intuitively, it represents the fact that rays cast from the view direction (Table 3(a)) would be scattered one or more times and eventually leave the surface (Table 3(d)). However, common analytical BRDFs do not model multiple scattering on the microsurface; the rays that bounce multiple times are removed from the BRDF by the shadowing function, as shown in Table 3(c) and described in Section 4.2. This is why common BRDF models do not integrate to 1 and do not satisfy the White Furnace Test equation.

The Weak White Furnace Test The White Furnace Test cannot be used to validate common BRDF models, which incorporate only the first scattering event. However, we can design another less restrictive test that must be satisfied by common microfacet-based BRDFs. We can verify that the distribution of rays reflected just after the first bounce and *before* leaving the surface is normalized (Table 3(b)). This can be achieved by replacing masking-shadowing by masking alone ($G_2(\omega_o, \omega_i, \omega_h) = G_1(\omega_o, \omega_h)$). Without Fresnel and shadowing, the BRDF from Equation (18) becomes:

$$\rho(\omega_o, \omega_i) = \frac{G_1(\omega_o, \omega_h)\, D(\omega_h)}{4\, |\omega_g \cdot \omega_o|\, |\omega_g \cdot \omega_i|},$$

and after cancellation of $|\omega_g \cdot \omega_i|$, the *Weak White Furnace Test* equation is given by:

$$\boxed{\int_\Omega \frac{G_1(\omega_o, \omega_h)\, D(\omega_h)}{4\, |\omega_g \cdot \omega_o|}\, d\omega_i = 1}. \tag{19}$$

This condition is only met with an appropriate masking function G_1. Under the assumption of the model (normal/masking independence) only Smith's masking function satisfies this condition. This is because Smith's masking function is not an approximation but the exact normalization coefficient required for the distribution of visible normals used in the BRDF, as shown in Section 4.1.

In Appendix C, we provide MATLAB code to numerically compute Equation (19) with Beckmann and GGX distributions and their associated Smith masking functions.

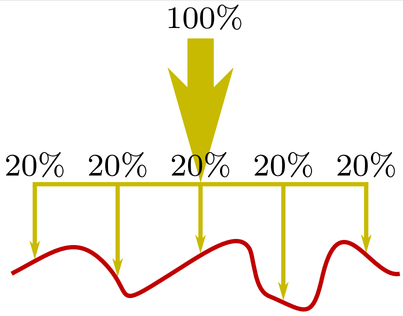
4.4 Summary

A frequently asked question concerning BRDF normalization is: "*Microfacet-based BRDFs do not integrate to 1. Shouldn't they be perfectly normalized?*"

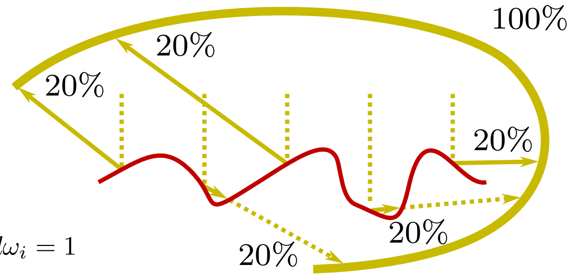
In this section we answered this question by developing the following ideas:

- The BRDF is constructed from the distribution of visible normals.
- The distribution of visible normals has to be normalized to ensure that the BRDF conserves energy.
- The normalization coefficient of the distribution of visible normals is the masking function.
- Microfacet-based BRDFs should theoretically be normalized and integrate exactly to 1.
- The shadowing function in microfacet-based BRDFs is used to separate the first scattering events from the multiple scattering events on the microsurface. Shadowing sets to 0 the scattering events of order greater than 1 and leaves the BRDF artificially unnormalized in the absence of a term to model multiple scattering events.
- The standard form of microfacet-based BRDFs is normalized by the masking function and without Fresnel and shadowing. This is what we call the "Weak White Furnace Test".
- If the model assumes normal/masking independence, only Smith's masking function satisfies "Weak White Furnace Test".

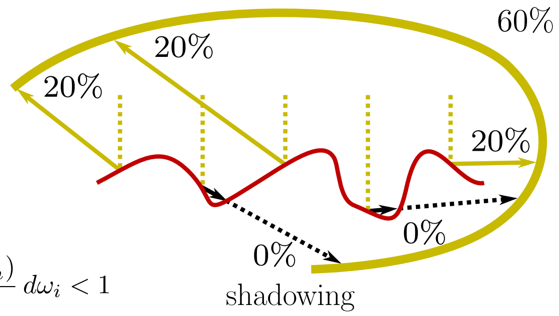
Note that the Weak White Furnace Test, in which shadowing is not incorporated, is a simple way to verify that the masking function is well defined. It is important to note that this does not mean that common BRDF models should be used without shadowing. Shadowing is what separates energy reflected after the first bounce from energy reflected after multiple bounces, which is not incorporated into common BRDF models.

<table>
<tr>
<td>(a) Casting rays onto the microsurface</td>
<td>
</td>
</tr>
<tr>
<td>(b) Without Fresnel and shadowing, the rays are always reflected somewhere and their energy does not change.

⇒ The distribution of reflected rays is normalized.

$$\int_\Omega \rho(\omega_o, \omega_i)\, |\omega_g \cdot \omega_i|\, d\omega_i = \int_\Omega \frac{G_1(\omega_h, \omega_o)\, D(\omega_h)}{4\, |\omega_g \cdot \omega_o|}\, d\omega_i = 1$$</td>
<td>
</td>
</tr>
<tr>
<td>(c) Shadowing in the BRDF models the fact that rays occluded by the microsurface are discarded.

⇒ The distribution of reflected rays is NOT normalized.

$$\int_\Omega \rho(\omega_o, \omega_i)\, |\omega_g \cdot \omega_i|\, d\omega_i = \int_\Omega \frac{G_2(\omega_o, \omega_i, \omega_h)\, D(\omega_h)}{4\, |\omega_g \cdot \omega_o|}\, d\omega_i < 1$$</td>
<td>
</td>
</tr>
<tr>
<td>(d) In the real world, rays are not discarded but scatter several times before leaving the surface.

⇒ A complete BRDF model should incorporate multiple scattering and would be normalized.

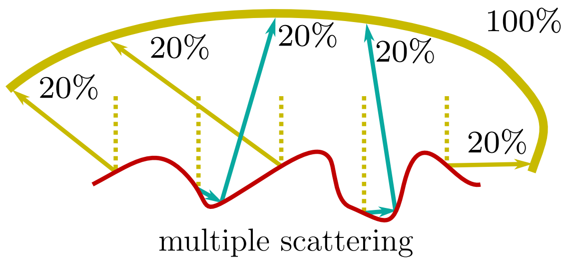
$$\int_\Omega \rho(\omega_o, \omega_i)\, |\omega_g \cdot \omega_i|\, d\omega_i = 1$$</td>
<td>
</td>
</tr>
</table>

Table 3: Normalization of microfacet-based BRDFs.

5 Smith's Joint Masking-Shadowing Function

In this section, we review the use of the Smith masking function with the light direction (i.e., as a shadowing function) and its joint form with the masking function. We recall four different forms of the joint masking-shadowing function. Each of these use the function Λ defined in Section 3—evaluated for the view ($\Lambda(\omega_o)$) and light ($\Lambda(\omega_i)$) directions—and combine them in different ways, producing different properties as a result.

Separable Masking and Shadowing The most simple and widely used variant of the masking-shadowing function is the separable form popularized by Walter et al. [WMLT07]. In this instance, masking and shadowing are supposed to be independent, and are computed separately and multiplied together:

$$\begin{aligned} G_2(\omega_o, \omega_i, \omega_n) &= G_1(\omega_o, \omega_n)\, G_1(\omega_i, \omega_n) \\ &= \frac{\chi^+(\omega_o \cdot \omega_n)}{1 + \Lambda(\omega_o)} \frac{\chi^+(\omega_i \cdot \omega_n)}{1 + \Lambda(\omega_i)}. \end{aligned} \tag{20}$$

This form does not model correlations between masking and shadowing, and therefore always overestimates shadowing since some correlation always exists, as explained in the next section.

