Understanding the Masking-Shadowing Function in Microfacet-Based BRDFs

Eric Heitz
INRIA ; CNRS ; Univ. Grenoble Alpes

Abstract

We provide a new presentation of the masking-shadowing functions (or geometric attenuation factors) in microfacet-based BRDFs and answer some common questions about their applications. Our main motivation is to define a correct (geometrically indicated), physically based masking function for application in microfacet models, as well as the properties that function should exhibit. Indeed, several different masking functions are often presented in the literature and making the right choice is not always obvious. We start by showing that physically based masking functions are constrained by the projected area of the visible microsurface onto the outgoing direction. We use this property to derive the distribution of visible normals from the microsurface, whose normalization factor is the masking function. We then show how the common form of microfacet-based BRDFs emerges from this distribution. As a consequence, the masking function is related to the correct normalization of microfacet-based BRDFs. However, while the correct masking function satisfies these normalization constraints, its explicit form is can only be determined for a given microsurface profile.

Our derivation emphasizes that under the assumptions of their respective microsurface profiles, both Smith's function and the V-cavity masking function are correct. However, we show that the V-cavity microsurface yields results that miss the effect of occlusion, making it analogous to the shading of a normal map instead of a displacement map. This observation explains why the V-cavity model yields incorrect glossy highlights at grazing view angles.

We also review other common masking functions, which are not associated with a microsurface profile and thus are not physically based. The insights gained from these observations motivate new research directions in the field of microfacet theory. For instance, we show that masking functions are stretch invariant and we show how this property can be used to derive the masking function for anisotropic microsurfaces in a straightforward way. We also discuss future work such as the incorporation of multiple scattering on the microsurface into BRDF models.

 ISSN 2331-7418

1. Introduction

Microfacet theory was originally developed in the field of optical physics to study scattering on statistical surfaces [Beckmann and Spizzichino 1963]. In the graphics community, we use it to derive physically based bidirectional reflectance distribution functions (BRDFs) [Cook and Torrance 1982; Oren and Nayar 1994; Walter et al. 2007], which are used widely in both real-time and production rendering. Today, microfacet theory is a fundamental background topic in computer graphics. For instance, for the past two years, the SIGGRAPH course on physically based rendering began with an introduction to microfacet theory [McAuley et al. 2012; McAuley et al. 2013], with the goal of providing the main intuitions derived from the underlying physics. Other considerations, such as flexibility for artistic direction and computational efficiency, are also discussed throughout that course. Microfacets are an area of continuing development because the combination of different components in microfacet-based BRDFs offer a wide range of possibilities. So, the right choices for each of those components are frequently not obvious and in our experience, are a common source of confusion in the field.

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: 2.5, 3.6, and 4.4. Implementors may wish to jump directly to those sections. The remainder of the document is for readers seeking to develop their intuition for 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 one model over another; we aim at providing background knowledge about 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 computer graphics community; we focus on understanding their physical properties.

On The Meaning Of "Physically Based" Regarding Microfacet Models A physical model is a simplified representation of a system or a physical phenomenon that permits to analyze it, explain it, and make predictions about its behavior.

In microfacet theory, the model subject to the study is, at the macroscopic scale, a flat geometric surface, whose interface is, at the microscopic scale, rough and composed of microfacets. This representation is used to explain and make predictions regarding the scattering events occurring at the geometric surface interface, i.e. how the light that intersects the geometric surface is reflected and scattered in other directions.

A meaningful microsurface model is described by a distribution of normals, which models how the microfacets are statistically oriented, and a microsurface profile, which models how the microfacets are organized on the microsurface. Microfacet BRDFs whose equations are derived from such a meaningful microsurface model are called "physically based" precisely because they are based on a microsurface model. Reciprocally, if there is no microsurface model from which the BRDF equations can be derived, then it is not a "physically based" microfacet model. The masking and shadowing functions are part of microfacet BRDFs. They give the probability that a microfacet is visible either in the outgoing direction (masking) or in the incident direction (shadowing). As for the BRDF, microfacet masking and shadowing functions can be called "physically based" only if they are derived from a microsurface model.

In this article, we explain how microsurface models are formally described and how physically based masking and shadowing functions are derived from them. We also show how this leads to the derivation of the associated physically based BRDFs.

However, it should be noted that that microfacet models are simply that: *models*. They are always based on some assumptions regarding the microsurface optical behavior, e.g. only geometrical optics, perfect mirror or diffuse reflection, no multiple scattering, etc. Thus, one needs to keep in mind that calling them "physically based" does not mean that they can precisely predict measurements from a real physical surface. In cases where those assumptions are wrong, it could even be that empirical models are sometimes more accurate than the mathematically rigorous "physically based" ones when compared to measured data.

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

- The Smith [1967] masking function is one of the most famous ones from the computer graphics literature. However, what is less known is that at the end of his article Smith points out that his function has the property of ensuring the conservation of the *visible projected area*, a property that is expected from a correct masking function.

- Ashikhmin et al. [2000] also observe that the visible projected area is a quantity that is conserved from the geometric surface to the microsurface. They use this knowledge to derive the general equation for a correct masking term, which ensures correct normalization and energy conservation. By doing this, they were actually reinventing the Smith masking function without being aware of it. Indeed, 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. [2005] propose a study of the reflectance of the sea. They model

the sea with a Gaussian rough surface (Beckmann distribution) and compute a normalized BRDF incorporating the Smith masking and shadowing functions. During the derivation, they observe that on Gaussian surfaces, the normalization coefficients of the BRDF and the Smith masking functions have similar expressions that cancel. They note that this property is convenient for computational purposes, but they do not provide a physical reason as to why this might (or must) occur.

In this document, we propose a unified microfacet framework where all those previous results (from the masking function to the entire BRDF) are directly derived from the conservation of the visible projected area.

In Section 2, we introduce the microfacet statistical quantities and derive the equation of the conservation of the visible projected area, that is satisfied by correct masking functions.

In Section 3, we introduce the *distribution of visible normals* and show how common BRDF models can be derived from this distribution. The reason why microfacet BRDFs require shadowing is that they only model the first scattering event occurring on the microsurface. Common microfacet BRDFs do not model multiple scattering and are not normalized for this reason, i.e., they do not integrate to exactly 1 (even when they model a perfectly reflective surface). Starting from this observation, we propose a normalization test—which 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 4, we instantiate the equations derived in the previous sections with the Smith and V-cavity microsurface profiles and compare the properties of their respective BRDFs. While our derivations do not provide new results, they have the advantage of emphasizing that the results are exact rather than approximate, and show how masking is related to the concept of the visible projected area on arbitrary stochastic surfaces. We also review other common masking functions, which are not derived from a microsurface model and thus are not exact nor physically based.

In Section 5, we demonstrate for the first time the *stretch 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 distributions of normals. This eases generalization to anisotropy of several previous results.

In Section 6, we discuss the properties of the Smith masking function used for shadowing and recall several masking-shadowing models that handle different types of correlation.

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 from our investigation.

Ω	spherical domain (4π steradians)
(θ, ϕ)	spherical coordinates: $\omega = (\cos\phi\sin\theta, \sin\phi\sin\theta, \cos\theta)$
$\omega_o = (x_o, y_o, z_o)$	outgoing direction
$\omega_i = (x_i, y_i, z_i)$	incident direction
$\omega_h = \frac{\omega_o + \omega_i}{\|\omega_o + \omega_i\|}$	half vector
$\mathcal{G}$	geometric surface
$\omega_g = (0,0,1)$	geometric normal
$\mathcal{M}$	microsurface
p_m	microsurface point
$\omega_m = (x_m, y_m, z_m)$	microsurface normal
$\omega_1 \cdot \omega_2$	dot product
$\lvert\omega_1 \cdot \omega_2\rvert$	absolute value of the dot product
$\langle\omega_1, \omega_2\rangle$	clamped dot product: 0 if $\omega_1 \cdot \omega_2 < 0$
$\chi^+(a)$	Heaviside function: 1 if $a > 0$ and 0 if $a \leq 0$
$\delta(a)$	Dirac delta distribution: $\int_{-\infty}^{+\infty} \delta(a)\, da = 1$

Table 1. Mathematical notation.

$L(\omega_o)$	radiance in direction ω_o	$\mathrm{W/sr/m^2}$
$L(\omega_o, p_m)$	radiance in direction ω_o at point p_m	$\mathrm{W/sr/m^2}$
$L(\omega_o, \mathcal{M})$	radiance in direction ω_o on microsurface $\mathcal{M}$	$\mathrm{W/sr/m^2}$
$\rho(\omega_o, \omega_i)$	BRDF	$1/\mathrm{sr}$
$\rho_{\mathcal{M}}(\omega_o, \omega_i, \omega_z)$	micro-BRDF of normal ω_m	$1/\mathrm{sr}$
$D(\omega_m)$	distribution of normals	$\mathrm{m^2/sr}$
$D_{\omega_o}(\omega_m)$	distribution of visible normals from ω_o	$1/\mathrm{sr}$
$G_1(\omega_o, p_m)$	spatial masking function at point p_m (binary value)	-
$G_1(\omega_i, p_m)$	spatial shadowing function at point p_m (binary value)	-
$G_2(\omega_o, \omega_i, p_m)$	spatial masking-shadowing function at point p_m (binary value)	-
$G_1(\omega_o, \omega_m)$	statistical masking function of normal ω_m (in $[0,1]$)	-
$G_1(\omega_i, \omega_m)$	statistical shadowing function of normal ω_m (in $[0,1]$)	-
$G_2(\omega_o, \omega_i, \omega_m)$	statistical masking-shadowing function of normal ω_m (in $[0,1]$)	-

Table 2. Physical quantities and the associated units ('-' denotes unitless quantities).

2. Derivation of the Masking Function

In this section, we show how the projected area of the microsurface can be used to introduce a constraint on physically based masking functions, following Ashikhmin et al. [2000]. We start by defining the concept of projected area (2.1) and show why it is essential to the measure of radiance. Then, we define the statistical framework of microfacet theory (2.2). The conservation of the projected area (2.3) gives a new microfacet equation that we use to constrain the masking function (2.4). This constraint, tied to the choice of a microsurface profile, leads to the derivation of physically based masking functions.

2.1. Measuring Radiance on a Surface

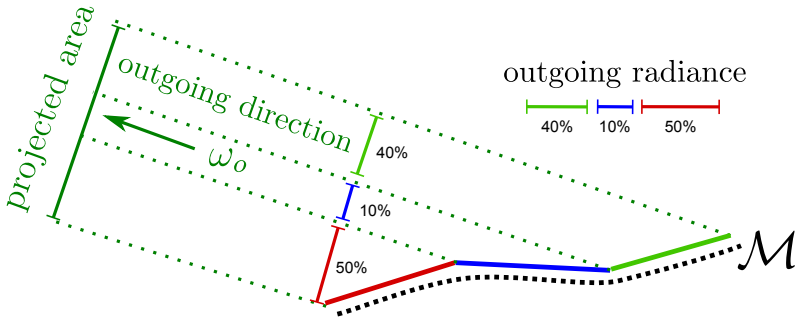


Figure 1. The outgoing radiance of surface $\mathcal{M}$ is the average of the radiances from each point of the surface, weighted by their projected-area fractions towards the outgoing direction.

Radiance is the energy density traveling through an area from a solid angle. It is measured in watts per steradian per square meter ($\mathrm{W/sr/m^2}$). The outgoing radiance $L(\omega_o, \mathcal{M})$ of a given surface $\mathcal{M}$ in direction ω_o, is the integral of the radiances $L(\omega_o, p_m)$ from each patch with center point p_m on the surface, measured from outgoing direction ω_o, and weighted by its projected area observed from that outgoing direction (as shown in Figure 1):

$$L(\omega_o, \mathcal{M}) = \frac{\int_{\mathcal{M}} \text{projected area}(p_m)\, L(\omega_o, p_m)\, dp_m}{\int_{\mathcal{M}} \text{projected area}(p_m)\, dp_m}. \tag{1}$$

The area of each surface point projected in the outgoing direction is a view-dependent weighting factor and the integral $\int_{\mathcal{M}} \text{projected area}(p_m)\, dp_m$, is the normalization coefficient of the projected-area fractions. Note that this normalization coefficient gives the entire expression radiance units; without it, the result would be missing the area units in the denominator.

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

2.2. Microfacet Statistics

We consider a planar region of a surface, which we call the "geometric surface" $\mathcal{G}$, whose area is 1m^2 by convention: $\int_{\mathcal{G}} dp_g = 1\text{m}^2$. The microfacet model assumes that the true surface is offset from this in the form of a collection of microfacets, which we call the "microsurface" $\mathcal{M}$. To be precise: if ω_g is the normal of the geometry $\mathcal{G}$, then $\mathcal{M}$ is the set of microfacet points that project onto $\mathcal{G}$ along ω_g. Each point p_m of the microsurface $\mathcal{M}$ has a normal vector $\omega_m(p_m)$, i.e. $\omega_m : \mathcal{M} \to \Omega$ is a function from a point on the microsurface to the the surface normal vector at that point. We express the three coordinates of this vector as (x_m, y_m, z_m).

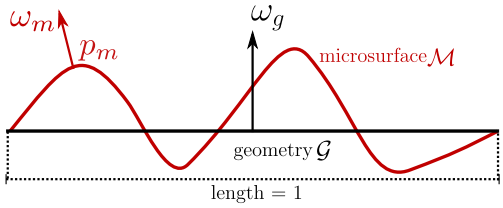


Figure 2. The geometric surface and the microsurface.

Microfacet theory is a statistical model of the scattering properties of the microsurface. Thus, writing *statistical* rather than *spatial* equations is more convenient for this study. In microfacet theory, the statistics are defined in the space of the normals, which is the spherical domain Ω.

The Distribution of Normals To relate integrals over the microsurface to integrals over the sphere—i.e., to convert from spatial to statistical integrals—we need a tool that measures the change-of-area as we switch domains. The *distribution of normals*[1] provides this. It is expressed in square meters per steradian (m^2/sr) and defined as

$$D(\omega) = \int_{\mathcal{M}} \delta_\omega(\omega_m(p_m))\, dp_m, \tag{2}$$

where the units of the Dirac delta distribution are $1/\text{sr}$, the inverse of those of its argument. Consider some region $\Omega' \subset \Omega$ of the unit sphere. Now consider the subset $\mathcal{M}' \subset \mathcal{M}$ of the microsurface that contains all of the points $p_m \in \mathcal{M}$ where the normal $\omega_m(p_m)$ is an element of Ω', such that

$$p_m \in \mathcal{M}' \iff \omega_m(p_m) \in \Omega'. \tag{3}$$

The distribution of normals has the property that its integral, over any region $\Omega' \subset \Omega$ of the unit sphere, gives the area of the set of all points $\mathcal{M}' \subset \mathcal{M}$ whose normals lie

[1] We use this somewhat awkward phrase because "normal distribution" is widely used to mean the distribution function for a Gaussian random variable. Also, note that D is actually the distribution of normals *per square meter of the geometric surface* and this is why it is measured in m^2/sr and not in $1/\text{sr}$, contrary to what is stated in [Walter et al. 2007].

in Ω':

$$\int_{\mathcal{M}'} dp_m = \int_{\Omega'} D(\omega_m)\, d\omega_m. \tag{4}$$

As a consequence, the integral of the distribution of normals is the area of the microsurface:

$$\text{microsurface area} = \int_{\mathcal{M}} dp_m = \int_{\Omega} D(\omega_m)\, d\omega_m \tag{5}$$

Spatial and Statistical Equations As a consequence of the definition of D, if $f(\omega_m)$ is any function of the microsurface normals, then the spatial integration of f can be replaced by a statistical integration:

$$\int_{\mathcal{M}} f(\omega_m(p_m))\, dp_m = \int_{\Omega} f(\omega_m)\, D(\omega_m)\, d\omega_m, \tag{6}$$

where the left-hand side is the *spatial integral* and the right-hand side is the *statistical integral*. This property is used in Figure 3(a), where f is the dot product.