Height-Correlated Masking and Shadowing A more accurate form of the masking-shadowing function models the correlation between masking and shadowing due to the height of the microsurface [RDP05]. Intuitively, the more a microfacet is elevated within the microsurface, the more the probabilities of being visible for the view direction (unmasked) and for the light direction (unshadowed) increase at the same time. Thus, masking and shadowing are correlated because of the elevation of the microfacets. This correlation is accounted for in the following form of the joint masking-shadowing function:

$$G_2(\omega_o, \omega_i, \omega_n) = \frac{\chi^+(\omega_o \cdot \omega_n)\ \chi^+(\omega_i \cdot \omega_n)}{1 + \Lambda(\omega_o) + \Lambda(\omega_i)}. \tag{21}$$

This form is correct when the view and light directions are far away, but overestimates shadowing when the directions are close. We suggest to use Equation (21) in practice, because it is more accurate than the separable from of Equation (20) whilst having an equivalent computational complexity. We recall the derivation of this form in Appendix B.

Direction-Correlated Masking and Shadowing Masking and shadowing are also strongly correlated when the view and light directions are close. Typically, when $\omega_o = \omega_i$, masking and shadowing are totally correlated because microfacets visible to the viewer are also visible to the light. In this case, the shadowing should be removed from the BRDF because shadowed microfacets are not visible to the light, and thus they are also not visible to the viewer. This is known as the "hotspot effect": when the view and light directions are parallel, shadows disappear. This does not mean that shadows no longer exist, only that they are not visible from this specific view direction. Since the BRDF models the radiance measured in the view direction, if shadowing exists on the surface but is not visible then it should not be part of the BRDF.

On a surface, full correlation is reached when ω_o and ω_i have the same azimuthal angle. In this case, the masking-shadowing function can be replaced by the minimum of masking and shadowing. Ashikhmin et al. [APS00] account for directional correlation by blending the separable form of Equation (20) with the case where both directions are fully correlated:

$$\begin{aligned} &G_2(\omega_o, \omega_i, \omega_n) \\ &= \lambda(\phi)\, G_1(\omega_o, \omega_n)\, G_1(\omega_i, \omega_n) + (1 - \lambda(\phi))\, \min(G_1(\omega_o, \omega_n), G_1(\omega_i, \omega_n)). \end{aligned} \tag{22}$$

where $\lambda(\phi)$ is an empirical factor similar to Ginneken et al.'s, which is presented next. Because the authors do not have Smith's analytical expression for function Λ, they have to compute masking and shadowing separately. This is why they have to blend the separable and the fully uncorrelated forms and cannot incorporate height correlation into their model.

Height-Direction-Correlated Masking and Shadowing The directional correlation between masking and shadowing can be modeled by adding a factor to the shadowing term in the height-onlycorrelated form:

$$G_2(\omega_o, \omega_i, \omega_n) = \frac{\chi^+(\omega_o \cdot \omega_n)\ \chi^+(\omega_i \cdot \omega_n)}{1 + \max\left(\Lambda(\omega_o), \Lambda(\omega_i)\right) + \lambda(\omega_o, \omega_i)\ \min\left(\Lambda(\omega_o), \Lambda(\omega_i)\right)}. \tag{23}$$

Here, masking and shadowing are fully correlated when the view and light directions are parallel and $\lambda = 0$. The correlation decreases as the angle between the light and view directions increases, and as λ increases up to 1. In this case, masking and shadowing are no longer directionally correlated and the formula returns to the elevation-correlated form.

Ginneken et al. [vGSK98] proposed an empirical factor $\lambda = \frac{4.41\,\phi}{4.41\phi+1}$—which depends on ϕ, the azimuthal angle difference between ω_o and ω_i—and is independent of the surface roughness. Heitz et al. recently presented a more in-depth study of this problem and an analytic approximation for $\lambda(\omega_o, \omega_i)$, which incorporates surface roughness when D is a Beckmann distribution [HBP13]. The result was given for isotropic Beckmann distributions only, but the stretching invariance presented in Section 3 can be used to easily generalize this result to anisotropic Beckmann. This form models exactly the correlation of masking and shadowing and is thus more accurate than the forms presented in Equations (20), (21), and (22). The derivation of practical forms for λ and generalization non-Gaussian distributions are open problems.

6 The V-Cavity Model

In this section, we discuss the masking model based on V-cavities [CT82, ON94], which is the most common alternative to the Smith masking function. We show that this model satisfies the Weak White Furnace Test and is thus mathematically well designed. However, we also show that the surface profile from this model is unrealistic and is similar to a normal map, where microfacets have no geometrical existence. We measure how this impacts the shape of the BRDF.

6.1 The V-Cavity Masking Function

Figure 6 illustrates the scattering model with V-cavity surfaces. Rather than modeling the scattering on one surface with a distribution of normals, this model computes the scattering on separate surfaces and averages their contributions. Each surface is composed of two normals $\omega_n = (x_n, y_n, z_n)$ and $\omega_n{}' = (-x_n, -y_n, z_n)$ and the contribution of each surface is weighted by $D(\omega_n)$ in the final BRDF.

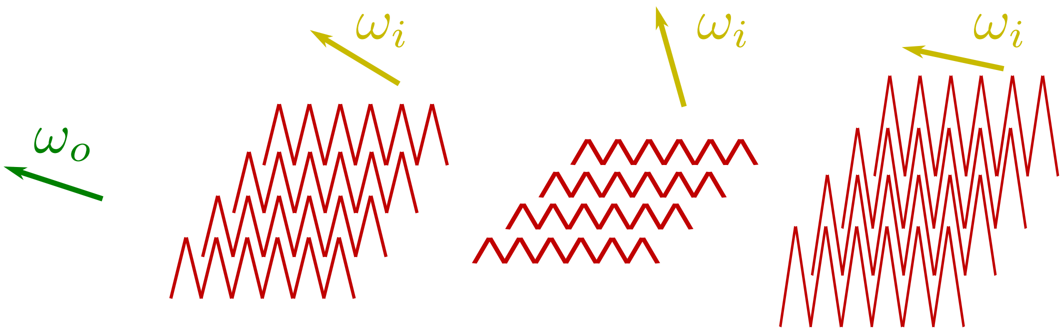


Figure 6: The V-cavity scattering model. Instead of modeling the scattering on one surface, the model computes the scattering on separate surfaces and blends the results.

A geometric demonstration is usually presented to derive the masking function of a V-cavity surface. We can derive the same result without going through trigonometric calculus by applying the conservation of the visible projected area like in Section 2.2. V-cavity surfaces have only two symmetric normals ω_n and $\omega_n{}'$. The normal distribution of this surface is:

$$D(\omega) = \frac{1}{2}\frac{\delta_{\omega_n}(\omega)}{\omega_n \cdot \omega_g} + \frac{1}{2}\frac{\delta_{\omega_n{}'}(\omega)}{\omega_n{}' \cdot \omega_g}.$$

We verify that the normalization is correct:

$$\begin{aligned}
\int_\Omega \langle \omega, \omega_g \rangle D(\omega) d\omega &= \int_\Omega \langle \omega, \omega_g \rangle \left(\frac{1}{2}\frac{\delta_{\omega_n}(\omega)}{\omega_n \cdot \omega_g} + \frac{1}{2}\frac{\delta_{\omega_n{}'}(\omega)}{\omega_n{}' \cdot \omega_g} \right) d\omega \\
&= \frac{1}{2}\frac{\omega_n{}' \cdot \omega_g}{\omega_n{}' \cdot \omega_g} + \frac{1}{2}\frac{\omega_n{}' \cdot \omega_g}{\omega_n{}' \cdot \omega_g} \\
&= \frac{1}{2} + \frac{1}{2} = 1.
\end{aligned}$$

To derive the masking term, we use the conservation of the visible projected area presented in Equation (5):

$$\begin{aligned}\cos\theta_o &= \int_\Omega G_1(\omega_o,\omega)\langle\omega_o,\omega\rangle D(\omega)d\omega \\ &= \frac{1}{2}G_1(\omega_o,\omega_n)\frac{\langle\omega_o,\omega_n\rangle}{\omega_n\cdot\omega_g} + \frac{1}{2}G_1(\omega_o,{\omega_n}')\frac{\langle\omega_o,{\omega_n}'\rangle}{{\omega_n}'\cdot\omega_g}\end{aligned}$$

There are two possible configurations as shown in Figure 7. In the first case the two normals are visible and there is no masking ($G_1(\omega_o,\omega_n) = 1$ and $G_1(\omega_o,{\omega_n}') = 1$). Otherwise, ${\omega_n}'$ is backface-culled ($G_1(\omega_o,{\omega_n}') = 0$) and we have:

$$\cos\theta_o = \frac{1}{2}G_1(\omega_o,\omega_n)\frac{\langle\omega_o,\omega_n\rangle}{\omega_n\cdot\omega_g},$$

whose solution is:

$$\begin{aligned}G_1(\omega_o,\omega_n) &= 2\,\frac{\cos\theta_o\,(\omega_n\cdot\omega_g)}{\langle\omega_o,\omega_n\rangle} \\ &= 2\,\frac{(\omega_n\cdot\omega_g)(\omega_o\cdot\omega_g)}{\langle\omega_o,\omega_n\rangle}.\end{aligned}$$

The result of these two configurations can be expressed in a single formula:

$$G_1(\omega_o,\omega_n) = \min\left(1, 2\,\frac{(\omega_n\cdot\omega_g)(\omega_o\cdot\omega_g)}{\langle\omega_o,\omega_n\rangle}\right), \tag{24}$$

which is the well-known V-cavity masking function used by Cook and Torrance [CT82].

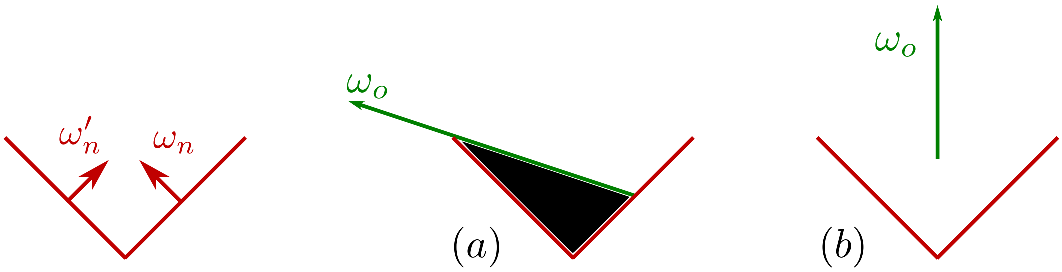


Figure 7: masking on a V-cavity surface. Either one of the two normal is backface-culled and the other is partially masked (a) or the two normals are visible and the masking function evaluates to 1.

6.2 The V-Cavity BRDF Model

The Distribution of Visible Normals We investigate what is the consequence of using the V-cavity masking function in the BRDF. As we have seen in Section 4.2, the BRDF is constructed from the distribution of visible normals defined by Equation (13). We substitute G_1 given in Equation 24:

$$D_{\omega_o}(\omega_n) = \min\left(1, 2\,\frac{(\omega_n\cdot\omega_g)(\omega_o\cdot\omega_g)}{\langle\omega_o,\omega_n\rangle}\right)\frac{\langle\omega_o,\omega_n\rangle\,D(\omega_n)}{\cos\theta_o}.$$

The form is complicated to study because of the $\min(1,-)$ term. However, we will see that the main difference introduced by the V-cavity model happens near the silhouettes. Thus, to simplify the study and go straight to the point we will focus the derivation only on what happens at grazing angles, where $\theta_o \approx \frac{\pi}{2}$. Indeed, at grazing angles, we are always in configuration (a) from Figure 7, where one of the two normals is backface-culled. In this case, we can drop the $\min(1,-)$:

$$\begin{aligned} D_{\omega_o}(\omega_n) &= 2\,\frac{(\omega_n\cdot\omega_g)(\omega_o\cdot\omega_g)}{\langle\omega_o,\omega_n\rangle}\,\frac{\langle\omega_o,\omega_n\rangle\,D(\omega_n)}{\cos\theta_o} \\ &= 2\;\chi^+(\omega_o\cdot\omega_n)\,(\omega_n\cdot\omega_g)\,D(\omega_n)\,d\omega_n. \end{aligned} \tag{25}$$

Note that the clamped dot product $\langle\omega_o,\omega_n\rangle$ simplifies out but the heaviside $\chi^+(\omega_o\cdot\omega_n)$ is left to make sure that backface-culled normals are still removed from the distribution. We validate this result by verifying that the distribution of visible normals is normalized. We compute:

$$\int_\Omega D_{\omega_o}(\omega_n)\,d\omega_n = 2\int_\Omega \chi^+(\omega_o\cdot\omega_n)\,(\omega_n\cdot\omega_g)\,D(\omega_n)d\omega_n.$$

Since the view is almost orthogonal, the heaviside function truncates the integral almost at the middle of the distribution. Also, a V-cavity surface implies that the distribution of normals is symmetrical, i.e. $D(\omega_n) = D(\omega_n{}')$. This implies that the heaviside function cuts the normal distribution into two equal parts and we get:

$$\begin{aligned} \int_\Omega \chi^+(\omega_o\cdot\omega_n)\,(\omega_n\cdot\omega_g)\,D(\omega_n)\,d\omega_n &= \frac{1}{2}\int_\Omega(\omega_n\cdot\omega_g)\,D(\omega_n)\;d\omega_n \\ &= \frac{1}{2}, \end{aligned}$$

where we use the fact that the normal distribution is normalized $\int_\Omega(\omega_n\cdot\omega_g)\,D(\omega_n)\,d\omega_n = 1$. We get:

$$\begin{aligned} \int_\Omega D_{\omega_o}(\omega_n)\,d\omega_n &= 2\int_\Omega \chi^+(\omega_o\cdot\omega_n)\,(\omega_n\cdot\omega_g)\,D(\omega_n)\,d\omega_n \\ &= 2\,\frac{1}{2} = 1. \end{aligned}$$

This shows that the distribution of visible normals from Equation 25 is normalized for grazing view angles. A more technical derivation can show that the distribution is normalized for any view angle. Another way to validate this model is to use the Weak White Furnace Test. We evaluate Equation 19 by using G_1 from Equation 24. The result of the integral is always 1 and the model with V-cavities is thus mathematically well designed and energy conserving.

The Microsurface Profile Nevertheless, while the distribution of visible normals of V-cavities is mathematically well defined, it is not physically plausible and models a non-realistic surface profile at grazing view angles.

There are two kind of normals: the ones that are backface-culled are removed by the heaviside term and the ones that are not backface-culled are weighted by $(\omega_n\cdot\omega_g)\,D(\omega_n)$. Note that the factor $(\omega_n\cdot\omega_g)$ is the Jacobian of the projection of a microfacet onto the macrosurface, as shown in Table 1(a). Thus, the microfacets are weighted exactly as if they were projected onto the macrosurface before being projected onto the view direction. As a result, we are simulating a geometrically flat microsurface: the microfacets can perturbate the reflection of light but they don't exist geometrically. This microsurface model is not realistic because it behaves more like a normal map than a displacement map, as shown in Figure 8.

This effect was expected: rather than simulating one microsurface, the V-cavity model simulates one surface per pair of normals and averages the results of the simulation. On one single surface, highly visible normals would occupy more projected area than less visible normals and thus have more important weights. However, this does not happen with V-cavities because different normals are simulated separately and are weighted by the normal distribution. There is no view dependence in the weighting (except for the backface culling). This is why the V-cavity model poorly incorporates the effect of visibility and ends up simulating something close to a normal map.

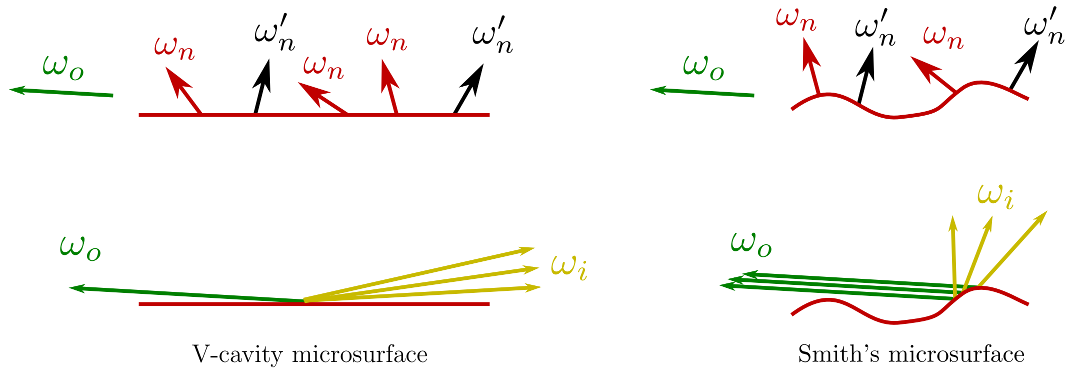


Figure 8: (top) V-cavity surfaces exhibit a normal map behavior at grazing angles: non-backface-culled normals have the same visibility than the macrosurface as if they had no geometrical existence. (bottom) For grazing view directions, the lobe reflected from the V-cavity surface is too low compared to the lobe reflected by physical surfaces.