Statistical Functions If $g(p_m)$ is a spatial function defined on the microsurface, we can define the related statistical function $g(\omega_m)$ as

$$g(\omega) = \frac{\int_{\mathcal{M}} \delta_\omega(\omega_m(p_m))\, g(p_m)\, dp_m}{\int_{\mathcal{M}} \delta_\omega(\omega_m(p_m))\, dp_m}. \tag{7}$$

The statistical functions can be used in statistical integrals in the following way:

$$\int_{\mathcal{M}} g(p_m)\, dp_m = \int_{\Omega} g(\omega_m)\, D(\omega_m)\, d\omega_m. \tag{8}$$

This property is used in Figure 3(c), where g is the masking function G_1, which we introduce in Section 2.3.

2.3. Microfacet Projections

(a) Projected area of the microsurface onto the geometric normal ω_g The area of the microsurface projected onto the geometric normal is the area of the geometric surface (Figure 3(a)), whose area is 1m^2 by convention. Hence, the projection of the distribution of normals onto the geometry is normalized:

$$\int_\Omega (\omega_m \cdot \omega_g)\, D(\omega_m)\, d\omega_m = \int_{\mathcal{M}} (\omega_m(p_m) \cdot \omega_g)\, dp_m = \int_{\mathcal{G}} dp_g = 1\text{m}^2. \tag{9}$$

(b) Projected area of the geometric surface onto the outgoing direction ω_o The geometric surface area is 1m^2 and its projected area onto the outgoing direction (Figure 3(b)) is the area multiplied by the cosine of the angle of incidence θ_o:

$$\text{projected area} = (\omega_o \cdot \omega_g)\ .\ \text{area} = \cos\theta_o\ .\ 1\text{m}^2. \tag{10}$$

(c) Projected area of the visible microsurface onto the outgoing direction ω_o We now show that the projected area of the geometric surface onto the outgoing direction is also the projected area of the *visible* microsurface (Figure 3(c)). It is the sum of the projected area of each visible microfacet. The projected area of a microfacet with normal ω_m is the geometric projection factor $\langle \omega_o, \omega_m \rangle$. Note that here we use the clamped dot product $\langle -, - \rangle$ because backfacing 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 *spatial masking function* $G_1(\omega_o, p_m)$ that has binary values: it evaluates to 0 if point p_m is masked and to 1 if it is visible. This gives

$$\text{projected area} = \int_{\mathcal{M}} G_1(\omega_o, p_m)\, \langle \omega_o, \omega_m(p_m) \rangle\, dp_m. \tag{11}$$

The *statistical masking function* $G_1(\omega_o, \omega_m)$ has the range $[0, 1]$ and gives the fraction of microfacets with normal ω_m that are visible along the outgoing direction ω_o:

$$G_1(\omega_o, \omega) = \frac{\int_{\mathcal{M}} \delta_\omega(\omega_m(p_m))\, G_1(\omega_o, p_m)\, dp_m}{\int_{\mathcal{M}} \delta_\omega(\omega_m(p_m))\, dp_m}. \tag{12}$$

The statistical equation is given by

$$\text{projected area} = \int_\Omega G_1(\omega_o, \omega_m)\, \langle \omega_o, \omega_m \rangle\, D(\omega_m)\, d\omega_m. \tag{13}$$

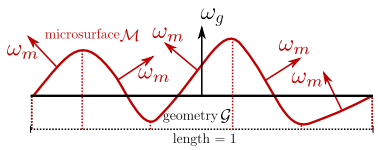
(a) Projected area of the microsurface onto the geometric normal ω_g (in m^2)

Spatial equation:

$$\int_{\mathcal{M}} (\omega_m(p_m) \cdot \omega_g)\, dp_m = \int_{\mathcal{G}} dp_g = 1$$

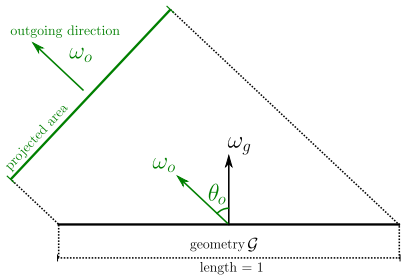
Statistical equation:

$$\int_{\Omega} (\omega_m \cdot \omega_g)\, D(\omega_m)\, d\omega_m = 1$$



(b) Projected area of the geometric surface onto the outgoing direction ω_o (in m^2)

projected area = $(\omega_o \cdot \omega_g)$. area $= \cos\theta_o$



(c) Projected area of the visible microsurface onto the outgoing direction ω_o (in m^2)

Spatial equation:

$$\text{projected area} = \int_{\mathcal{M}} G_1(\omega_o, p_m)\, \langle \omega_o, \omega_m(p_m) \rangle\, dp_m$$

Statistical equation:

$$\text{projected area} = \int_{\Omega} G_1(\omega_o, \omega_m)\, \langle \omega_o, \omega_m \rangle\, D(\omega_m)\, d\omega_m$$

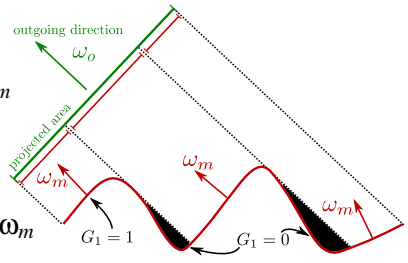


Figure 3. Projections in microfacet theory.

2.4. A Constraint on the Masking Function

Figure 3 emphasizes a fundamental property of microfacet theory: the projected area of the visible microsurface from Equation (13) is exactly the projected area of the geometric surface given in Equation (10). This equivalence imposes a constraint on the statistical masking function, which is formalized by the following equation:

$$\cos\theta_o = \int_\Omega G_1(\omega_o, \omega_m)\, \langle \omega_o, \omega_m \rangle\, D(\omega_m)\, d\omega_m. \tag{14}$$

Physically based masking functions G_1 should always satisfy this constraint. However, this constraint does not entirely determine G_1, since for a fixed outgoing direction ω_o, the masking function is two-dimensional—$G_1(\omega_o, \omega_m)$ is defined for each normal—and there are an infinite number of functions for G_1 that satisfy the equation. In order to reduce the number of solutions to one, we introduce a second constraint: we choose a *microsurface profile*.

An intuitive way to think about this is that the distribution of normals is like a histogram, describing only the *proportion* of each normal on the microsurface. It does not provide information on how they are organized, however—for this, we need a microsurface profile. Furthermore, as Figure 4 illustrates, the choice of profile can have a strong impact on the shape of the resulting BRDF.

Once the microsurface profile has been chosen, the masking function is completely determined, and its exact form can be derived. This is covered Section 4, which reviews the exact form of G_1 obtained with the Smith and V-cavity microsurface profiles.

2.5. 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 based?*"
In this section, we have shown that:

- The projected area of the visible microsurface is equal to the projected area of the geometric surface onto any projection direction.
- The masking function is constrained by this equality. More formally, physically based masking functions always satisfy Equation (14).
- The masking function is not entirely determined by this contraint, however.
- The masking function is entirely determined once the microsurface profile has been chosen.
- The microsurface profile impacts the shape of the BRDF.

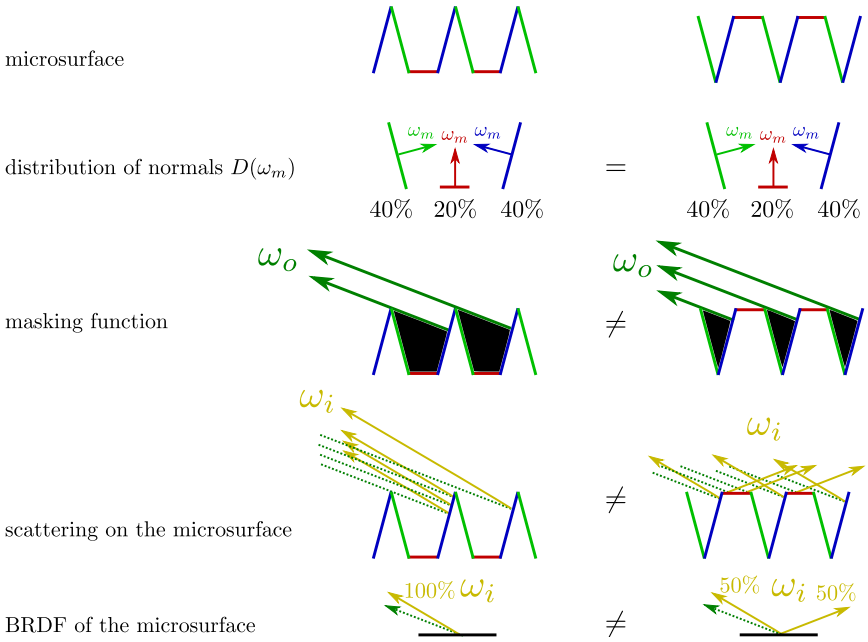


Figure 4. Microsurfaces with the same distribution of normals but with different profiles result in different BRDFs.

3. Microfacet-Based BRDFs

In this section, we define the distribution of visible normals (3.1) and we show how microfacet models are constructed from this distribution in the general case (3.2), and in the specific cases of specular (3.3) and diffuse (3.4) microfacets. We show that the 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 (3.5).

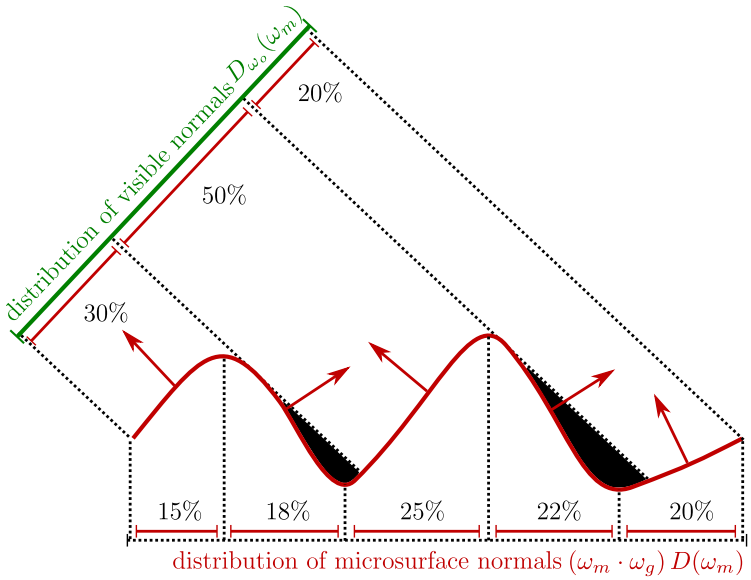


Figure 5. The distribution of microsurface normals $(\omega_m \cdot \omega_g) D(\omega_m)$ is an intrinsic surface property, while the distribution of visible normals $D_{\omega_o}(\omega_m)$ is view dependent.

3.1. Distribution of Visible Normals

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

$$L(\omega_o, \mathcal{M}) = \frac{1}{\cos\theta_o} \int_\Omega L(\omega_o, \omega_m)\, G_1(\omega_o, \omega_m)\, \langle \omega_o, \omega_m \rangle\, D(\omega_m)\, d\omega_m, \qquad (15)$$

where $L(\omega_o, \mathcal{M})$ is the outgoing radiance from the microsurface, $L(\omega_o, \omega_m)$ is the outgoing radiance from the microfacets with normal ω_m, and the factor $\frac{1}{\cos\theta_o}$ is here to normalize the integral by the projected area of the geometric surface. We can see that the outgoing radiance from the microsurface is the sum of the outgoing radiance from each microfacet weighted by what we call *the distribution of visible normals*, as illustrated in Figure 5. It is the distribution of normals weighted by the projected area

of each normal (the clamped cosine, $\langle \omega_o, \omega_m \rangle$) and by the masking function:

$$D_{\omega_o}(\omega_m) = \frac{G_1(\omega_o, \omega_m) \langle \omega_o, \omega_m \rangle D(\omega_m)}{\cos \theta_o}. \tag{16}$$

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

$$L(\omega_o, \mathcal{M}) = \int_\Omega L(\omega_o, \omega_m) D_{\omega_o}(\omega_m) \, d\omega_m, \tag{17}$$

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, by using the result of Equation (14), we can replace $\cos \theta_o$ in Equation (16) and verify that the distribution of normals is normalized:

$$\begin{aligned} \int_\Omega D_{\omega_o}(\omega_m) \, d\omega_m &= \int_\Omega \frac{G_1(\omega_o, \omega_m) \langle \omega_o, \omega_m \rangle D(\omega_m)}{\cos \theta_o} \, d\omega_m \\ &= \frac{\int_\Omega G_1(\omega_o, \omega_m) \langle \omega_o, \omega_m \rangle D(\omega_m) \, d\omega_m}{\int_\Omega G_1(\omega_o, \omega_m) \langle \omega_o, \omega_m \rangle D(\omega_m) \, d\omega_m} \\ &= 1, \end{aligned} \tag{18}$$

and the average outgoing radiance from Equations (15) and (17) can thus be expressed in the same form as Equation (1), emphasizing the correct normalization:

$$L(\omega_o, \mathcal{M}) = \frac{\int_\Omega L(\omega_o, \omega_m) G_1(\omega_o, \omega_m) \langle \omega_o, \omega_m \rangle D(\omega_m) \, d\omega_m}{\int_\Omega G_1(\omega_o, \omega_m) \langle \omega_o, \omega_m \rangle D(\omega_m) \, d\omega_m}. \tag{19}$$

3.2. Construction of the BRDF

We now construct the BRDF upon the distribution of visible normals. The radiance $L(\omega_o, \omega_m)$ of each microfacet can be expressed in terms of the micro-BRDF $\rho_{\mathcal{M}}(\omega_o, \omega_i, \omega_m)$ associated with each microfacet and integrated with the incident radiance $L(\omega_i)$ over the domain of the incident directions Ω_i (we reserve Ω for the space of the normals):

$$L(\omega_o, \omega_m) = \int_{\Omega_i} \frac{dL(\omega_o, \omega_m)}{d\omega_i} \, d\omega_i = \int_{\Omega_i} \rho_{\mathcal{M}}(\omega_o, \omega_i, \omega_m) \langle \omega_i, \omega_m \rangle L(\omega_i) \, d\omega_i, \tag{20}$$

where the micro-BRDF $\rho_{\mathcal{M}}(\omega_o, \omega_i, \omega_m)$ is defined as the ratio of the differential outgoing radiance $dL(\omega_o, \omega_m)$ to the differential incoming irradiance $\langle \omega_i, \omega_m \rangle L(\omega_i) \, d\omega_i$:

$$\rho_{\mathcal{M}}(\omega_o, \omega_i, \omega_m) = \frac{dL(\omega_o, \omega_m)}{\langle \omega_i, \omega_m \rangle L(\omega_i) \, d\omega_i}. \tag{21}$$