The Shape of The BRDF The more the view angle is grazing, the more the surface profile tends to exhibit this normal map behavior. The consequence for the BRDF is that the reflected lobes tend to be too low: on a real surface, normals oriented toward the view direction are more weighted in the BRDF because their projected area is more important. Because of this, the reflected lobe tends to be shifted toward the view direction, as shown in Figure 8. This shifting effect is not present with normal maps because the microfacets have no geometrical existence: they all have the same projected area. Table 4 shows the reflected lobes of an isotropic Beckmann BRDF with the V-cavity and Smith masking-shadowing functions, and with measured data. We see that, with the Smith masking function, the reflected lobe is shifted toward the view direction as the roughness increases. For very important roughnesses, the lobe is even mainly backscattering. This effect, present in the measured data, was expected, because the normals oriented toward the view direction are the most visible. However, this effect does not emerge with V-cavities.

6.3 Summary

Another frequently asked question concerning the masking function is: "*Is the V-cavity masking function wrong?*"
The typical answer to this question is: "*It should be wrong because it does not depend on the normal distribution."*

In this section we answered this question by developing the following ideas:

- The V-cavity masking function per-normal does not depend on the roughness.
- However, the average of the V-cavity masking function does depend on the roughness. This is because the masking function and the normals are not assumed to be independent, contrary to Smith's model. The more the surface roughness increases, the more the average masking of the BRDF increases.
- The V-cavity model is mathematically well defined, satisfies the Weak White Furnace Test, and is thus energy conserving.
- The V-cavity masking function can be used with any kind of symmetric normal distribution and guarantees correct normalization.
- However, the surface profile assumed by the V-cavity model has a response close to a normal map with flat microfacets at grazing view angles. It is not physically realistic.
- The consequence is that at grazing angles and with important roughness, the BRDF reflected lobe is too low compared to what is expected from a realistic material.

There is no definitive answer to the question of choosing V-cavities or Smith. Both are mathematically well defined. V-cavities are cheaper and generic, they mathematically work with any kind of normal distribution, but are less realistic. In contrast, the Smith-based model is physically accurate but requires specific derivations and sometimes expensive evaluations. The choice is thus a matter of tradeoff between realism and complexity.

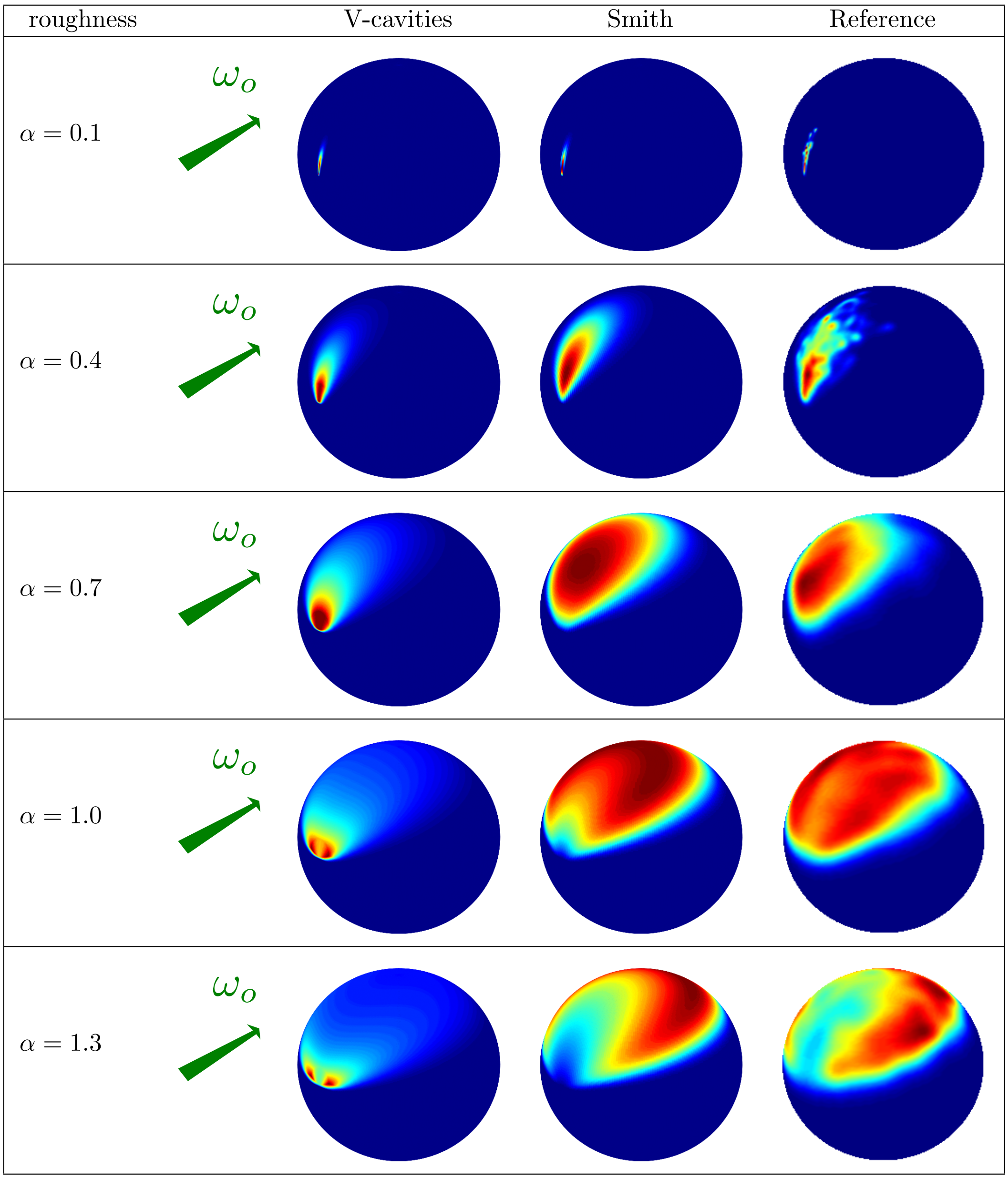


Table 4: (left and middle) The reflected lobe of the BRDF with an isotropic Beckmann distribution and different masking functions at grazing view-angle $\theta_o = 1.5$. (right) Reference computed with Monte Carlo raytracing on a procedural surface.

7 Discussion and Future Work

In this section, we discuss several ideas resulting from the derivations presented in this paper, and provide some thoughts for possible future work.

Normal/Masking Independence We have seen in Section 2.3 that normal/masking independence is a core assumption of microfacet theory in computer graphics, and is equivalent to assuming that the heights and the slopes of the microsurface are uncorrelated. However, this is not physically realistic because it is means that the autocorrelation function is 0 beyond the local microfacet, which implies a random 3D set of microfacets rather than a continuous surface—as shown in Figure 9. This is reminiscent of "metal flakes", which can be found in some metallic car paints [RMS+08], but real-world continuous surfaces have wider autocorrelation functions. Bourlier et al. compared Smith's masking function to the numerically measured masking function on random rough surfaces with different autocorrelation functions (Gaussian and Lorentzian) [BSB00]. The conclusion of their investigation was that the error introduced by neglecting correlation on random surfaces is, on average, small and noticeable only at observation angles such as $\tan(\theta)/\alpha > 0.5$, where $\sigma^2 = \frac{\alpha^2}{2}$ is the slope variance. Smith's masking function tends to produce slight overestimations in this case. Given that Smith's masking function is overall accurate even on correlated surfaces and given that there is no analytical solution to the correlated masking function, it seems reasonable to stick to Smith's masking function in a computer graphics context. However, as pointed out by Ashikhmin et al., the effect of correlation on non-random surfaces with repetitive or structured patterns (e.g. fabric) can be of high importance and must be incorporated into dedicated models [APS00].