Next, we differentiate Equation (17) with respect to the incoming irradiance and substitute $dL(\omega_o, \omega_m)$ via Equation (21):

$$\begin{aligned} dL(\omega_o, \mathcal{M}) &= \int_\Omega dL(\omega_o, \omega_m)\, D_{\omega_o}(\omega_m)\, d\omega_m \\ &= L(\omega_i)\, d\omega_i \int_\Omega \rho_{\mathcal{M}}(\omega_o, \omega_i, \omega_m)\, \langle \omega_i, \omega_m \rangle\, D_{\omega_o}(\omega_m)\, d\omega_m, \end{aligned} \tag{22}$$

where $L(\omega_i)\, d\omega_i$ can be moved outside of the integral because it does not depend on ω_m. Since the macro-BRDF is defined by the equation

$$dL(\omega_o, \mathcal{M}) = \rho(\omega_o, \omega_i) \cos\theta_i L(\omega_i)\, d\omega_i, \tag{23}$$

we arrive at the following:

$$\begin{aligned} \rho(\omega_o, \omega_i) &= \frac{dL(\omega_o, \mathcal{M})}{\cos\theta_i L(\omega_i)\, d\omega_i} \\ &= \frac{1}{\cos\theta_i} \int_\Omega \rho_{\mathcal{M}}(\omega_o, \omega_i, \omega_m)\, \langle \omega_i, \omega_m \rangle\, D_{\omega_o}(\omega_m)\, d\omega_m. \end{aligned} \tag{24}$$

By substituting $D_{\omega_o}(\omega_m)$ from Equation (16) we get

$$\begin{aligned} \rho(\omega_o, \omega_i) &= \frac{1}{\cos\theta_o \cos\theta_i} \int_\Omega \rho_{\mathcal{M}}(\omega_o, \omega_i, \omega_m)\, \langle \omega_o, \omega_m \rangle\, \langle \omega_i, \omega_m \rangle\, G_1(\omega_o, \omega_m)\, D(\omega_m)\, d\omega_m \\ &= \frac{1}{|\omega_g \cdot \omega_o|\, |\omega_g \cdot \omega_i|} \int_\Omega \rho_{\mathcal{M}}(\omega_o, \omega_i, \omega_m)\, \langle \omega_o, \omega_m \rangle\, \langle \omega_i, \omega_m \rangle\, G_1(\omega_o, \omega_m)\, D(\omega_m)\, d\omega_m. \end{aligned} \tag{25}$$

An important observation is that this equation is only modeling how rays are reflected just after the first bounce *before* leaving the vicinity of the surface (Figure 6(b)). However, a BRDF model must instead describe how rays are distributed *after* leaving the surface, following all microscattering. The distribution before and after leaving the vicinity of the surface is not the same, because some reflected rays hit the microsurface again and are reflected in another direction before leaving (Figure 6(d)). Since the BRDF model derived here only accounts for the first bounce on the surface, rays involving multiple bounces (shown in black in Figure 6(c)) have to be removed from the model, which is achieved by introducing a *shadowing* function. In practice, we replace the masking function G_1 by a masking-shadowing function G_2:

$$\begin{aligned} &\rho(\omega_o, \omega_i) \\ &= \frac{1}{|\omega_g \cdot \omega_o|\, |\omega_g \cdot \omega_i|} \int_\Omega \rho_{\mathcal{M}}(\omega_o, \omega_i, \omega_m)\, \langle \omega_o, \omega_m \rangle\, \langle \omega_i, \omega_m \rangle\, G_2(\omega_o, \omega_i, \omega_m)\, D(\omega_m)\, d\omega_m. \end{aligned} \tag{26}$$

Next, we will instantiate this equation for the specific cases where the microfacets are perfect mirrors (3.3) or perfect Lambertian diffusers (3.4).

3.3. Construction of the BRDF with specular microfacets

The micro-BRDF for mirror-like microfacets is

$$\begin{aligned}\rho_{\mathcal{M}}(\omega_o,\omega_i,\omega_m) &= \left\|\frac{\partial\omega_h}{\partial\omega_i}\right\| \frac{F(\omega_o,\omega_h)\,\delta_{\omega_h}(\omega_m)}{|\omega_i\cdot\omega_h|} \\ &= \frac{F(\omega_o,\omega_h)\,\delta_{\omega_h}(\omega_m)}{4\,|\omega_i\cdot\omega_h|^2},\end{aligned} \tag{27}$$

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 [Walter et al. 2007], and F is the Fresnel term. By substituting $\rho_{\mathcal{M}}(\omega_o,\omega_i,\omega_m)$ from Equation (27) and $D_{\omega_o}(\omega_m)$ from Equation (16) into Equation (26), we arrive at

$$\rho(\omega_o,\omega_i) = \frac{1}{|\omega_g\cdot\omega_o|\,|\omega_g\cdot\omega_i|}\int_\Omega \frac{F(\omega_o,\omega_h)\,\delta_{\omega_h}(\omega_m)}{4\,|\omega_i\cdot\omega_h|^2}\,\langle\omega_o,\omega_m\rangle\,\langle\omega_i,\omega_m\rangle\,G_2(\omega_o,\omega_i,\omega_m)\,D(\omega_m)\,d\omega_m. \tag{28}$$

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

$$\rho(\omega_o,\omega_i) = \frac{F(\omega_o,\omega_h)\,G_2(\omega_o,\omega_i,\omega_h)\,D(\omega_h)}{4\,|\omega_g\cdot\omega_o|\,|\omega_g\cdot\omega_i|}. \tag{29}$$

We have arrived at the well-known equation for specular microfacet-based BRDFs [Walter et al. 2007].

3.4. Construction of the BRDF with diffuse microfacets

The micro-BRDF for diffuse microfacets is constant:

$$\rho_{\mathcal{M}}(\omega_o,\omega_i,\omega_m) = \frac{1}{\pi}. \tag{30}$$

In Equation (26), by substituting $\rho_{\mathcal{M}}(\omega_o,\omega_i,\omega_m)$ from Equation (30) and $D_{\omega_o}(\omega_m)$ from Equation (16) we obtain

$$\rho(\omega_o,\omega_i) = \frac{1}{\pi}\frac{1}{|\omega_g\cdot\omega_o|\,|\omega_g\cdot\omega_i|}\int_\Omega \langle\omega_o,\omega_m\rangle\,\langle\omega_i,\omega_m\rangle\,G_2(\omega_o,\omega_i,\omega_m)\,D(\omega_m)\,d\omega_m. \tag{31}$$

This equation has no analytical solution. Oren and Nayar [Oren and Nayar 1994] propose an analytical fit of this function in the case where D is a spherical Gaussian—not to be confused with the Beckmann distribution—and where G_2 is the V-cavity masking-shadowing function.

3.5. 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:

$$s(\omega_o, \omega_i) = \rho(\omega_o, \omega_i) + t(\omega_o, \omega_i). \tag{32}$$

If we had a surface that absorbed no incident radiance, then the radiance of the rays would be perfectly preserved during scattering. 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:

$$\int_{\Omega_i} s(\omega_o, \omega_i) \, |\omega_g \cdot \omega_i| \, d\omega_i = 1 \; \forall \omega_o. \tag{33}$$

If the Fresnel term is always 1, then rays are never transmitted (they never penetrate the surface), so 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:

$$\int_{\Omega_i} \rho(\omega_o, \omega_i) \, |\omega_g \cdot \omega_i| \, d\omega_i = 1. \tag{34}$$

Intuitively, this represents the fact that rays cast from the outgoing direction (Figure 6(a)) would be scattered one or more times and eventually leave the surface (Figure 6(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 Figure 6(c) and described in Section 3.2. This is why common BRDF models do not integrate to 1 even when parameterized on a "perfect reflector" microsurface, 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 (Figure 6(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 (29) becomes

$$\rho(\omega_o, \omega_i) = \frac{G_1(\omega_o, \omega_h) \, D(\omega_h)}{4 \, |\omega_g \cdot \omega_o| \, |\omega_g \cdot \omega_i|}, \tag{35}$$

and, after substituting the *White Furnace Test* into Equation (34), the $|\omega_g \cdot \omega_i|$ terms cancel and the *Weak White Furnace Test* equation is given by

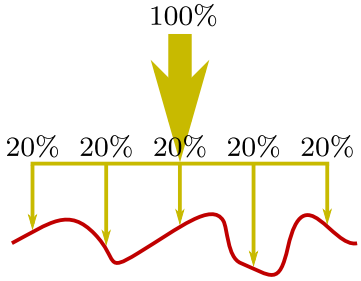
$$\boxed{\int_{\Omega_i} \frac{G_1(\omega_o, \omega_h) D(\omega_h)}{4 |\omega_g \cdot \omega_o|} d\omega_i = 1.} \tag{36}$$

This condition is only met with an appropriate masking function G_1 that satisfies Equation (14). In Appendix C, we provide MATLAB code to numerically compute Equation (36) with Beckmann and GGX distributions and their associated Smith masking functions.

We can define the same test for BRDFs with diffuse microfacets, by substituting $G_2(\omega_o, \omega_i, \omega_h) = G_1(\omega_o, \omega_h)$ into Equation (31) and by integrating over the incident directions:

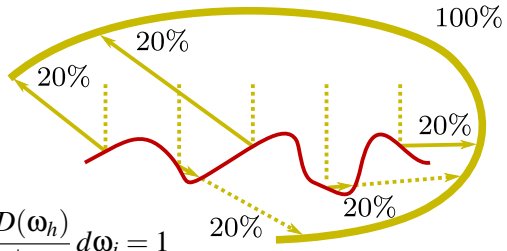
$$\begin{aligned} &\int_{\Omega_i} \frac{1}{\pi} \frac{1}{|\omega_g \cdot \omega_o| \, |\omega_g \cdot \omega_i|} \int_{\Omega} \langle \omega_o, \omega_m \rangle \langle \omega_i, \omega_m \rangle G_1(\omega_o, \omega_m) D(\omega_m) \, d\omega_m \, |\omega_g \cdot \omega_i| \, d\omega_i \\ &= \frac{1}{\pi} \frac{1}{|\omega_g \cdot \omega_o|} \int_{\Omega_i} \int_{\Omega} \langle \omega_o, \omega_m \rangle \langle \omega_i, \omega_m \rangle G_1(\omega_o, \omega_m) D(\omega_m) \, d\omega_m \, d\omega_i = 1. \end{aligned} \tag{37}$$

(a) Casting rays onto the microsurface



(b) Without Fresnel and shadowing, the rays are always reflected somewhere and their energy does not change.

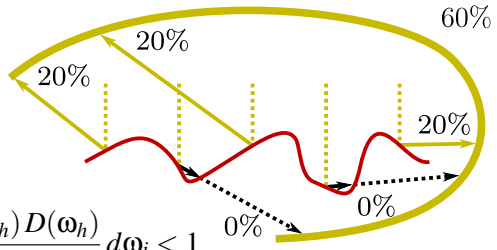
⇒ The distribution of reflected rays is normalized.



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

(c) Shadowing in the BRDF models discards rays occluded by the microsurface.

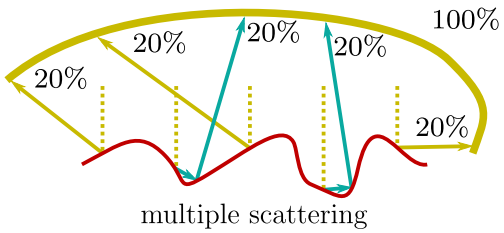
⇒ The distribution of reflected rays is NOT normalized.



$$\int_{\Omega_i} \rho(\omega_o,\omega_i)\,|\omega_g\cdot\omega_i|\,d\omega_i = \int_{\Omega_i} \frac{G_2(\omega_o,\omega_i,\omega_h)\,D(\omega_h)}{4\,|\omega_g\cdot\omega_o|}\,d\omega_i < 1$$

(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 therefore be normalized.



$$\int_{\Omega_i} \rho(\omega_o,\omega_i)\,|\omega_g\cdot\omega_i|\,d\omega_i = 1$$

Figure 6. Normalization of specular microfacet-based BRDFs.

3.6. 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 have 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 be normalized, i.e., integrate to exactly 1 for a non-absorbing, non-transmissive material.
- The shadowing function in microfacet-based BRDFs is used to separate the first scattering event from multiple scattering events on the microsurface. Shadowing discards (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. Physically based masking functions always satisfy Equations (36) and (37) . This is what we call the "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 physically valid. 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.

4. Common Physically Based and Non-Physically Based Masking Functions

In Sections 2 and 3 we derived general results for the masking function—Equations (14), (18) and (36)—without making any assumption about the type of microsurface. In this section, we review the Smith (4.1) and V-cavity (4.2) microsurface profiles, derive the closed form of their respective masking functions, and discuss their properties. We also review other common masking functions that have no associated microsurface profile and thus are not physically based (4.3).

4.1. The Smith Microsurface Profile

Normal/Masking Independence The Smith microsurface profile assumes that the microsurface is not autocorrelated, i.e., that there is no correlation between the height (or the normal) at one point of the microsurface and the height (or the normal) at any neighboring point, even the closest ones. This implies a random set of microfacets rather than a continuous surface—as shown in Figure 7(right), where the heights and the normals of the microsurface are independent random variables.

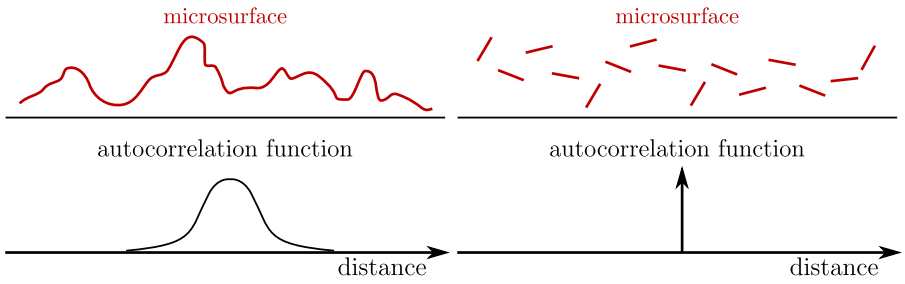


Figure 7. Microsurfaces and their autocorrelation functions. (left) A real-world continuous microsurface with large autocorrelation distance. (right) A uncorrelated surface, where each microfacet is not correlated to its neighborhood, as Smith models.

The consequence of this model is that the probability $G_1(\omega_o, \omega_m)$ given by the masking function is independent of the normal orientation ω_m for normals that are not backfacing ($\omega_o \cdot \omega_m > 0$). The intuition is that the normal ω_m is a *local* property of the microfacet, while the potential occlusion responsible for masking occurs elsewhere on the microsurface and is thus a *distant* property of the microfacet (albeit, where the distance is still on the micro-scale). Since the microsurface is not autocorrelated, local properties are independent of distant properties and the masking function can be expressed in the separable form

$$G_1(\omega_o, \omega_m) = G_1^{\text{local}}(\omega_o, \omega_m)\, G_1^{\text{dist}}(\omega_o), \tag{38}$$

where the local masking function is the binary discard of backfacing microfacets:

$$G_1^{\text{local}}(\omega_o, \omega_m) = \chi^+(\omega_o \cdot \omega_m), \tag{39}$$

and the distant masking function $G_1^{\text{dist}}(\omega_o)$ is the probability of occlusion by a distant point of the microsurface, which is independent of the local orientation ω_m.

Derivation of the Masking Function By expanding the masking function from Equation (38) into Equation (14), we get

$$\begin{aligned}\cos\theta_o &= \int_\Omega G_1(\omega_o,\omega_m)\,\langle\omega_o,\omega_m\rangle\,D(\omega_m)\,d\omega_m\\ &= \int_\Omega G_1^{\text{local}}(\omega_o,\omega_m)\,G_1^{\text{dist}}(\omega_o)\,\langle\omega_o,\omega_m\rangle\,D(\omega_m)\,d\omega_m\\ &= \int_\Omega \chi^+(\omega_o\cdot\omega_m)\,G_1^{\text{dist}}(\omega_o)\,\langle\omega_o,\omega_m\rangle\,D(\omega_m)\,d\omega_m\\ &= G_1^{\text{dist}}(\omega_o)\int_\Omega \langle\omega_o,\omega_m\rangle\,D(\omega_m)\,d\omega_m,\end{aligned} \tag{40}$$

where $G_1^{\text{dist}}(\omega_o)$ can be taken out of the integral since it does not depend on ω_m, and we remove the step function $\chi^+(\omega_o\cdot\omega_m)$ since it is redundant due to the clamped dot product that already evaluates to 0 when $\omega_o\cdot\omega_m<0$. The masking function for the non-backfacing normals from outgoing direction ω_o is thus

$$G_1^{\text{dist}}(\omega_o) = \frac{\cos\theta_o}{\int_\Omega \langle\omega_o,\omega_m\rangle\,D(\omega_m)\,d\omega_m}, \tag{41}$$

and so the complete masking function is therefore

$$\boxed{G_1(\omega_o,\omega_m) = \chi^+(\omega_o\cdot\omega_m)\,\frac{\cos\theta_o}{\int_\Omega \langle\omega_o,\omega_m\rangle\,D(\omega_m)\,d\omega_m}.} \tag{42}$$

This is the integral form of the exact masking function under normal/masking independence by Ashikhmin et al. [2000]. They use this integral expression to precompute the masking function, which they store in a table for efficient evaluation at runtime.

The Smith Masking Function In the literature, the Smith masking function is often expressed as a fraction $\frac{1}{1+\Lambda(\omega_o)}$ involving a function $\Lambda(\omega_o)$. This function is expressed as an integral over the slopes of the microsurface and its form is derived with a ray-tracing formulation of the masking probability [Walter et al. 2007]. The drawback of this formulation is that it does not emphasize the exactness of the result. This is why the Smith masking function is often considered to be approximate.

We show that the derivation of Ashikhmin et al. leads to the same result and has the advantage of emphasizing its exactness. Indeed, by changing the integration domain from normal to slope space—we provide the detailed derivation in Appendix A—Equation (41) becomes

$$G_1^{\text{dist}}(\omega_o,\omega_m) = \frac{1}{1+\Lambda(\omega_o)},$$

and thus Equation (42) can be rewritten as

$$\boxed{G_1(\omega_o, \omega_m) = \frac{\chi^+(\omega_o \cdot \omega_m)}{1+\Lambda(\omega_o)},} \tag{43}$$

where $\frac{1}{1+\Lambda(\omega_o)}$ is the generalized form of the Smith masking function [Brown 1980; Walter et al. 2007], for which closed-form solutions are available for many stochastic surfaces, as shown in Section 5. Therefore, under the assumption of normal/masking independence, the Smith masking function is exact.

Properties Still, if we were to compare the analytical function with measured data, we would find that the predictions of the model are accurate but not exact. Indeed, Smith compared his formula to real-world measurements and discovered that it was a good fit, but an approximation nonetheless. However, the approximation does not reside in his derivation, because his formula is exact within the framework of his model. 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.

The uncorrelated microsurface assumed by the Smith model is reminiscent of "metal flakes", which can be found in some metallic car paints [Rump et al. 2008], but real-world continuous surfaces have wider autocorrelation functions. Bourlier et al. [2000] compared the Smith masking function to the numerically measured masking function on random rough surfaces with different autocorrelation functions (Gaussian and Lorentzian). The conclusion of their investigation was that the error introduced by neglecting correlation on random surfaces is noticeable only at observation angles such as $\tan(\theta)/\alpha > 0.5$, where $\sigma^2 = \frac{\alpha^2}{2}$ is the slope variance. The Smith masking function tends to produce slight overestimations in this case. Given that the Smith masking function is generally accurate even on correlated surfaces, and given that there is no analytical solution to the correlated masking function, it seems reasonable to apply the Smith masking function in a computer graphics context. However, as pointed out by Ashikhmin et al. [2000], 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.

The Smith Averaged Masking Functions Smith derived the masking function averaged over different quantities of the microsurface, such as the heights and the normals [Smith 1967]. The masking function $G_1(\omega_o, \omega_m)$ presented in Equation (43) 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 the normals involved in the BRDF, we can just average over them. However, the normals are not all processed in the same way, since backfacing normals are not considered. In this BRDF model, only what is visible to the external viewer matters, because only the radiance that can be measured by this viewer matters. If something exists on the surface but is not vis-

ible, then it won't be included in the BRDF. Smith derived a normal-averaged form of his masking function, which does not directly address this. He wanted to answer the question: "*What proportion of normals are masked?*", which is important when studying properties intrinsic to the surface in other situations, such as wave optics models. It is not important for geometric optics-based BRDF models, which are the focus of this article.

In a geometrical microfacet-based BRDF problem, we are actually interested in the slightly different question: "*What proportion of* ***non-backfacing*** *normals are masked?*"

4.2. The V-Cavity Microsurface Profile

In this section, we discuss the masking model based on V-cavities [Cook and Torrance 1982; Oren and Nayar 1994], which is the most common alternative to the Smith masking function. Figure 8 illustrates the scattering model with V-cavity microsurfaces. Rather than modeling the scattering on one microsurface with a distribution of normals, this model computes the scattering on separate microsurfaces and averages their contributions. Each microsurface is composed of two normals $\omega_m = (x_m, y_m, z_m)$ and $\omega_m{}' = (-x_m, -y_m, z_m)$ and the contribution of each microsurface is weighted by $\langle \omega_m, \omega_g \rangle D(\omega_m)$ in the final BRDF.

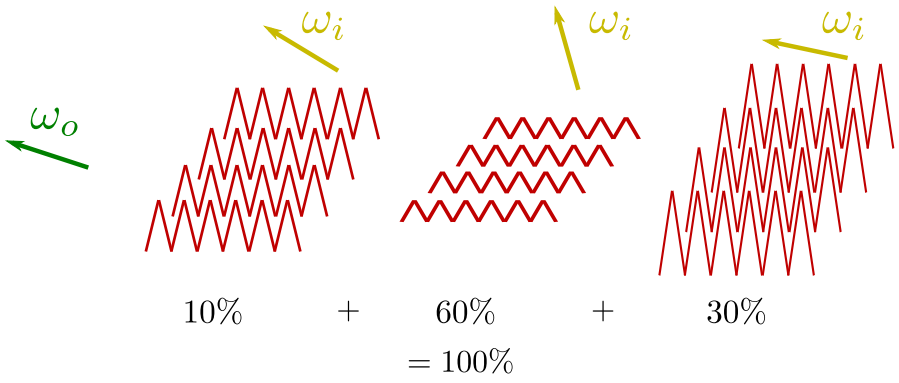


Figure 8. The V-cavity scattering model. Instead of modeling the scattering on one microsurface, the model computes the scattering on separate microsurfaces and blends the results.

A trigonometric derivation is usually presented to arrive at the masking function of a V-cavity microsurface. We can derive the same result simply from the property that the projected area of the visible microsurface is conserved, as discussed in Section 2.3. V-cavity microsurfaces have only the two symmetric normals, ω_m and $\omega_m{}'$. The distribution of normals of this microsurface is therefore

$$D(\omega) = \frac{1}{2}\frac{\delta_{\omega_m}(\omega)}{\omega_m \cdot \omega_g} + \frac{1}{2}\frac{\delta_{\omega_m{}'}(\omega)}{\omega_m{}' \cdot \omega_g}, \tag{44}$$

and we can verify that the normalization is correct:

$$\begin{aligned}\int_\Omega \langle \omega, \omega_g\rangle D(\omega) d\omega &= \int_\Omega \langle \omega, \omega_g\rangle \left(\frac{1}{2}\frac{\delta_{\omega_m}(\omega)}{\omega_m \cdot \omega_g} + \frac{1}{2}\frac{\delta_{\omega_m{}'}(\omega)}{\omega_m{}' \cdot \omega_g}\right) d\omega \\ &= \frac{1}{2}\frac{\omega_m{}' \cdot \omega_g}{\omega_m{}' \cdot \omega_g} + \frac{1}{2}\frac{\omega_m{}' \cdot \omega_g}{\omega_m{}' \cdot \omega_g} \\ &= \frac{1}{2} + \frac{1}{2} = 1. \end{aligned} \tag{45}$$

To derive the masking term, we use the conservation of the visible projected area presented in Equation (14):

$$\begin{aligned}\cos\theta_o &= \int_\Omega G_1(\omega_o, \omega)\langle \omega_o, \omega\rangle D(\omega) d\omega \\ &= \frac{1}{2} G_1(\omega_o, \omega_m)\frac{\langle \omega_o, \omega_m\rangle}{\omega_m \cdot \omega_g} + \frac{1}{2} G_1(\omega_o, \omega_m{}')\frac{\langle \omega_o, \omega_m{}'\rangle}{\omega_m{}' \cdot \omega_g}. \end{aligned} \tag{46}$$

There are two possible configurations, as shown in Figure 9. In the first case, the two normals are visible and there is no masking ($G_1(\omega_o, \omega_m) = 1$ and $G_1(\omega_o, \omega_m{}') = 1$). Otherwise, $\omega_m{}'$ is backfacing ($G_1(\omega_o, \omega_m{}') = 0$) and we have

$$\cos\theta_o = \frac{1}{2} G_1(\omega_o, \omega_m)\frac{\langle \omega_o, \omega_m\rangle}{\omega_m \cdot \omega_g}, \tag{47}$$

whose solution is

$$\begin{aligned} G_1(\omega_o, \omega_m) &= 2\frac{\cos\theta_o\,(\omega_m \cdot \omega_g)}{\langle \omega_o, \omega_m\rangle} \\ &= 2\frac{(\omega_m \cdot \omega_g)(\omega_o \cdot \omega_g)}{\langle \omega_o, \omega_m\rangle}. \end{aligned} \tag{48}$$

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

$$G_1(\omega_o, \omega_m) = \min\left(1, 2\frac{(\omega_m \cdot \omega_g)(\omega_o \cdot \omega_g)}{\langle \omega_o, \omega_m\rangle}\right), \tag{49}$$

which is the well-known V-cavity masking function used by Cook and Torrance [1982].

Validation We will now verify that this model satisfies Equation (18), i.e., that the distribution of visible normals is normalized. First, we will substitute G_1 (from Equation 49) into Equation (16):

$$D_{\omega_o}(\omega_m) = \min\left(1, 2\frac{(\omega_m \cdot \omega_g)(\omega_o \cdot \omega_g)}{\langle \omega_o, \omega_m\rangle}\right)\frac{\langle \omega_o, \omega_m\rangle D(\omega_m)}{\cos\theta_o}. \tag{50}$$

This form is complicated to study because of the $\min(1, -)$ term. However, the main difference between the V-cavity model and the Smith model occurs at grazing angles, where $\theta_o \approx \frac{\pi}{2}$. Indeed, at grazing angles, we are always in configuration (a) from

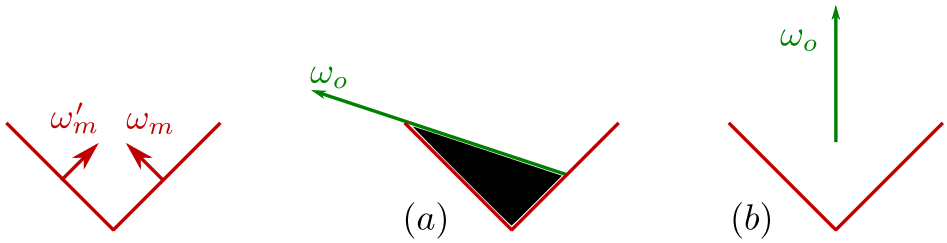


Figure 9. Masking on a V-cavity microsurface. Either one of the two normal is backfacing and the other is partially masked (a) or the two normals are visible and the masking function evaluates to 1.

Figure 9, where one of the two normals is backfacing. In this case, we can drop the $\min(1,-)$:

$$\begin{aligned} D_{\omega_o}(\omega_m) &= 2\,\frac{(\omega_m\cdot\omega_g)(\omega_o\cdot\omega_g)}{\langle\omega_o,\omega_m\rangle}\,\frac{\langle\omega_o,\omega_m\rangle\,D(\omega_m)}{\cos\theta_o} \\ &= 2\,\chi^+(\omega_o\cdot\omega_m)\,(\omega_m\cdot\omega_g)\,D(\omega_m)\,d\omega_m. \end{aligned} \tag{51}$$

Note that the clamped dot product $\langle\omega_o,\omega_m\rangle$ cancels out, but the heaviside $\chi^+(\omega_o\cdot\omega_m)$ is left to ensure that backfacing normals are still removed from the distribution. We now validate this result by verifying Equation (18)—i.e., the distribution of visible normals is normalized—by computing

$$\int_\Omega D_{\omega_o}(\omega_m)\,d\omega_m = 2\int_\Omega \chi^+(\omega_o\cdot\omega_m)\,(\omega_m\cdot\omega_g)\,D(\omega_m)d\omega_m. \tag{52}$$

Since the outgoing direction is almost orthogonal to the geometric normal at a grazing angle, 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. $D(\omega_m) = D(\omega_m{}')$. This means that the heaviside function cuts the distribution of normals into two equal parts, yielding

$$\begin{aligned} \int_\Omega \chi^+(\omega_o\cdot\omega_m)\,(\omega_m\cdot\omega_g)\,D(\omega_m)\,d\omega_m &= \frac{1}{2}\int_\Omega (\omega_m\cdot\omega_g)\,D(\omega_m)\;d\omega_m \\ &= \frac{1}{2}, \end{aligned} \tag{53}$$

(recall that $\int_\Omega(\omega_m\cdot\omega_g)\,D(\omega_m)\,d\omega_m = 1$, i.e. the distribution of normals is normalized). Using this result in Equation (52) yields

$$\begin{aligned} \int_\Omega D_{\omega_o}(\omega_m)\,d\omega_m &= 2\int_\Omega \chi^+(\omega_o\cdot\omega_m)\,(\omega_m\cdot\omega_g)\,D(\omega_m)\,d\omega_m \\ &= 2\,\frac{1}{2} = 1. \end{aligned} \tag{54}$$

This shows that the distribution of visible normals from Equation (51) is normalized for grazing angles of incidence. A more technical derivation can show that the distribution is normalized for any angle of incidence.

Another way to validate this model is to use the Weak White Furnace Test. To do this, we evaluate Equation (36) with G_1 from Equation (49). In practice, this evaluation can be performed via numerical integration. The result of the integral is always 1, thus the model with V-cavities is mathematically well designed and energy conserving.

Properties 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 angles of incidence.

There are two kinds of normals: those that are backfacing, which are removed by the heaviside term; and those that are not backfacing, which yield a radiance contribution weighted by $(\omega_m \cdot \omega_g) D(\omega_m)$. Note that the factor $(\omega_m \cdot \omega_g)$ is the Jacobian of the projection of a microfacet onto the geometric surface, as shown in Figure 3(a). Thus, the microfacets are weighted exactly as if they were projected onto the geometric surface before being projected onto the outgoing direction. As a result, we are simulating a geometrically flat microsurface: the microfacets can perturb the reflection of light, but they do not exist geometrically. This microsurface model is therefore unrealistic because it behaves more like a normal map than a displacement map, as shown in Figure 10.