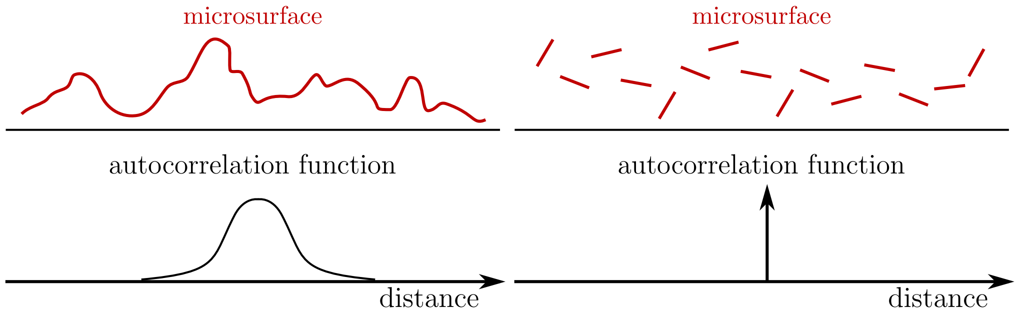


Figure 9: Microsurfaces and their autocorrelation functions. (left) A real-world continuous microsurface with large autocorrelation distance. (right) A uncorrelated surface where each microfacet is not correlated to its neighborhood, as modeled in microfacet theory.

Deriving the Smith Masking Function for Other Commonly Used Models We have seen that the Smith masking function must be used in microfacet-based BRDF models. Closed analytical forms can be derived for the Beckmann and GGX distributions, as reviewed in Section 2. However, the masking function does not always integrate analytically for other commonly used normal distributions.

An important example is the Phong distribution. Walter et al. proposed to use Smith's masking function for the Beckmann distribution since they have a similar appearance for small roughnesses. However, the more that the roughness increases, the more the error becomes important [WMLT07]. It would be interesting to derive an analytical approximation for function Λ dedicated to the Phong distribution. Walter et al. proposed such an approximation for Beckmann because it is cheaper than the analytical solution. It is easy to do so for Beckmann because the information contained in the distribution is only 1D, because Beckmann is shape invariant, as discussed in Section 3. Indeed, the Beckmann distribution depends only on the ratio $a = \tan(\theta)/\alpha$. All of the information required for the distribution, and thus for the masking function, depends only on variable a. This is why the function Λ used in masking can be encoded as a 1D function of variable a, which is efficiently represented as a rational polynomial for the Beckmann distribution. Doing the same for the Phong distribution is less straightforward because it cannot be represented as a 1D function of a, because it is not shape invariant. However, it is certainly possible to merge θ and α into another intermediate quantity that the Phong Λ function would be a 1D function of, or find an accurate 2D fit instead.

Another example is the generalization of GGX distribution called GTR [Bur12], whose masking function has yet to be found.

Correlation of Masking-Shadowing As we have seen in Section 5, multiplying masking and shadowing together is a very rough approximation because these effects can be correlated. Deriving accurate and practical forms of the correlated masking-shadowing function for arbitrary normal distributions is an open problem.

Multiple Scattering Modeling multiple scattering is one possible way to introduce effects that are poorly represented by our common BRDF model. For instance, Beckmann, Phong and even GGX are known to have overly short "tails" compared to measured materials [Bur12]. The first reflex in the CG community is to keep the standard BRDF formulation and tweak the normal distribution. For instance, Bagher et al. [BSH12] use a shifted Gamma distribution to fit measured materials. This distribution is complicated to compute and to integrate, and furthermore, they have to tweak the Fresnel term to make their model fit the data. In the end, their model performs well as a fitting tool, but it no longer makes physical sense. In the same way, Burley [Bur12] generalizes GGX to GTR to create a BRDF with a longer tail, in order to more accurately represent measured materials. But the masking function is not available for GTR, so instead he uses a tweaked masking function, violating the fundamental link with the normal distribution. It seems that we have almost reached the limit of what is feasible with this model. Yet, in the race for physical accuracy we keep pushing it further, sometimes even at the cost of violating the model's physical basis, which is counterproductive.

Rather than continuing to invent more complicated ways to parameterize the model, we should ask ourselves whether certain effects present in measured data are simply missing from the model, and therefore look to extend it instead. Modeling multiple scattering seems like a good candidate here, and in fact it has already been investigated in the physics literature [BB04]. However, these models are quite complicated, because the physics community aims for accuracy rather than for ease of implementation. A first attempt to model it in a simple and practical way for computer graphics applications would be to combine the knowledge of energy conservation and empirical observations. In Section 4.2, we showed that shadowing is introduced in common microfacet BRDFs because they only model the first scattering event. An interesting future research avenue would be to introduce a BRDF model with multiple scattering:

$$\rho(\omega_o, \omega_i) = \rho_1(\omega_o, \omega_i) + \rho_{2^+}(\omega_o, \omega_i)$$

where $\rho_1(\omega_o, \omega_i)$ would be the usual BRDF term modeling the first scattering that incorporates shadowing, and $\rho_{2^+}(\omega_o, \omega_i)$ would be a new multiple scattering term. We know that a multiple scattering BRDF model passes the White Furnace Test (when Fresnel is set to 1):

$$\int_\Omega \rho(\omega_o, \omega_i)\, |\omega_g \cdot \omega_i|\, d\omega_i = \int_\Omega \rho_1(\omega_o, \omega_i)\, |\omega_g \cdot \omega_i|\, d\omega_i + \int_\Omega \rho_{2^+}(\omega_o, \omega_i)\, |\omega_g \cdot \omega_i|\, d\omega_i = 1,$$

and that the energy present in the multiple scattering term would be completely determined by the energy loss due to shadowing in the first scattering term:

$$E_{2^+} = \int_\Omega \rho_{2^+}(\omega_o, \omega_i)\, |\omega_g \cdot \omega_i|\, d\omega_i = 1 - \int_\Omega \rho_1(\omega_o, \omega_i)\, |\omega_g \cdot \omega_i|\, d\omega_i$$

The shape of ρ_{2^+} could be investigated, for instance, by computing Monte-Carlo simulations on rough surface samples. If its shape turns out to be simple, then as a first approximation we could model ρ_{2^+} with an analytical function (e.g. like a single lobe) of norm E_{2^+}. When Fresnel is not 1 (when the surface transmits) then E_{2^+} should depend on F as well. As a first approximation, it could be multiplied by the average visible value of Fresnel, which could be precomputed via:

$$\bar{F}_{\omega_o} = \int_\Omega F(\omega_o, \omega_n)\, D_{\omega_o}(\omega_n)\, d\omega_n,$$

and stored in a look-up table. Note that multiplying by $\bar{F}_{\omega_o}$ would rescale the energy present in E_{2^+} according to the ratio of rays transmitted after the first bounce only. Perhaps the average Fresnel value after multiple bounces could be precomputed as well. In general, since multiple scattering tends to smooth out functions, one can reasonably expect it to be efficiently represented and stored with simple analytical functions or small precomputed look-up textures.

8 Conclusion

In this document, we recalled how the masking function is linked to the normal distribution function by the visible projected area. By using this knowledge and the usual normal/masking independence assumption, we have shown that the Smith masking function is the only valid one. We have shown that the masking function is stretching invariant and how this property can be used to generalize known results for anisotropic normal distributions. Upon that, we defined the visible normal distribution, which we used to derive the common form of the BRDF, emphasizing the link with normalization and energy conservation. During this derivation, we introduced shadowing and we reviewed different shadowing models. We have shown that shadowing has to be part of the common form of the BRDF model, which only incorporates the first scattering event that occurs on the microsurface. We introduced the Weak White Furnace Test, which can be used to verify that BRDFs of this kind are well defined. We reviewed the V-cavity model and we showed why it is mathematically well defined but not realistic.

In the last section, we discussed the limitations of the BRDF model from the 70's that the graphics community is still using today. Finally, we suggested that by extending the model, it should be possible to represent more effects present in measured materials in a simple and practical way instead of continuing to explore its parametrization by introducing new normal distribution functions with growing complexity and less practicability.

Acknowledgements

I would like to thank Sébastien Lagarde who encouraged me to write this report and gave the first feedback, Stephen Hill and Naty Hoffman for their numerous advices and corrections, and Jonathan Dupuy for our endless discussions about microfacet theory.

References

[APS00] Michael Ashikmin, Simon Premože, and Peter Shirley. A microfacet-based brdf generator. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '00, pages 65–74, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co.

[BB04] C Bourlier and G Berginc. Multiple scattering in the high-frequency limit with second-order shadowing function from 2d anisotropic rough dielectric surfaces. *Waves in Random Media*, 14(3):253–276, 2004.

[Bro80] G. Brown. Shadowing by non-gaussian random surfaces. *IEEE Trans. on Antennas and Propagation*, 28(6):788 – 790, nov 1980.

[BS63] P. Beckmann and A. Spizzichino. *The scattering of electromagnetic waves from rough surfaces*. International series of monographs on electromagnetic waves. Pergamon Press, 1963.

[BSB00] C. Bourlier, Joseph Saillard, and G. Berginc. Effect of correlation between shadowing and shadowed points on the wagner and smith monostatic one-dimensional shadowing functions. *Antennas and Propagation, IEEE Transactions on*, 48(3):437–446, 2000.

[BSH12] M. M. Bagher, C. Soler, and N. Holzschuch. Accurate fitting of measured reflectances using a shifted gamma micro-facet distribution. *Comp. Graph. Forum*, 31(4):1509–1518, June 2012.

[Bur12] Brent Burley. Physically-based shading at disney. In *ACM SIGGRAPH 2012 Courses*, SIGGRAPH '12, pages 10:1–7. ACM, 2012.

[CT82] R. L. Cook and K. E. Torrance. A reflectance model for computer graphics. *ACM Trans. Graph.*, 1(1):7–24, January 1982.

[DHI+13] Jonathan Dupuy, Eric Heitz, Jean-Claude Iehl, Pierre Poulin, Fabrice Neyret, and Victor Ostromoukhov. Linear Efficient Antialiased Displacement and Reflectance Mapping. *ACM Transactions on Graphics*, 32(6), September 2013.

[HBP13] Eric Heitz, Christophe Bourlier, and Nicolas Pinel. Correlation effect between transmitter and receiver azimuthal directions on the illumination function from a random rough surface. *Waves in Random and Complex Media*, 23(3):318–335, 2013.

[MHH+12] Stephen McAuley, Stephen Hill, Naty Hoffman, Yoshiharu Gotanda, Brian Smits, Brent Burley, and Adam Martinez. Practical physically-based shading in film and game production. In *ACM SIGGRAPH 2012 Courses*, SIGGRAPH '12, pages 10:1–7. ACM, 2012.

[MHM+13] S. McAuley, S. Hill, A. Martinez, R. Villemin, M. Pettineo, D. Lazarov, D. Neubelt, B. Karis, C. Hery, Naty Hoffman, and H. Zap Andersson. Physically based shading in theory and practice. In *ACM SIGGRAPH 2013 Courses*, SIGGRAPH '13, pages 22:1–8. ACM, 2013.

[OB10] Marc Olano and Dan Baker. Lean mapping. In *Proceedings of the 2010 ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, I3D '10, pages 181–188, New York, NY, USA, 2010. ACM.

[ON94] Michael Oren and Shree K. Nayar. Generalization of Lambert's reflectance model. In *Proc. SIGGRAPH '94*, pages 239–246, 1994.

[RDP05] Vincent Ross, Denis Dion, and Guy Potvin. Detailed analytical approach to the gaussian surface bidirectional reflectance distribution function specular component applied to the sea surface. *J. Opt. Soc. Am. A*, 22(11):2442–2453, Nov 2005.

[RMS+08] Martin Rump, Gero Müller, Ralf Sarlette, Dirk Koch, and Reinhard Klein. Photo-realistic rendering of metallic car paint from image-based measurements. *Computer Graphics Forum*, 27(2):527–536, April 2008.

[Smi67] B. Smith. Geometrical shadowing of a random rough surface. *IEEE Trans. on Antennas and Propagation*, 15:668–671, 1967.

[vGSK98] Bram van Ginneken, Marigo Stavridi, and Jan J. Koenderink. Diffuse and specular reflectance from rough surfaces. *Appl. Opt.*, 37(1):130–139, Jan 1998.

[WMLT07] B. Walter, S. R. Marschner, H. Li, and K. E. Torrance. Microfacet models for refraction through rough surfaces. In *Proc. Eurographics Symposium on Rendering*, EGSR'07, pages 195–206, 2007.

A Derivation of the Masking Function

In this section, we derive $\bar{G}_1^+(\omega_o)$ (denoted $\bar{G}_1^+$ for convenience) starting from Equation (8):

$$\cos\theta_o = \bar{G}_1^+ \int_\Omega \langle\omega_o, \omega_n\rangle\, D(\omega_n)\, d\omega_n.$$

Slope/Normal Transformations The most complicated step consists of computing the integral, which is defined in the space of the normals. It is more convenient to solve this integral in slope space. We recall that the surface slope associated with a normal $\omega_n = (x_n, y_n, z_n)$ is defined by:

$$\tilde{n}(\omega_n) = (x_{\tilde{n}}, y_{\tilde{n}}) = (-x_n/z_n, -y_n/z_n)$$

and reciprocally:

$$\omega_n(\tilde{n}) = (x_n, y_n, z_n) = \frac{1}{\sqrt{x_{\tilde{n}}^2 + y_{\tilde{n}}^2 + 1}}(-x_{\tilde{n}}, -y_{\tilde{n}}, 1)$$

and that the slope distribution P_{22} is linked to the normal distribution by the relationship:[4]

$$P_{22}(\tilde{n})\, d\tilde{n} = (\omega_n \cdot \omega_g)\, D(\omega_n)\, d\omega_n.$$

By using this change of variable in Equation (8), we write:

$$\int_\Omega \langle\omega_o, \omega_n\rangle\, D(\omega_n)\, d\omega_n = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \frac{\langle\omega_o, \omega_n(\tilde{n})\rangle}{\omega_g \cdot \omega_n(\tilde{n})}\, P_{22}(\tilde{n})\, d\tilde{n},$$

where $[-\infty, +\infty]^2$ is the cartesian 2D space where the slopes are defined. Since $\omega_g = (0, 0, 1)$, we get:

$$\omega_g \cdot \omega_n(\tilde{n}) = \frac{1}{\sqrt{x_{\tilde{n}}^2 + y_{\tilde{n}}^2 + 1}},$$

The clamped dot product can be expanded as:

$$\langle\omega_o, \omega_n(\tilde{n})\rangle = \frac{\chi^+(-x_o x_{\tilde{n}} - y_o y_{\tilde{n}} + z_o)\,(-x_o x_{\tilde{n}} - y_o y_{\tilde{n}} + z_o)}{\sqrt{x_{\tilde{n}}^2 + y_{\tilde{n}}^2 + 1}}$$

and so the integral becomes:

$$\begin{aligned}&\int_\Omega \langle\omega_n, \omega_o\rangle\, D(\omega_n)\, d\omega_n \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \chi^+(-x_o x_{\tilde{n}} - y_o y_{\tilde{n}} + z_o)\,(-x_o x_{\tilde{n}} - y_o y_{\tilde{n}} + z_o)\, P_{22}(x_{\tilde{n}}, y_{\tilde{n}})\, dx_{\tilde{n}} dy_{\tilde{n}}\end{aligned}$$

[4] The Jacobian of the normal to slope transformation is $\left\|\frac{\partial\omega_n}{\partial\tilde{n}}\right\| = |\omega_n \cdot \omega_g|^3$ and we use it to derive the slope distribution $P_{22}(\tilde{n}) = |\omega_n \cdot \omega_g|^4\, D(\omega_n)$.

Without loss of generality, we can assume that the view direction is aligned to the x-axis (i.e., $\omega_o = (\sin\theta_0, 0, \cos\theta_o)$):

$$
\begin{aligned}
&\int_\Omega \langle \omega_n, \omega_o \rangle \, D(\omega_n) \, d\omega_n \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \chi^+(-\sin\theta_o x_{\tilde{n}} + \cos\theta_o) \, (-\sin\theta_o x_{\tilde{n}} + \cos\theta_o) \, P_{22}(x_{\tilde{n}}, y_{\tilde{n}}) \, dx_{\tilde{n}} dy_{\tilde{n}} \\
&= \int_{-\infty}^{+\infty} \chi^+(-\sin\theta_o x_{\tilde{n}} + \cos\theta_o) \, (-\sin\theta_o x_{\tilde{n}} + \cos\theta_o) \left(\int_{-\infty}^{+\infty} P_{22}(x_{\tilde{n}}, y_{\tilde{n}}) \, dy_{\tilde{n}} \right) dx_{\tilde{n}} \\
&= \int_{-\infty}^{+\infty} \chi^+(-\sin\theta_o x_{\tilde{n}} + \cos\theta_o) \, (-\sin\theta_o x_{\tilde{n}} + \cos\theta_o) \, P_2(x_{\tilde{n}}) \, dx_{\tilde{n}}
\end{aligned}
$$

where $P_2(x_{\tilde{n}}) = \int_{-\infty}^{+\infty} P_{22}(x_{\tilde{n}}, y_{\tilde{n}}) \, dy_{\tilde{n}}$ is the 1D slope distribution in the view direction (aligned with the x-axis). Since:

$$
-\sin\theta_o x_{\tilde{n}} + \cos\theta_o > 0 \Rightarrow x_{\tilde{n}} < \cot\theta_o,
$$

we can drop the Heaviside function by changing the integration domain:

$$
\begin{aligned}
&\int_{-\infty}^{+\infty} \chi^+(-\sin\theta_o x_{\tilde{n}} + \cos\theta_o) \, (-\sin\theta_o x_{\tilde{n}} + \cos\theta_o) \, P_2(x_{\tilde{n}}) \, dx_{\tilde{n}} \\
&= \int_{-\infty}^{\cot\theta_o} (-\sin\theta_o x_{\tilde{n}} + \cos\theta_o) \; P_2(x_{\tilde{n}}) \, dx_{\tilde{n}}.
\end{aligned}
$$