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

The more the angle of incidence is grazing the microsurface, the more the microsurface profile tends to exhibit this normal map behavior. The consequence is that the peak of the BRDF tends to be too low: on a real microsurface, normals oriented toward the outgoing direction have a higher contribution to the BRDF because their projected area is greater. Because of this, the reflected directions tend to be shifted toward the outgoing direction, as shown in Figure 10. This shifting effect is not present with normal maps because the microfacets have no geometrical existence: they all have the same projected area.

Figure 11 shows the BRDFs produced by an isotropic Beckmann distribution with the V-cavity and Smith masking-shadowing functions, and results computed from a numerical simulation of the scattering of a procedural stochastic microsurface model with matching Gaussian statistics. We see that, with the Smith masking function, the

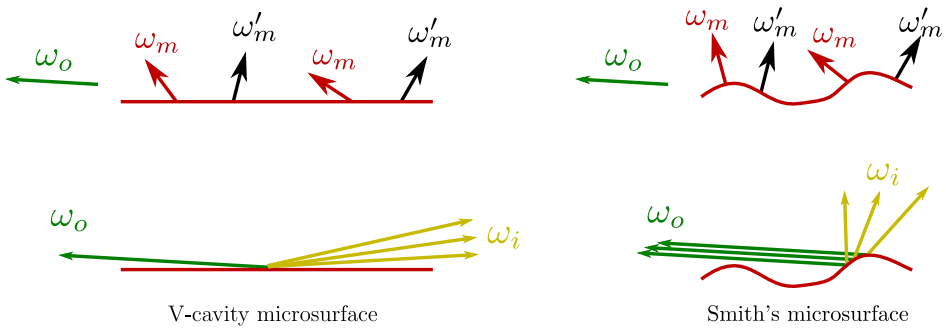


Figure 10. (top) V-cavity surfaces exhibit a normal map behavior at grazing angles: non-backfacing normals have the same visibility as the geometric surface, as if they had no geometrical existence. (bottom) With the same grazing incidence ω_o, the directions ω_i reflected by the V-cavity microsurface are too low on average compared to the directions ω_i reflected by physical surfaces.

distribution is shifted toward the outgoing direction as the roughness increases. For very high roughness values, the BRDF is even mainly backscattering. This effect—present in the measured data—is to be expected, because the normals oriented toward the outgoing direction are the most visible. In contrast, this effect does not emerge from the V-cavity model.

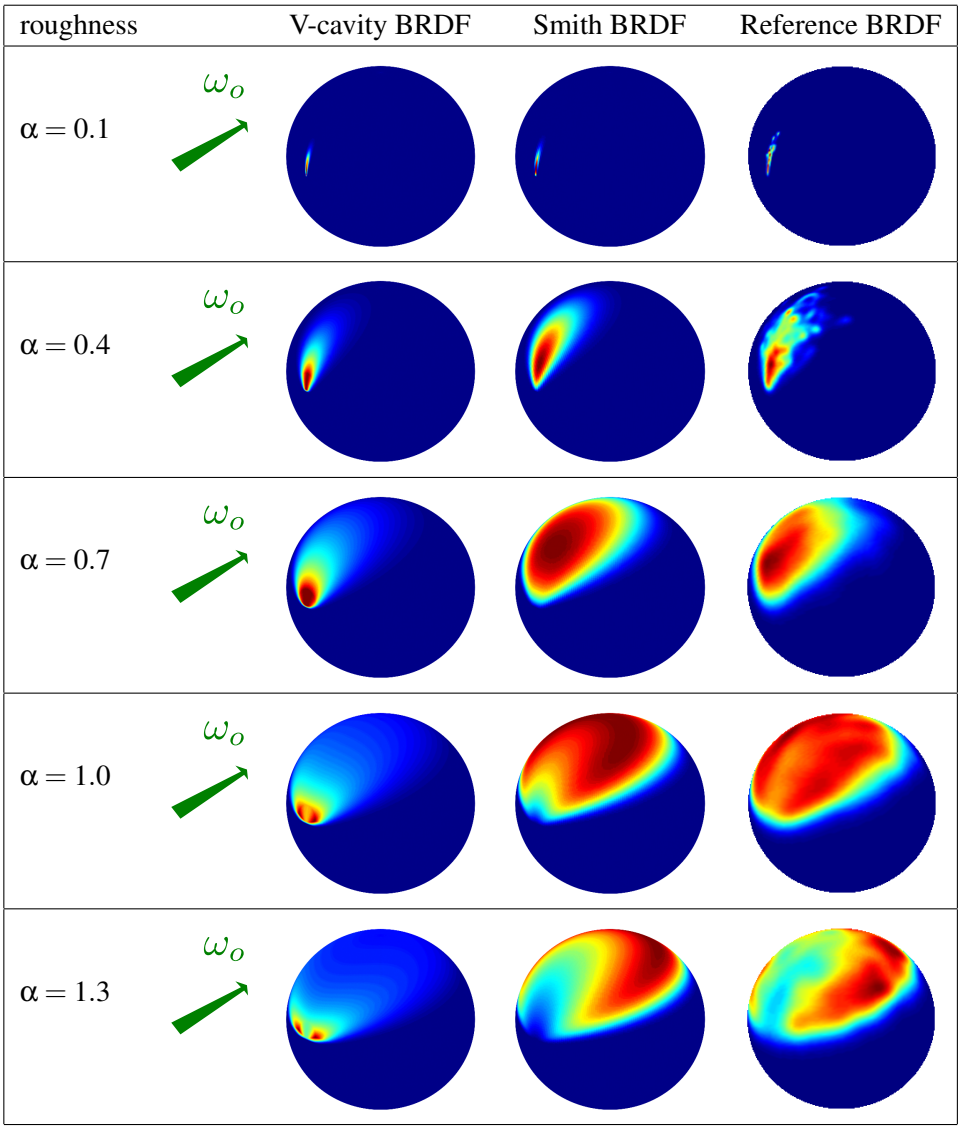


Figure 11. (left and middle) The BRDFs associated with an isotropic Beckmann distribution and different masking functions at grazing incidence ($\theta_o = 1.5$). (right) Reference computed with Monte Carlo raytracing on a procedural surface with Gaussian statistics parametrized by the roughness parameter α.

4.3. Non-Physically Based Masking Functions

Definition Recall that we classify a masking function as "physically based" when it has been derived from a microsurface model or measured on a physical microsurface. A physically based masking function always satisfies Equations (14), (18), (36) and (37). Reciprocally, a "non-physically based" masking function does not satisfy those equations, i.e., there is no possible microsurface from which it can be derived (or measured).

The Implicit Masking Function In order to simplify the expression of the specular microfacet-based BRDF given in Equation (29), many models remove the masking-shadowing function and the denominator expression $|\omega_g \cdot \omega_o|\,|\omega_g \cdot \omega_i|$ assuming that they cancel out [McAuley et al. 2013]. The masking-shadowing is thus implicitly defined by

$$G_2(\omega_o, \omega_i, \omega_m) = G_1(\omega_o, \omega_m)\, G_1(\omega_i, \omega_m), \tag{55}$$

where the associated separate masking and shadowing functions are

$$G_1(\omega_o, \omega_m) = \chi^+(\omega_o \cdot \omega_m)\, \langle \omega_o, \omega_g \rangle, \tag{56}$$

$$G_1(\omega_i, \omega_m) = \chi^+(\omega_i \cdot \omega_m)\, \langle \omega_i, \omega_g \rangle. \tag{57}$$

This masking function is "plausible" since $G_1(\omega_o, \omega_m) = 1$ when $\omega_o = \omega_g$ and decreases to $G_1(\omega_o, \omega_m) = 0$ as the incidence angle goes to $\frac{\pi}{2}$. However, this masking function is not physically based. Indeed, it does not satisfy the conservation of the projected area from Equation (14), which implies that there is no physical microsurface model from which this masking function can be derived.

The Schlick-Smith Masking Function Schlick [1994] proposed an approximation to the Smith masking function, which is often referred to as the "Schlick-Smith" masking function. This masking function has three problems.

First, the roughness parameter m used across this paper is inconsistent. Equation (18) of the paper refers to the root mean square (RMS) slope of the microsurface, i.e. $m = \sigma$, the roughness descriptor often used in the physics literature and especially in the original Smith [1967] paper. This is consistent with the definition of variable h in the same equation. However, within Equation (20) of the paper, m refers to the roughness parameter from the Beckmann distribution $m = \alpha = \sqrt{2}\sigma$, which is the RMS slope scaled by a factor of $\sqrt{2}$. As a result, there is a mismatch between the roughness used in the masking function and the roughness used in the distribution of normals.

Whilst this inconsistency is easily solved, a more critical issue is that his reformulation of the Smith masking function (Equation (18) in the paper) is wrong. He states "After several equivalences, the original expression of G can be written...". Whilst

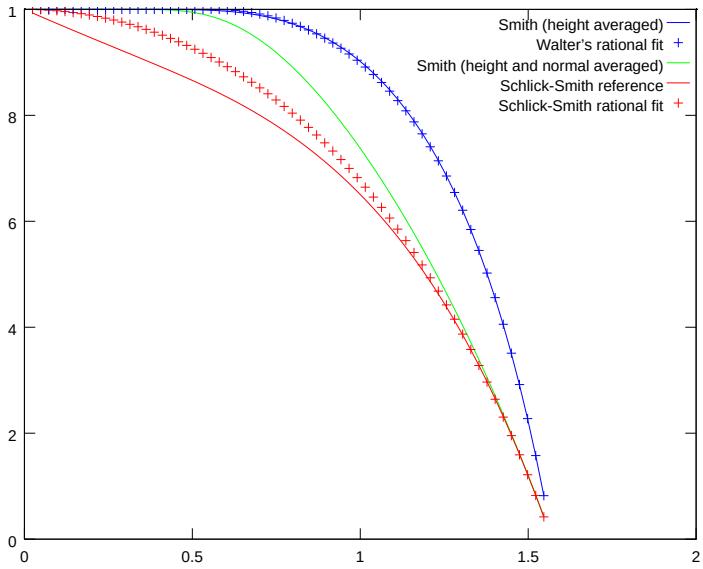


Figure 12. The Smith and the Schlick-Smith masking functions and their associated rational approximations for varying angle θ_o and fixed $\alpha = 1$. The Schlick-Smith reference and the associated rational approximation do not match the original Smith function.

those derivations are not provided, it appears that Schlick rearranged the equations provided by He et al. [1991] (Equations (24) and (25) of that paper), which contain a typo: the exponential term from the Smith Λ function is missing.

Furthermore, at the end of Section 4.1 (paragraph "The Smith Averaged Masking Functions"), we explained that the Smith masking function to use in this geometric optics BRDF model is the one that is averaged over the heights of the microsurface. However, Schlick followed He et al.'s use of the height-and-normal-averaged version of the Smith masking function, which was appropriate in the context of their *wave optics* model.

As a consequence of these two errors, neither Schlick's original formula nor his fitted approximation match the correct Smith masking function, as shown by the plot in Figure 12. This means that the "Schlick-Smith" masking function is not physically based, since it does not ensure conservation of the projected area from Equation (14).

The Kelemen Masking Function Kelemen et al. [2001] proposed a cheap way to replace the masking-shadowing function of the V-cavity model and the denominator of the specular microfacet-based BRDF given in Equation (29) by approximating

$$\frac{G_2(\omega_o, \omega_i, \omega_h)}{|\omega_g \cdot \omega_o|\,|\omega_g \cdot \omega_i|} = \frac{1}{|\omega_o \cdot \omega_h|\,|\omega_i \cdot \omega_h|}, \tag{58}$$

which is equivalent to approximating the masking and the shadowing functions by

$$G_1(\omega_o, \omega_h) = \frac{|\omega_g \cdot \omega_o|}{|\omega_o \cdot \omega_h|}, \tag{59}$$

$$G_1(\omega_i, \omega_h) = \frac{|\omega_g \cdot \omega_i|}{|\omega_i \cdot \omega_h|}. \tag{60}$$

This masking function is close to the V-cavity masking function: the multiplication by $|\omega_g \cdot \omega_o|$ is the projected area of the geometric surface and the division by $|\omega_o \cdot \omega_h|$ removes the projected area of the microfacet, i.e. the projected area of the microfacet is replaced by the projected area of the geometric surface, which tends to simulate a microsurface with flat microfacets like a normal map.

However, unlike the V-cavity masking function, the Kelemen masking function does not ensure conservation of the projected area from Equation (14). While it remains a good approximation to the V-cavity masking function, there is no physical model from which Kelemen's equation can be derived and it is thus not physically based.

4.4. Summary

In this section, we have shown that:

- The Smith and V-cavity masking functions are both physically based, but assume different microsurface profiles.
- Both ensure conservation of the projected area from Equation (14).
- Both are the normalization coefficient of the distribution of visible normals, Equation (18).
- Both satisfy the Weak White Furnace Test given by Equations (36) and (37).
- In contrast, the implicit masking function, the Schlick-Smith masking function, and the Kelemen masking function do not satisfy Equations (14), (18), (36) and (37). They are thus not physically based, i.e. there are no microsurface profiles from which their equations can be derived.

Masking function	Equations (14), (18), (36) and (37)	Physically based
Smith	✓	✓
V-cavity	✓	✓
Implicit	✗	✗
Schlick-Smith	✗	✗
Kelemen	✗	✗

Table 3. Properties of the different masking functions.

The Smith Masking Function A typical belief is that: "*The Smith masking function is a good approximation, because it depends on the distribution of normals.*"
We have shown that this answer is correct but the reason invoked is wrong, by developing the following ideas:

- Choosing the Smith model implies making the choice of assuming a microsurface on which 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 the Smith masking function.

The point here is that the reason why one would choose the Smith masking function is not because it is a physically plausible approximation parametrized by

the distribution of normals. The real reason to choose it is that Smith's formula is the exact masking function under the assumption of the chosen microsurface profile (i.e. normal/masking independence). The fact that it is physically plausible and is parametrized by the distribution of normals are not directly the reasons to choose it, but are some of the expected side effects of using a correct physically based masking function.

The V-Cavity Masking Function Another misconception is that: "*V-cavity masking function wrong because it does not depend on the distribution of normals.*"
In this section, we showed that this is reasoning is faulty, but that there are important limitations to the approximation yielded by the V-cavity model because:

- The V-cavity masking function per-normal does not depend on the distribution of normals.
- However, the average of the V-cavity masking function does depend on the distribution of normals. This is because the masking function and the normals are not assumed to be independent, contrary to the Smith model. The more the surface roughness increases, the more the average masking of the BRDF increases.
- The V-cavity masking function can be used with any kind of symmetric distribution of normals and guarantees correct normalization.
- However, the surface profile assumed by the V-cavity model has a response close to that of a normal map, with flat microfacets at grazing angles of incidence. This makes it physically less realistic than the Smith model.
- The consequence is that at grazing angles and with high roughness, the BRDF lobe is too low compared to what is expected from a realistic material.

There is no definitive answer to the question of whether to choose V-cavities or Smith, as both are based on a microsurface profile and are mathematically well defined. V-cavities are less computationally expensive and are generic—they work, mathematically, with any distribution of normals—but are less realistic. In contrast, the Smith-based model is more accurate, but requires specific derivations and sometimes-expensive evaluations. The choice is thus a matter of tradeoff between realism and performance.