Now we can return to Equation (8):

$$
\cos\theta_o = \bar{G}_1^+ \int_{-\infty}^{\cot\theta_o} (-\sin\theta_o x_{\tilde{n}} + \cos\theta_o) \; P_2(x_{\tilde{n}}) \, dx_{\tilde{n}}.
$$

By dividing by $\sin\theta_o$ on both sides, we get:

$$
\cot\theta_o = \bar{G}_1^+ \int_{-\infty}^{\cot\theta_o} (-x_{\tilde{n}} + \cot\theta_o) \; P_2(x_{\tilde{n}}) \, dx_{\tilde{n}}.
$$

Since microfacet distributions are centered, the average slope in any direction is zero ($\int_{-\infty}^{+\infty} P_2(x_{\tilde{n}}) \, dx_{\tilde{n}} = 0$) and we can introduce this term in the equation:

$$
\cot\theta_o = \bar{G}_1^+ \int_{-\infty}^{+\infty} x_{\tilde{n}} \, P_2(x_{\tilde{n}}) \, dx_{\tilde{n}} + \bar{G}_1^+ \int_{-\infty}^{\cot\theta_o} (-x_{\tilde{n}} + \cot\theta_o) \; P_2(x_{\tilde{n}}) \, dx_{\tilde{n}}
$$

and by using $\cot\theta_o = (1 - \bar{G}_1^+) \cot\theta_o + \bar{G}_1^+ \cot\theta_o$:

$$
\begin{aligned}
(1 - \bar{G}_1^+) \cot\theta_o + \bar{G}_1^+ \cot\theta_o &= \bar{G}_1^+ \int_{-\infty}^{+\infty} x_{\tilde{n}} \, P_2(x_{\tilde{n}}) \, dx_{\tilde{n}} + \bar{G}_1^+ \int_{-\infty}^{\cot\theta_o} (-x_{\tilde{n}} + \cot\theta_o) \; P_2(x_{\tilde{n}}) \, dx_{\tilde{n}} \\
(1 - \bar{G}_1^+) \cot\theta_o &= \bar{G}_1^+ \int_{-\infty}^{+\infty} x_{\tilde{n}} \, P_2(x_{\tilde{n}}) \, dx_{\tilde{n}} + \bar{G}_1^+ \int_{-\infty}^{\cot\theta_o} (-x_{\tilde{n}} + \cot\theta_o) \; P_2(x_{\tilde{n}}) \, dx_{\tilde{n}} - \bar{G}_1^+ \cot\theta_o
\end{aligned}
$$

and since P_2 integrates to 1 we have $\bar{G}_1^+ \cot\theta_o = \bar{G}_1^+ \int_{-\infty}^{+\infty} \cot\theta_o \, P_2(x_{\tilde{n}})\, dx_{\tilde{n}}$:

$$\begin{aligned}
(1-\bar{G}_1^+)\cot\theta_o &= \bar{G}_1^+ \int_{-\infty}^{+\infty} x_{\tilde{n}}\, P_2(x_{\tilde{n}})\, dx_{\tilde{n}} + \bar{G}_1^+ \int_{-\infty}^{\cot\theta_o} \left(-x_{\tilde{n}} + \cot\theta_o\right) P_2(x_{\tilde{n}})\, dx_{\tilde{n}} \\
&\quad - \bar{G}_1^+ \int_{-\infty}^{+\infty} \cot\theta_o \, P_2(x_{\tilde{n}})\, dx_{\tilde{n}} \\
&= \bar{G}_1^+ \left(\int_{-\infty}^{+\infty} x_{\tilde{n}}\, P_2(x_{\tilde{n}})\, dx_{\tilde{n}} - \int_{-\infty}^{\cot\theta_o} x_{\tilde{n}}\, P_2(x_{\tilde{n}})\, dx_{\tilde{n}} \right) \\
&\quad + \bar{G}_1^+ \left(\int_{-\infty}^{\cot\theta_o} \cot\theta_o \, P_2(x_{\tilde{n}})\, dx_{\tilde{n}} - \int_{-\infty}^{+\infty} \cot\theta_o \, P_2(x_{\tilde{n}})\, dx_{\tilde{n}} \right) \\
&= \bar{G}_1^+ \int_{\cot\theta_o}^{+\infty} x_{\tilde{n}}\, P_2(x_{\tilde{n}})\, dx_{\tilde{n}} - \bar{G}_1^+ \int_{\cot\theta_o}^{+\infty} \cot\theta_o \, P_2(x_{\tilde{n}})\, dx_{\tilde{n}} \\
&= \bar{G}_1^+ \int_{\cot\theta_o}^{\infty} \left(x_{\tilde{n}} - \cot\theta_o\right) P_2(x_{\tilde{n}})\, dx_{\tilde{n}}
\end{aligned}$$

By dividing by $\bar{G}_1^+$ on each side, we get:

$$\frac{(1-\bar{G}_1^+)}{\bar{G}_1^+} = \frac{1}{\cot\theta_o} \int_{\cot\theta_o}^{\infty} \left(x_{\tilde{n}} - \cot\theta_o\right) P_2(x_{\tilde{n}})\, dx_{\tilde{n}}$$

which leads to the final form:

$$\boxed{\bar{G}_1^+(\omega_o) = \frac{1}{1+\Lambda(\omega_o)}}$$

where function Λ is defined by:

$$\boxed{\Lambda(\omega_o) = \frac{1}{\cot\theta_o} \int_{\cot\theta_o}^{\infty} \left(x_{\tilde{n}} - \cot\theta_o\right) P_2(x_{\tilde{n}})\, dx_{\tilde{n}}}$$

Our derivation, based on the projected area, has lead us to the generalized form of Smith's masking term [Bro80, WMLT07].

B Derivation of the Height-Correlated Masking and Shadowing Function

In this section, we recall the derivation of the height correlated form of the joint masking-shadowing function [RDP05, HBP13, DHI+13] presented in Equation (21):

$$G_2(\omega_o, \omega_i, \omega_n) = \frac{\chi^+(\omega_o \cdot \omega_n)\ \chi^+(\omega_i \cdot \omega_n)}{1 + \Lambda(\omega_o) + \Lambda(\omega_i)}.$$

The microsurface is defined by the normal distribution $D(\omega_n)$, and the associated slope distribution is $P_{22}(\tilde{n})$ as presented in Appendix A. We introduce $P_1(\xi)$, the height distribution of the microsurface. Note that the slopes of the microsurface are simply the gradients of the heights: $\tilde{n} = \nabla\xi$. Smith's derivation [Smi67, WMLT07] gives the probability that a point at height ξ with nonbackface-culled normal ω_n is visible from direction ω_o:

$$G_1(\omega_o, \omega_n, \xi_0) = \chi^+(\omega_o \cdot \omega_n) \left(\int_{-\infty}^{\xi_0} P_1(\xi)\, d\xi \right)^{\Lambda(\omega_o)}.$$