5. Stretch Invariance of the Masking Function

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

5.1. Masking Probability Invariance

Figure 13 shows the effect that stretching a 1D configuration has on the masking of a microsurface, for a given outgoing 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 unoccluded rays are still unoccluded. This is a key property: the masking probability is invariant to configuration stretching when all of the slopes involved in the configuration are scaled at the same time. This includes the slopes of the microsurface and the slope associated with the outgoing direction. They are all scaled by the inverse of the stretching factor. The distribution of slopes width is thus also stretched by the inverse stretching factor.

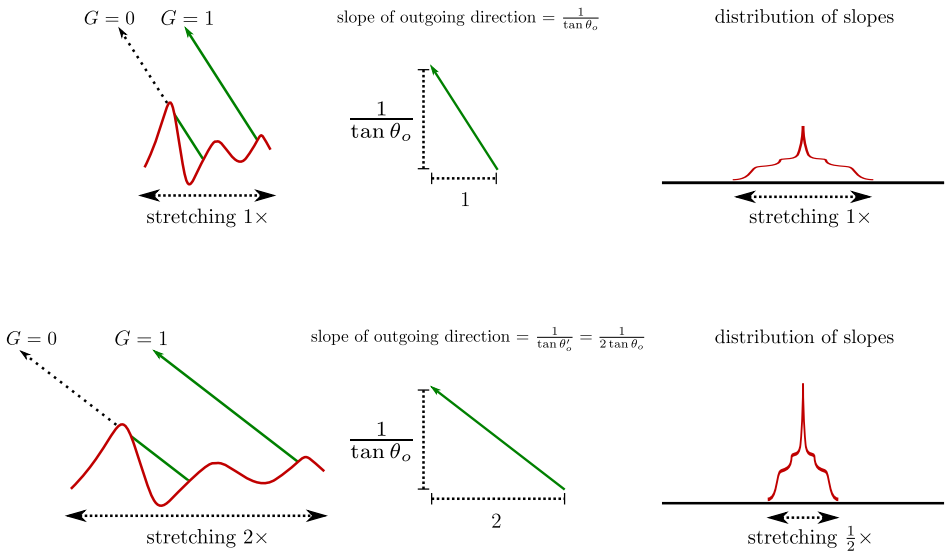


Figure 13. Stretching a 1D configuration by a factor of 2 does not change the masking probability G, but all of the slopes of the configuration are scaled by a factor of $\frac{1}{2}$. This includes the slopes of the microsurface as well as the slope associated with the outgoing direction.

5.2. The Distribution of Slopes

If the microsurface is a heightfield, the distribution of heights of the microsurface is often denoted $P^1(h)$. The slopes of the microsurface are the gradients of the heights:

$(x_{\tilde{m}}, y_{\tilde{m}}) = \nabla h$, i.e. they measure how the microsurface heights vary spatially. The distribution of slopes of the microsurface is denoted $P^{22}(x_{\tilde{m}}, y_{\tilde{m}})$ where

$$\tilde{m} = (x_{\tilde{m}}, y_{\tilde{m}}) = \left(-\frac{x_m}{z_m}, -\frac{y_m}{z_m}\right) = -\tan\theta_m(\cos\phi_m, \sin\phi_m) \tag{61}$$

is the slope associated with the normal $\omega_m = (x_m, y_m, z_m)$ and reciprocally

$$\omega_m = \frac{(-x_{\tilde{m}}, -y_{\tilde{m}}, 1)}{\sqrt{x_{\tilde{m}}^2 + y_{\tilde{m}}^2 + 1}}. \tag{62}$$

The distribution of slopes is necessarily normalized:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P^{22}(x_{\tilde{m}}, y_{\tilde{m}})\, dx_{\tilde{m}} dy_{\tilde{m}} = 1, \tag{63}$$

and the distribution of normals is constructed from it:

$$D(\omega_m) = \frac{P^{22}(x_{\tilde{m}}, y_{\tilde{m}})}{\cos^4\theta_m}. \tag{64}$$

When the roughness parameters must be explicit, we use the notation $D(\omega_m, \alpha)$ and $P^{22}(x_{\tilde{m}}, y_{\tilde{m}}, \alpha)$ for isotropic distributions and $D(\omega_m, \alpha_x, \alpha_y)$ and $P^{22}(x_{\tilde{m}}, y_{\tilde{m}}, \alpha_x, \alpha_y)$ for anisotropic distributions.

5.3. Isotropic Shape-Invariant Distributions of Slopes

Shape Invariance Several isotropic parametric distributions of slopes P^{22} 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 distribution of slopes depends only on the ratio $\frac{\tan\theta_m}{\alpha}$ between the slope amplitude $\tan\theta_m = \sqrt{x_{\tilde{m}}^2 + y_{\tilde{m}}^2}$ of a normal of angle θ_m, and the roughness parameter α:

$$P^{22}(x_{\tilde{m}}, y_{\tilde{m}}, \alpha) = \frac{1}{\alpha^2} f\left(\frac{\sqrt{x_{\tilde{m}}^2 + y_{\tilde{m}}^2}}{\alpha}\right) = \frac{1}{\alpha^2} f\left(\frac{\tan\theta_m}{\alpha}\right), \tag{65}$$

where f is a 1D function that defines the shape of the distribution. These distributions of slopes are *shape invariant*, because distributions that exhibit this property always have the same shape f and are only stretched and scaled by the roughness parameter:

$$P^{22}(x_{\tilde{m}}, y_{\tilde{m}}, \alpha) = \frac{1}{\lambda^2} P^{22}\left(\frac{x_{\tilde{m}}}{\lambda}, \frac{y_{\tilde{m}}}{\lambda}, \frac{\alpha}{\lambda}\right), \quad \text{for any } \lambda > 0. \tag{66}$$

As shown in Figure 13, with isotropic shape-invariant distributions of slopes, stretching the configuration is equivalent to scaling the roughness parameters α and the slope of the outgoing vector by the same factor. It implies that the masking function depends

only on the variable $a = \frac{1}{\alpha \tan\theta_o}$, where $\frac{1}{\tan\theta_o}$ is the slope of the outgoing direction[2]. Beckmann and GGX distributions are shape invariant and this is why their associated functions Λ depend only on a, where Λ appears in the Smith masking function given in Equation (43).

Beckmann Distribution

$$P^{22}(x_{\tilde{m}}, y_{\tilde{m}}) = \frac{1}{\pi\alpha^2} \exp\left(-\frac{x_{\tilde{m}}^2 + y_{\tilde{m}}^2}{\alpha^2}\right), \tag{67}$$

$$D(\omega_m) = \frac{\chi^+(\omega_m \cdot \omega_g)}{\pi\alpha^2 \cos^4\theta_m} \exp\left(-\frac{\tan^2\theta_m}{\alpha^2}\right), \tag{68}$$

$$\Lambda(\omega_o) = \frac{\mathrm{erf}(a) - 1}{2} + \frac{1}{2a\sqrt{\pi}} \exp\left(-a^2\right). \tag{69}$$

where $a = \frac{1}{\alpha \tan\theta_o}$. Walter et al. [Walter et al. 2007] 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

$$P^{22}(x_{\tilde{m}}, y_{\tilde{m}}) = \frac{1}{\pi\alpha^2 \left(1 + \frac{x_{\tilde{m}}^2 + y_{\tilde{m}}^2}{\alpha^2}\right)^2}, \tag{70}$$

$$D(\omega_m) = \frac{\chi^+(\omega_m \cdot \omega_g)}{\pi\alpha^2 \cos^4\theta_m \left(1 + \frac{\tan^2\theta_m}{\alpha^2}\right)^2}, \tag{71}$$

$$\Lambda(\omega_o) = \frac{-1 + \sqrt{1 + \frac{1}{a^2}}}{2}, \tag{72}$$

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

Shape-variant distributions It should be noted that not all distributions are shape invariant. For instance, the Phong distribution is not because it cannot be expressed in the form of Equation (65). In other words, as the roughness changes, the shape of the Phong distribution changes.

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

5.4. Anisotropic Shape-Invariant Distributions of Slopes

Shape Invariance The same shape-invariant distributions can be anisotropic if the shape is stretched with azimuth-dependent factors. The slopes are weighted separately in each direction and Equation (65) is replaced by

$$P^{22}(x_{\tilde{m}}, y_{\tilde{m}}, \alpha_x, \alpha_y) = \frac{1}{\alpha_x \alpha_y} f\left(\sqrt{\frac{x_{\tilde{m}}^2}{\alpha_x^2} + \frac{y_{\tilde{m}}^2}{\alpha_y^2}}\right) = \frac{1}{\alpha_x \alpha_y} f\left(\tan\theta_m \sqrt{\frac{\cos^2\phi_m}{\alpha_x^2} + \frac{\sin^2\phi_m}{\alpha_y^2}}\right), \tag{73}$$

where $(-\tan\theta_m \cos\phi_m, -\tan\theta_m \sin\phi_m) = (x_{\tilde{m}}, y_{\tilde{m}})$ is the slope and α_x and α_y are the stretching coefficients of the distribution in the x- and y-axis, respectively. The shape invariance property is written

$$P^{22}(x_{\tilde{m}}, y_{\tilde{m}}, \alpha_x, \alpha_y) = \frac{1}{\lambda_x \lambda_y} P^{22}\left(\frac{x_{\tilde{m}}}{\lambda_x}, \frac{y_{\tilde{m}}}{\lambda_y}, \frac{\alpha_x}{\lambda_x}, \frac{\alpha_y}{\lambda_y}\right), \quad \text{for any } \lambda_x, \lambda_y > 0. \tag{74}$$

Derivation of the Masking Function Figure 14 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 outgoing 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

$$\alpha_x' = \alpha_x \frac{\alpha_y}{\alpha_x} = \alpha_y, \tag{75}$$

$$\alpha_y' = \alpha_y. \tag{76}$$

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

$$\omega_o' = \left(\frac{\alpha_x}{\alpha_y} x_o, y_o, z_o\right) = \left(\frac{\alpha_x}{\alpha_y} \cos\phi_o \sin\theta_o, \sin\phi_o \sin\theta_o, \cos\theta_o\right), \tag{77}$$

$$\frac{1}{\tan\theta_o'} = \frac{z_o}{\sqrt{\frac{\alpha_x^2}{\alpha_y^2} x_o^2 + y_o^2}} = \frac{1}{\sqrt{\frac{\alpha_x^2}{\alpha_y^2} \cos^2\phi_o + \sin^2\phi_o} \tan\theta_o}. \tag{78}$$

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

$$\begin{aligned} a' = \frac{1}{\alpha_y \tan\theta'_o} &= \frac{1}{\alpha_y \sqrt{\cos^2\phi_o \frac{\alpha_x^2}{\alpha_y^2} + \sin^2\phi_o} \tan\theta_o} \\ &= \frac{1}{\sqrt{\cos^2\phi_o \alpha_x^2 + \sin^2\phi_o \alpha_y^2} \tan\theta_o} \\ &= \frac{1}{\alpha_o \tan\theta_o}, \end{aligned} \tag{79}$$

where

$$\alpha_o = \sqrt{\cos^2\phi_o \alpha_x^2 + \sin^2\phi_o \alpha_y^2} \tag{80}$$

is the *roughness projected onto the outgoing direction*. This shows that the masking function associated with a given anisotropic shape-invariant distribution of slopes is essentially the same as the isotropic version. The only difference is that it is parameterized by the roughness of the anisotropic surface projected onto the outgoing direction. We use this property to derive the masking functions for the anisotropic Beckmann and GGX distributions.

Anisotropic Beckmann Distribution

$$P^{22}(x_{\tilde{m}}, y_{\tilde{m}}) = \frac{1}{\pi\alpha_x\alpha_y} \exp\left(-\frac{x_{\tilde{m}}^2}{\alpha_x^2} - \frac{y_{\tilde{m}}^2}{\alpha_y^2}\right), \tag{81}$$

$$D(\omega_m) = \frac{\chi^+(\omega_m \cdot \omega_g)}{\pi\alpha_x\alpha_y \cos^4\theta_m} \exp\left(-\tan^2\theta_m \left(\frac{\cos^2\phi_m}{\alpha_x^2} + \frac{\sin^2\phi_m}{\alpha_y^2}\right)\right), \tag{82}$$

$$\Lambda(\omega_o) = \frac{\mathrm{erf}(a) - 1}{2} + \frac{1}{2a\sqrt{\pi}} \exp\left(-a^2\right), \tag{83}$$

where $a = \frac{1}{\alpha_o \tan\theta_o}$ and α_o is defined in Equation 80. The approximation of Λ for the isotropic Beckmann distribution can be used as well.

Anisotropic GGX Distribution

$$P^{22}(x_{\tilde{m}}, y_{\tilde{m}}) = \frac{1}{\pi\alpha_x\alpha_y \left(1 + \frac{x_{\tilde{m}}^2}{\alpha_x^2} + \frac{y_{\tilde{m}}^2}{\alpha_y^2}\right)^2}, \tag{84}$$

$$D(\omega_m) = \frac{\chi^+(\omega_m \cdot \omega_g)}{\pi\alpha_x\alpha_y \cos^4\theta_m \left(1 + \tan^2\theta_m \left(\frac{\cos^2\phi_m}{\alpha_x^2} + \frac{\sin^2\phi_m}{\alpha_y^2}\right)\right)^2}, \tag{85}$$

$$\Lambda(\omega_o) = \frac{-1 + \sqrt{1 + \frac{1}{a^2}}}{2}, \tag{86}$$

where $a = \frac{1}{\alpha_o \tan\theta_o}$ and α_o is defined in Equation 80.

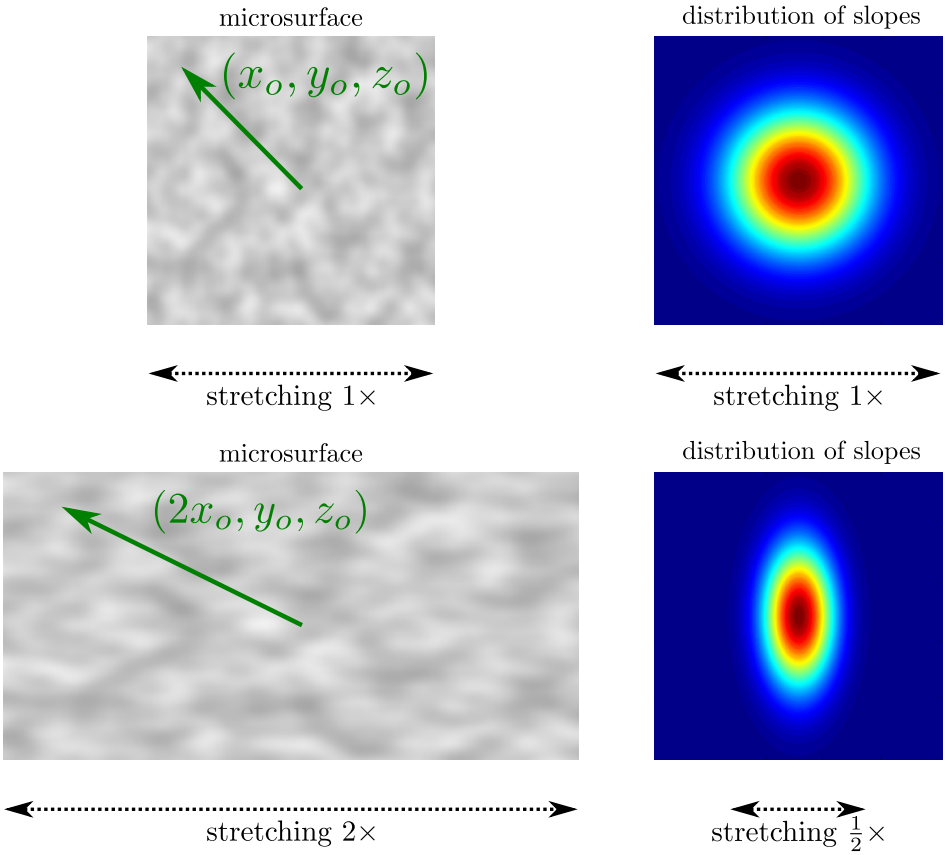


Figure 14. Stretching a 2D configuration by a factor of 2. The coordinates of the outgoing vector are stretched as well.

5.5. More Generalization

Arbitrary Shape-Invariant Distributions An important property of shape-invariant distributions is that all of 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 one's own shape-invariant anisotropic distribution of normals by choosing an arbitrary 1D function f and set

$$D(\omega_m) = \frac{c}{\alpha_x \alpha_y \cos\theta_m^4} f\left(\tan^2\theta_m \left(\frac{\cos^2\phi_m}{\alpha_x^2} + \frac{\sin^2\phi_m}{\alpha_y^2}\right)\right), \tag{87}$$

where c would be the constant 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 Q be a symmetric positive-definite matrix[3]:

$$Q^{-1} = \begin{bmatrix} \alpha_x^2 & r_{xy}\alpha_x\alpha_y \\ r_{xy}\alpha_x\alpha_y & \alpha_y^2 \end{bmatrix}, \tag{88}$$

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:

$$\begin{aligned} ||\tilde{m}|| &= \sqrt{\langle \tilde{m}, \tilde{m} \rangle} \\ &= \sqrt{\tilde{m}^T Q \tilde{m}}. \end{aligned} \tag{89}$$

Recall that the 2D vector $\tilde{m} = -\tan\theta_m(\cos\phi_m, \sin\phi_m)$ defined in Equation (61) is the slope associated with the normal ω_m. The norm of the slope, $||\tilde{m}||$, describes the stretching that occurs in slope space and is the argument of the distribution shape f. In the simpler case of no correlation ($r_{xy} = 0$), the formulas for the norm and the projected roughness in the outgoing direction are[4]

$$\begin{aligned} ||\tilde{m}||^2 &= \tan\theta_m(\cos\phi_m, \sin\phi_m)^T Q \tan\theta_m(\sin\phi_m, \cos\phi_m) \\ &= \tan^2\theta_m \left(\frac{\cos^2\phi_m}{\alpha_x^2} + \frac{\sin^2\phi_m}{\alpha_y^2}\right), \end{aligned} \tag{90}$$

$$\alpha_o^2 = \cos^2\phi_o \alpha_x^2 + \sin^2\phi_o \alpha_y^2. \tag{91}$$

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

[4] We provide the squared results for convenience.

In the general case ($r_{xy} \neq 0$), we instead have

$$||\tilde{m}||^2 = \tan\theta_m(\cos\phi_m, \sin\phi_m)^T Q \tan\theta_m(\sin\phi_m, \cos\phi_m) \tag{92}$$
$$= \tan^2\theta_m \left(\frac{\cos^2\phi_m\alpha_y^2 + \sin^2\phi_m\alpha_x^2 - 2\cos\phi_m\sin\phi_m r_{xy}\alpha_x\alpha_y}{\alpha_x^2\alpha_y^2 - r_{xy}^2\alpha_x^2\alpha_y^2} \right),$$
$$\alpha_o^2 = \cos^2\phi_o\alpha_x^2 + \sin^2\phi_o\alpha_y^2 + 2\cos\phi_o\sin\phi_o r_{xy}\alpha_x\alpha_y. \tag{93}$$

For instance, in LEADR mapping, a correlated Beckmann distribution is used [Dupuy et al. 2013]. Note that setting the correlation coefficient $r_{xy} \in [-1,1]$ to non-zero values affects the constant normalization factor of D.

Vertical Shearing and Non-Centered Distributions Figure 15 shows that the masking function is also invariant under vertical shearing. Applying a vertical shear on the configuration is equivalent to offsetting all of the slopes of the configuration by a constant value. As before, this includes the slopes of the microsurface and the slope associated with the outgoing direction. We use the term *mesosurface* to refer to the average slope $\bar{\tilde{m}} = (\bar{x}_{\tilde{m}}, \bar{y}_{\tilde{m}})$ of the microsurface:

$$(\bar{x}_{\tilde{m}}, \bar{y}_{\tilde{m}}) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x_{\tilde{m}}, y_{\tilde{m}})\, P^{22}(x_{\tilde{m}}, y_{\tilde{m}})\, dx_{\tilde{m}} dy_{\tilde{m}}, \tag{94}$$

which is represented in blue in the figure. This corresponds to the center of the distribution of slopes. If the distribution is not centered around the origin, i.e. $\bar{\tilde{m}} \neq (0,0)$, one has to include the offset $\bar{\tilde{m}}$ in the computation of the argument of the distribution shape f and of the factor a for the masking function:

$$||\tilde{m}||^2 = (\tan\theta_m(\cos\phi_m, \sin\phi_m) + (\bar{x}_{\tilde{m}}, \bar{y}_{\tilde{m}}))^T Q (\tan\theta_m(\cos\phi_m, \sin\phi_m) + (\bar{x}_{\tilde{m}}, \bar{y}_{\tilde{m}}))$$
$$= \frac{(\tan\theta_m\cos\phi_m + \bar{x}_{\tilde{m}})^2\alpha_y^2}{\alpha_x^2\alpha_y^2 - r_{xy}^2\alpha_x^2\alpha_y^2} + \frac{(\tan\theta_m\sin\phi_m + \bar{y}_{\tilde{m}})^2\alpha_x^2}{\alpha_x^2\alpha_y^2 - r_{xy}^2\alpha_x^2\alpha_y^2}$$
$$- 2\frac{(\tan\theta_m\cos\phi_m + \bar{x}_{\tilde{m}})(\tan\theta_m\sin\phi_m + \bar{y}_{\tilde{m}}) r_{xy}\alpha_x\alpha_y}{\alpha_x^2\alpha_y^2 - r_{xy}^2\alpha_x^2\alpha_y^2}, \tag{95}$$
$$a = \frac{\frac{1}{\tan\theta_o} - (\cos\phi_o\bar{x}_{\tilde{m}} + \sin\phi_o\bar{y}_{\tilde{m}})}{\alpha_o}, \tag{96}$$
$$\alpha_o^2 = \cos^2\phi_o\alpha_x^2 + \sin^2\phi_o\alpha_y^2 + 2\cos\phi_o\sin\phi_o r_{xy}\alpha_x\alpha_y. \tag{97}$$

Note that vertical shearing does not affect the projected roughness α_o and the normalization factor of the distribution. This makes sense because stretching changes the shape of the distribution, and thus the roughness, while shearing offsets the distribution without changing its shape. It's tempting to believe that, because the roughness and the normalization factor are invariant under shearing—which alters all the slopes, and hence the normal vectors—these might also be invariant under a rotation of the

normals. They are not, because the mapping from normal vector to facet slope does not transform rotations of vectors to translations of slope values.

Typically, the distribution of slopes is centered around 0, which means that the mesosurface is aligned with the geometric surface. However, this assumption is wrong when the macrogeometry is amplified by 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 [Olano and Baker 2010], a multi-scale non-centered Gaussian distribution of slopes 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 ensure that everything is still well defined. Fortunately, the vertical shear invariance shows that the masking function of a non-centered microsurface is the same as the masking function of a centered microsurface with offset slopes. This property was used in LEADR mapping [Dupuy et al. 2013], where microfacet theory is extended to non-centered distributions.

Another important consideration 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 $\frac{\omega_{\tilde{n}}\cdot\omega_o}{\omega_{\tilde{n}}\cdot\omega_g}$, where $\omega_{\tilde{n}}$ is the normal of the mesosurface. In the case where the mesosurface is the geometric surface, we have $\omega_{\tilde{m}} = \omega_g$ and we get back to $\frac{\omega_{\tilde{m}}\cdot\omega_o}{\omega_{\tilde{m}}\cdot\omega_g} = \frac{\cos\theta_o}{1}$, so this is consistent. More details are available in the LEADR mapping paper.

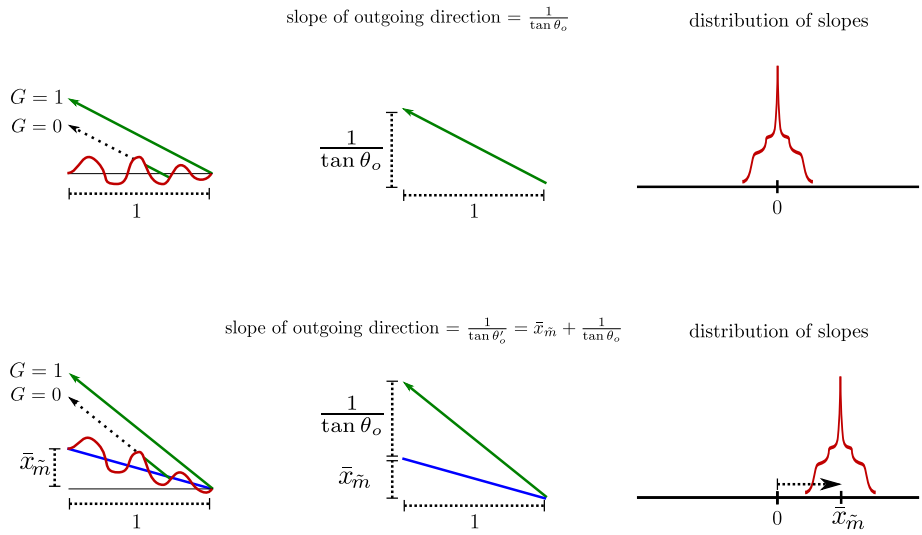


Figure 15. Vertical shearing of a 1D configuration. This does not change masking probability G_1, but all of the slopes of the configuration are offset by a constant factor $\bar{x}_{\tilde{m}}$. This includes the slopes of the microsurface as well as the slope associated with the outgoing direction. The distribution of slopes is shifted by an offset $\bar{x}_{\tilde{m}}$ and is no longer centered around 0.

6. The Smith 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: the masking-shadowing function. We recall four different forms of the masking-shadowing function. Each of these use the function Λ defined in Section 5—evaluated for the outgoing ($\Lambda(\omega_o)$) and incident ($\Lambda(\omega_i)$) directions—and combine them in different ways, producing different properties as a result.

Separable Masking and Shadowing The simplest and most widely used variant of the masking-shadowing function is the separable form popularized by Walter et al. [2007]. In this instance, masking and shadowing are supposed to be independent, and are computed separately and multiplied together:

$$\begin{aligned} G_2(\omega_o, \omega_i, \omega_m) &= G_1(\omega_o, \omega_m)\, G_1(\omega_i, \omega_m) \\ &= \frac{\chi^+(\omega_o \cdot \omega_m)}{1+\Lambda(\omega_o)} \frac{\chi^+(\omega_i \cdot \omega_m)}{1+\Lambda(\omega_i)}. \end{aligned} \tag{98}$$

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 [Ross et al. 2005]. Intuitively, the more a microfacet is elevated within the microsurface, the more the probabilities of being visible for the outgoing direction (unmasked) and for the incident direction (unshadowed) increase at the same time. Thus, masking and shadowing are correlated through the elevation of the microfacets. This correlation is accounted for in the following form of the joint masking-shadowing function:

$$G_2(\omega_o, \omega_i, \omega_m) = \frac{\chi^+(\omega_o \cdot \omega_m)\, \chi^+(\omega_i \cdot \omega_m)}{1+\Lambda(\omega_o)+\Lambda(\omega_i)}. \tag{99}$$

This form is accurate when the outgoing and incident directions are far away from each other, but overestimates shadowing when the directions are close. We suggest to use Equation (99) in practice, because it is more accurate than the separable from of Equation (98) 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 outgoing and incident directions are close to one another. Typically, when $\omega_o = \omega_i$, masking and shadowing are perfectly correlated because microfacets visible from direction ω_o are also visible from direction ω_i. In this case, the

shadowing should be removed from the BRDF because shadowed microfacets are not visible from direction ω_i, and thus they are also not visible from ω_o. This is known as the "hotspot effect": when the view and light directions are parallel, shadows disappear[5]. Since the BRDF models the radiance measured along the outgoing direction, if shadowing exists on the surface but is not visible then it should not be part of the BRDF.

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. [2000] account for directional correlation by blending the separable form of Equation (98) with the case where both directions are fully correlated:

$$\begin{aligned} &G_2(\omega_o, \omega_i, \omega_m) \\ &= \lambda(\phi)\, G_1(\omega_o, \omega_m)\, G_1(\omega_i, \omega_m) + (1 - \lambda(\phi))\, \min(G_1(\omega_o, \omega_m), G_1(\omega_i, \omega_m)). \end{aligned} \quad (100)$$

where $\lambda(\phi)$ is an empirical factor similar to Ginneken et al.'s, which is presented next. Because the authors do not have the Smith 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 incorporating a directional correlation factor λ into the height-correlated form:

$$G_2(\omega_o, \omega_i, \omega_m) = \frac{\chi^+(\omega_o \cdot \omega_m)\, \chi^+(\omega_i \cdot \omega_m)}{1 + \max(\Lambda(\omega_o), \Lambda(\omega_i)) + \lambda(\omega_o, \omega_i)\, \min(\Lambda(\omega_o), \Lambda(\omega_i))}. \quad (101)$$

Here, masking and shadowing are fully correlated when the outgoing and incident directions are parallel and $\lambda = 0$. The correlation decreases as the angle between the 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 height-correlated form.

Ginneken et al. [1998] 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 [Heitz et al. 2013]. The result was given for isotropic Beckmann distributions only, but the stretch invariance presented in Section 5 can be used to easily generalize this result to anisotropic Beckmann distributions. This form perfectly models the correlation of masking and shadowing and is

[5]This does not mean that shadows no longer exist, only that they are not visible from this specific view direction.

thus more accurate than the forms presented in Equations (98), (99), and (100). The derivation of practical forms for λ and generalization non-Gaussian distributions are open problems.

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.

Deriving the Smith Masking Function for Other Commonly Used Models We have seen that closed analytical forms of the Smith masking function 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 distributions of normals.

An important example is the Phong distribution[6] [Walter et al. 2007]. Walter et al. proposed to use the Smith masking function for the Beckmann distribution, since they have a similar appearance for low roughness values. However, the more the roughness increases, the more the error becomes significant. 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, since Beckmann is shape invariant, as discussed in Section 5. Indeed, shape invariant distributions and their associated masking functions only depend on the ratio $a = \frac{||\tilde{m}||}{\alpha}$ between the slope amplitude and the roughness. 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, since 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 [Burley 2012], whose masking function has yet to be found.

Correlation of Masking-Shadowing As we have seen in Section 6, 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 distributions of normals is an open problem.

[6]Here, we use "Phong distribution" to mean the distribution of normals recalled by Walter et al. It is not to be confused with the Phong BRDF, which is not a microfacet-based model, nor the original Blinn-Phong BRDF, which uses a non-normalized distribution.