The height averaged form is given by:

$$\begin{aligned}
G_1(\omega_o, \omega_n) &= \int_{-\infty}^{+\infty} G_1(\omega_o, \omega_n, \xi_0)\, P_1(\xi_0)\, d\xi_0 \\
&= \int_{-\infty}^{+\infty} \chi^+(\omega_o \cdot \omega_n) \left(\int_{-\infty}^{\xi_0} P_1(\xi)\, d\xi \right)^{\Lambda(\omega_o)} P_1(\xi_0)\, d\xi_0 \\
&= \chi^+(\omega_o \cdot \omega_n) \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{\xi_0} P_1(\xi)\, d\xi \right)^{\Lambda(\omega_o)} P_1(\xi_0)\, d\xi_0 \\
&= \frac{\chi^+(\omega_o \cdot \omega_n)}{1 + \Lambda(\omega_o)}.
\end{aligned}$$

which is Smith's masking function from Equation (10). Now, if we suppose that there is no directional correlation for masking from directions ω_o and ω_i, then the probability that a point at height ξ is visible from both directions is just the product of the probabilities:

$$\begin{aligned}
G_2(\omega_o, \omega_i, \omega_n, \xi_0) &= G_1(\omega_o, \omega_n, \xi_0)\, G_1(\omega_i, \omega_n, \xi_0) \\
&= \chi^+(\omega_o \cdot \omega_n) \left(\int_{-\infty}^{\xi_0} P_1(\xi)\, d\xi \right)^{\Lambda(\omega_o)} \chi^+(\omega_i \cdot \omega_n) \left(\int_{-\infty}^{\xi_0} P_1(\xi)\, d\xi \right)^{\Lambda(\omega_i)} \\
&= \chi^+(\omega_o \cdot \omega_n)\ \chi^+(\omega_i \cdot \omega_n) \left(\int_{-\infty}^{\xi_0} P_1(\xi)\, d\xi \right)^{\Lambda(\omega_o)+\Lambda(\omega_i)},
\end{aligned}$$

and the height averaged form is given by:

$$
\begin{aligned}
G_2(\omega_o, \omega_i, \omega_n) &= \int_{-\infty}^{+\infty} G_2(\omega_o, \omega_i, \omega_n)\, P_1(\xi_0)\, d\xi_0 \\
&= \int_{-\infty}^{+\infty} \chi^+(\omega_o \cdot \omega_n)\ \chi^+(\omega_i \cdot \omega_n) \left(\int_{-\infty}^{\xi_0} P_1(\xi)\, d\xi \right)^{\Lambda(\omega_o)+\Lambda(\omega_i)} P_1(\xi_0)\, d\xi_0 \\
&= \chi^+(\omega_o \cdot \omega_n)\ \chi^+(\omega_i \cdot \omega_n) \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{\xi_0} P_1(\xi)\, d\xi \right)^{\Lambda(\omega_o)+\Lambda(\omega_i)} P_1(\xi_0)\, d\xi_0 \\
&= \frac{\chi^+(\omega_o \cdot \omega_n)\ \chi^+(\omega_i \cdot \omega_n)}{1 + \Lambda(\omega_o) + \Lambda(\omega_i)},
\end{aligned}
$$

which is the height correlated masking-shadowing function presented in Equation (21).

C MATLAB Code for the Weak White Furnace Test

In this section, we provide code to numerically compute the integral in Equation (19):

$$\int_{\Omega} \frac{G_1(\omega_o, \omega_h)\, D(\omega_h)}{4\,|\omega_g \cdot \omega_o|}\, d\omega_i = 1,$$

with Beckmann and GGX distributions and their associated Smith masking functions.

```
function [] = TEST_BECKMANN(alpha, theta_o)

% view vector
V = [sin(theta_o) 0 cos(theta_o)];
% masking (rational approximation for Lambda)
a = 1 / (alpha * tan(theta_o));
if a < 1.6
  Lambda = (1 - 1.259*a + 0.396*a^2) / (3.535*a + 2.181*a^2);
else
  Lambda = 0;
endif
G = 1 / (1 + Lambda);

integral = 0;
dtheta = 0.05;
dphi = 0.05;
for theta = 0:dtheta:pi
for phi = 0:dphi:2*pi
  % reflected vector
  L = [cos(phi)*sin(theta) sin(phi)*sin(theta) cos(theta)];
  % half vector
  H = (V + L) / norm(V + L);
  % Beckmann distribution
  if H(3) > 0
    % angle associated with H
    theta_h = acos(H(3));
    D = exp(-(tan(theta_h)/alpha)^2) / (pi * alpha^2 * H(3)^4);
  else
    continue;
  endif
  % integrate
  integral += sin(theta) * D * G / abs(4 * V(3));
end
end

% display integral (should be 1)
integral *= dphi * dtheta
endfunction
```

```
function [] = TEST_BECKMANN_ANISO(alpha_x, alpha_y, theta_o, phi_o)

% view vector
V = [cos(phi_o)*sin(theta_o) sin(phi_o)*sin(theta_o) cos(theta_o)];
% alpha in view direction
alpha_o = sqrt(cos(phi_o)^2*alpha_x^2 + sin(phi_o)^2*alpha_y^2);
% masking (rational approximation for Lambda)
a = 1 / (alpha_o * tan(theta_o));
if a < 1.6
  Lambda = (1 - 1.259*a + 0.396*a^2) / (3.535*a + 2.181*a^2);
else
  Lambda = 0;
endif
G = 1 / (1 + Lambda);

integral = 0;
dtheta = 0.05;
dphi = 0.05;
for theta = 0:dtheta:pi
for phi = 0:dphi:2*pi
  % reflected vector
  L = [cos(phi)*sin(theta) sin(phi)*sin(theta) cos(theta)];
  % half vector
  H = (V + L) / norm(V + L);
  % Beckmann distribution
  if H(3) > 0
    % slope associated with H
    slope = [-H(1)/H(3) -H(2)/H(3)];
    D = exp(-(slope(1)/alpha_x)^2 - (slope(2)/alpha_y)^2);
    D /= pi * alpha_x * alpha_y * H(3)^4;
  else
    continue;
  endif
  % integrate
  integral += sin(theta) * D * G / abs(4 * V(3));
end
end

% display integral (should be 1)
integral *= dphi * dtheta
endfunction
```

```
function [] = TEST_GGX(alpha, theta_o)

% view vector
V = [sin(theta_o) 0 cos(theta_o)];
% masking
a = 1 / (alpha * tan(theta_o));
Lambda = (-1 + sqrt(1 + 1/a^2)) / 2;
G = 1 / (1 + Lambda);

integral = 0;
dtheta = 0.05;
dphi = 0.05;
for theta = 0:dtheta:pi
for phi = 0:dphi:2*pi
  % reflected vector
  L = [cos(phi)*sin(theta) sin(phi)*sin(theta) cos(theta)];
  % half vector
  H = (V + L) / norm(V + L);
  % GGX distribution
  if H(3) > 0
    % angle associated with H
    theta_h = acos(H(3));
    D = 1 / (1 + (tan(theta_h)/alpha)^2)^2;
    D /= pi * alpha^2 * H(3)^4;
  else
    D = 0;
  endif
  % integrate
  integral += sin(theta) * D * G / abs(4 * V(3));
end
end

% display integral (should be 1)
integral *= dphi * dtheta
endfunction
```

```
function [] = TEST_GGX_ANISO(alpha_x, alpha_y, theta_o, phi_o)

% view vector
V = [cos(phi_o)*sin(theta_o) sin(phi_o)*sin(theta_o) cos(theta_o)];
% alpha in view direction
alpha_o = sqrt(cos(phi_o)^2*alpha_x^2 + sin(phi_o)^2*alpha_y^2);
% masking
a = 1 / (alpha_o * tan(theta_o));
Lambda = (-1 + sqrt(1 + 1/a^2)) / 2;
G = 1 / (1 + Lambda);

integral = 0;
dtheta = 0.05;
dphi = 0.05;
for theta = 0:dtheta:pi
for phi = 0:dphi:2*pi
  % reflected vector
  L = [cos(phi)*sin(theta) sin(phi)*sin(theta) cos(theta)];
  % half vector
  H = (V + L) / norm(V + L);
  % GGX distribution
  if H(3) > 0
    % slope associated with H
    slope = [-H(1)/H(3) -H(2)/H(3)];
    D = 1/(1 + (slope(1)/alpha_x)^2 + (slope(2)/alpha_y)^2)^2;
    D /= pi * alpha_x * alpha_y * H(3)^4;
  else
    D = 0;
  endif
  % integrate
  integral += sin(theta) * D * G / abs(4 * V(3));
end
end

% display integral (should be 1)
integral *= dphi * dtheta
endfunction
```

$\boxed{\textit{Warning!}}$ The values `dtheta` and `dphi` used to discretize the BRDF in the numerical integration are hardcoded. In practice, setting them to 0.05 works well for `alpha` > 0.2. If `alpha` is smaller than 0.2 then `dtheta` and `dphi` must be set to smaller values as well, in order to correctly capture the sharp BRDF lobe.

RESEARCH CENTRE
GRENOBLE – RHÔNE-ALPES

Inovallée
655 avenue de l'Europe Montbonnot
38334 Saint Ismier Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399