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 [Burley 2012]. The first reflex in the CG community is to keep the standard BRDF formulation and tweak the distribution of normals. For instance, Bagher et al. [2012] 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 [2012] 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 distribution of normals. 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 stretching 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 [Bourlier and Berginc 2004]. However, these models are quite complicated, because the physics community aims for accuracy rather than 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 3.2, 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:

$$\rho(\omega_o, \omega_i) = \rho_1(\omega_o, \omega_i) + \rho_{2^+}(\omega_o, \omega_i), \tag{102}$$

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

$$\begin{aligned}\int_{\Omega_i} \rho(\omega_o, \omega_i) \, |\omega_g \cdot \omega_i| \, d\omega_i &= \int_{\Omega_i} \rho_1(\omega_o, \omega_i) \, |\omega_g \cdot \omega_i| \, d\omega_i + \int_{\Omega} \rho_{2^+}(\omega_o, \omega_i) \, |\omega_g \cdot \omega_i| \, d\omega_i \\ &= 1,\end{aligned} \tag{103}$$

and that the energy present in the multiple scattering term would be completely deter-

mined by the energy loss due to shadowing in the first scattering term:

$$E_{2^+} = \int_{\Omega_i} \rho_{2^+}(\omega_o, \omega_i) \, |\omega_g \cdot \omega_i| \, d\omega_i = 1 - \int_{\Omega_i} \rho_1(\omega_o, \omega_i) \, |\omega_g \cdot \omega_i| \, d\omega_i. \tag{104}$$

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. as 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 F, which could be precomputed via

$$\bar{F}_{\omega_o} = \int_{\Omega} F(\omega_o, \omega_m) \, D_{\omega_o}(\omega_m) \, d\omega_m, \tag{105}$$

and stored in a look-up table. Note that multiplying by $\bar{F}_{\omega_o}$ would rescale E_{2^+} according to the fraction 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 distribution of normals by the projected area of the visible microsurface. By using this knowledge and the choice of a microsurface profile, we have shown that the exact form of the masking function can be derived. We have made these derivations for the Smith microsurface profile, which assumes normal/masking independence, and for the V-cavity microsurface profile, which assumes separate microsurfaces made of V-cavities. We showed why this last model is mathematically well defined yet unrealistic. Upon that, we defined the distribution of visible normals, 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 reviewed different shadowing models. We have demonstrated that the masking function is stretch invariant and how this property can be used to generalize known results to anisotropic distributions of normals. 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.

In the last section, we discussed the limitations of the BRDF model [Cook and Torrance 1982] 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 distributions of normals, with growing complexity and therefore less practicality.

Acknowledgements

I would like to thank Sébastien Lagarde, who encouraged me to write this article and provided initial feedback, Stephen Hill and Naty Hoffman for their advice and numerous corrections, and Jonathan Dupuy for our endless discussions about microfacet theory. The article has also improved significantly since the submitted version thanks to JCGT's anonymous reviewers. This work was partially funded by the Galaxy/veRTIGE project #ANR-10-CORD-006 and the SimeOne project #ANR-10-JCJC-0207.

References

ASHIKMIN, M., PREMOŽE, S., AND SHIRLEY, P. 2000. A microfacet-based brdf generator. In *Proc. SIGGRAPH '00*, 65–74. 50, 53, 69, 70, 92

BAGHER, M. M., SOLER, C., AND HOLZSCHUCH, N. 2012. Accurate fitting of measured reflectances using a shifted gamma micro-facet distribution. *Comp. Graph. Forum 31*, 4 (June), 1509–1518. 94

BECKMANN, P., AND SPIZZICHINO, A. 1963. *The scattering of electromagnetic waves from rough surfaces*. International series of monographs on electromagnetic waves. Pergamon Press. 49

BOURLIER, C., AND BERGINC, G. 2004. 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. 94

BOURLIER, C., SAILLARD, J., AND BERGINC, G. 2000. Effect of correlation between shadowing and shadowed points on the wagner and smith monostatic one-dimensional shadowing functions. *IEEE Trans. on Antennas and Propagation 48*, 3, 437–446. 70

BROWN, G. 1980. Shadowing by non-gaussian random surfaces. *IEEE Trans. on Antennas and Propagation 28*, 6 (nov), 788 – 790. 70, 100

BURLEY, B. 2012. Physically-based shading at disney. In *ACM SIGGRAPH 2012 Courses*, ACM, SIGGRAPH '12, 10:1–7. 93, 94

COOK, R. L., AND TORRANCE, K. E. 1982. A reflectance model for computer graphics. *ACM Trans. Graph. 1*, 1 (Jan.), 7–24. 49, 71, 72, 95

DUPUY, J., HEITZ, E., IEHL, J.-C., POULIN, P., NEYRET, F., AND OSTROMOUKHOV, V. 2013. Linear Efficient Antialiased Displacement and Reflectance Mapping. *ACM Trans. Graph. 32*, 6 (Sept.). 89, 90, 101

HE, X. D., TORRANCE, K. E., SILLION, F. X., AND GREENBERG, D. P. 1991. A comprehensive physical model for light reflection. In *Proceedings of SIGGRAPH '91*, 175–186. 78

HEITZ, E., BOURLIER, C., AND PINEL, N. 2013. 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. 92, 101

KELEMEN, C., AND SZIRMAY-KALOS, L. 2001. A microfacet based coupled specular-matte brdf model with importance sampling. In *Eurographics Short Presentations*. 78

MCAULEY, S., HILL, S., HOFFMAN, N., GOTANDA, Y., SMITS, B., BURLEY, B., AND MARTINEZ, A. 2012. Practical physically-based shading in film and game production. In *ACM SIGGRAPH 2012 Courses*, ACM, SIGGRAPH '12, 10:1–7. 49

MCAULEY, S., HILL, S., MARTINEZ, A., VILLEMIN, R., PETTINEO, M., LAZAROV, D., NEUBELT, D., KARIS, B., HERY, C., HOFFMAN, N., AND ZAP ANDERSSON, H. 2013. Physically based shading in theory and practice. In *ACM SIGGRAPH 2013 Courses*, ACM, SIGGRAPH '13, 22:1–8. 49, 77

OLANO, M., AND BAKER, D. 2010. Lean mapping. In *Proceedings of the 2010 ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, ACM, New York, NY, USA, I3D '10, 181–188. 90

OREN, M., AND NAYAR, S. K. 1994. Generalization of Lambert's reflectance model. In *Proc. SIGGRAPH '94*, 239–246. 49, 63, 71

ROSS, V., DION, D., AND POTVIN, G. 2005. 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 (Nov), 2442–2453. 50, 91, 101

RUMP, M., MÜLLER, G., SARLETTE, R., KOCH, D., AND KLEIN, R. 2008. Photo-realistic rendering of metallic car paint from image-based measurements. *Comp. Graph. Forum 27*, 2 (Apr.), 527–536. 70

SCHLICK, C. 1994. An inexpensive brdf model for physically-based rendering. *Computer Graphics Forum 13*, 233–246. 77

SMITH, B. 1967. Geometrical shadowing of a random rough surface. *IEEE Trans. on Antennas and Propagation 15*, 668–671. 50, 70, 77, 101

VAN GINNEKEN, B., STAVRIDI, M., AND KOENDERINK, J. J. 1998. Diffuse and specular reflectance from rough surfaces. *Appl. Opt. 37*, 1 (Jan), 130–139. 92

WALTER, B., MARSCHNER, S. R., LI, H., AND TORRANCE, K. E. 2007. Microfacet models for refraction through rough surfaces. In *Proc. Eurographics Symposium on Rendering*, EGSR'07, 195–206. 49, 54, 63, 69, 70, 84, 91, 93, 100, 101

A. Derivation of the Masking Function

In this section, we derive $G_1^{\text{dist}}(\omega_o)$ (denoted G_1^{dist} for convenience) starting from Equation (41):

$$\cos\theta_o = G_1^{\text{dist}} \int_\Omega \langle \omega_o, \omega_m \rangle \, D(\omega_m) \, d\omega_m.$$

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 $\omega_m = (x_m, y_m, z_m)$ is defined by

$$\tilde{m}(\omega_m) = (x_{\tilde{m}}, y_{\tilde{m}}) = (-x_m/z_m, -y_m/z_m), \tag{106}$$

and reciprocally

$$\omega_m(\tilde{m}) = (x_m, y_m, z_m) = \frac{1}{\sqrt{x_{\tilde{m}}^2 + y_{\tilde{m}}^2 + 1}} (-x_{\tilde{m}}, -y_{\tilde{m}}, 1), \tag{107}$$

and that the distribution of slopes P^{22} is linked to the distribution of normals by the relationship[7]

$$P^{22}(\tilde{m}) \, d\tilde{m} = (\omega_m \cdot \omega_g) \, D(\omega_m) \, d\omega_m. \tag{108}$$

By using this change of variable in Equation (41), we can write

$$\int_\Omega \langle \omega_o, \omega_m \rangle \, D(\omega_m) \, d\omega_m = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\langle \omega_o, \omega_m(\tilde{m}) \rangle}{\omega_g \cdot \omega_m(\tilde{m})} \, P^{22}(\tilde{m}) \, d\tilde{m}, \tag{109}$$

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

$$\omega_g \cdot \omega_m(\tilde{m}) = \frac{1}{\sqrt{x_{\tilde{m}}^2 + y_{\tilde{m}}^2 + 1}}, \tag{110}$$

The clamped dot product can be expanded as

$$\langle \omega_o, \omega_m(\tilde{m}) \rangle = \frac{\chi^+(-x_o x_{\tilde{m}} - y_o y_{\tilde{m}} + z_o) \, (-x_o x_{\tilde{m}} - y_o y_{\tilde{m}} + z_o)}{\sqrt{x_{\tilde{m}}^2 + y_{\tilde{m}}^2 + 1}}, \tag{111}$$

and so the integral becomes

$$\begin{aligned} &\int_\Omega \langle \omega_o, \omega_m \rangle \, D(\omega_m) \, d\omega_m \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \chi^+(-x_o x_{\tilde{m}} - y_o y_{\tilde{m}} + z_o) \, (-x_o x_{\tilde{m}} - y_o y_{\tilde{m}} + z_o) \, P^{22}(x_{\tilde{m}}, y_{\tilde{m}}) \, dx_{\tilde{m}} dy_{\tilde{m}}. \end{aligned} \tag{112}$$

[7] We recall that the distribution of slopes and of normals are such that $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} P^{22}(\tilde{m}) \, d\tilde{m} = 1$ and $\int_\Omega (\omega_m \cdot \omega_g) \, D(\omega_m) \, d\omega_m = 1$.

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

$$\begin{aligned}
&\int_\Omega \langle \omega_o, \omega_m \rangle D(\omega_m)\, d\omega_m \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \chi^+(-\sin\theta_o x_{\tilde{m}} + \cos\theta_o)\,(-\sin\theta_o x_{\tilde{m}} + \cos\theta_o)\, P^{22}(x_{\tilde{m}}, y_{\tilde{m}})\, dx_{\tilde{m}} dy_{\tilde{m}} \\
&= \int_{-\infty}^{+\infty} \chi^+(-\sin\theta_o x_{\tilde{m}} + \cos\theta_o)\,(-\sin\theta_o x_{\tilde{m}} + \cos\theta_o) \left(\int_{-\infty}^{+\infty} P^{22}(x_{\tilde{m}}, y_{\tilde{m}})\, dy_{\tilde{m}}\right) dx_{\tilde{m}} \\
&= \int_{-\infty}^{+\infty} \chi^+(-\sin\theta_o x_{\tilde{m}} + \cos\theta_o)\,(-\sin\theta_o x_{\tilde{m}} + \cos\theta_o)\, P^{2-}(x_{\tilde{m}})\, dx_{\tilde{m}}, \qquad (113)
\end{aligned}$$

where $P^{2-}(x_{\tilde{m}}) = \int_{-\infty}^{+\infty} P^{22}(x_{\tilde{m}}, y_{\tilde{m}})\, dy_{\tilde{m}}$ is the 1D distribution of slopes in the view direction (aligned with the x-axis). Since

$$-\sin\theta_o x_{\tilde{m}} + \cos\theta_o > 0 \Rightarrow x_{\tilde{m}} < \cot\theta_o, \qquad (114)$$

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

$$\begin{aligned}
&\int_{-\infty}^{+\infty} \chi^+(-\sin\theta_o x_{\tilde{m}} + \cos\theta_o)\,(-\sin\theta_o x_{\tilde{m}} + \cos\theta_o)\, P^{2-}(x_{\tilde{m}})\, dx_{\tilde{m}} \\
&= \int_{-\infty}^{\cot\theta_o} (-\sin\theta_o x_{\tilde{m}} + \cos\theta_o)\, P^{2-}(x_{\tilde{m}})\, dx_{\tilde{m}}. \qquad (115)
\end{aligned}$$

Now we can return to Equation (41):

$$\cos\theta_o = G_1^{\text{dist}} \int_{-\infty}^{\cot\theta_o} (-\sin\theta_o x_{\tilde{m}} + \cos\theta_o)\, P^{2-}(x_{\tilde{m}})\, dx_{\tilde{m}}. \qquad (116)$$

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

$$\cot\theta_o = G_1^{\text{dist}} \int_{-\infty}^{\cot\theta_o} (-x_{\tilde{m}} + \cot\theta_o)\, P^{2-}(x_{\tilde{m}})\, dx_{\tilde{m}}. \qquad (117)$$

Since microfacet distributions are centered, the average slope in any direction is zero ($\int_{-\infty}^{+\infty} P^{2-}(x_{\tilde{m}})\, dx_{\tilde{m}} = 0$) and we can introduce this term into the equation:

$$\cot\theta_o = G_1^{\text{dist}} \int_{-\infty}^{+\infty} x_{\tilde{m}} P^{2-}(x_{\tilde{m}})\, dx_{\tilde{m}} + G_1^{\text{dist}} \int_{-\infty}^{\cot\theta_o} (-x_{\tilde{m}} + \cot\theta_o)\, P^{2-}(x_{\tilde{m}})\, dx_{\tilde{m}}, \qquad (118)$$

and by using $\cot\theta_o = (1 - G_1^{\text{dist}})\cot\theta_o + G_1^{\text{dist}}\cot\theta_o$:

$$\begin{aligned}
(1 - G_1^{\text{dist}})\cot\theta_o + G_1^{\text{dist}}\cot\theta_o &= G_1^{\text{dist}} \int_{-\infty}^{+\infty} x_{\tilde{m}} P^{2-}(x_{\tilde{m}})\, dx_{\tilde{m}} \\
&\quad + G_1^{\text{dist}} \int_{-\infty}^{\cot\theta_o} (-x_{\tilde{m}} + \cot\theta_o)\, P^{2-}(x_{\tilde{m}})\, dx_{\tilde{m}} \qquad (119) \\
(1 - G_1^{\text{dist}})\cot\theta_o &= G_1^{\text{dist}} \int_{-\infty}^{+\infty} x_{\tilde{m}} P^{2-}(x_{\tilde{m}})\, dx_{\tilde{m}} \\
&\quad + G_1^{\text{dist}} \int_{-\infty}^{\cot\theta_o} (-x_{\tilde{m}} + \cot\theta_o)\, P^{2-}(x_{\tilde{m}})\, dx_{\tilde{m}} - G_1^{\text{dist}}\cot\theta_o. \qquad (120)
\end{aligned}$$

Since P^{2-} integrates to 1 we have $G_1^{\text{dist}} \cot\theta_o = G_1^{\text{dist}} \int_{-\infty}^{+\infty} \cot\theta_o \, P^{2-}(x_{\tilde{m}})\, dx_{\tilde{m}}$:

$$\begin{aligned}
(1 - G_1^{\text{dist}}) \cot\theta_o &= G_1^{\text{dist}} \int_{-\infty}^{+\infty} x_{\tilde{m}} P^{2-}(x_{\tilde{m}})\, dx_{\tilde{m}} + G_1^{\text{dist}} \int_{-\infty}^{\cot\theta_o} (-x_{\tilde{m}} + \cot\theta_o)\, P^{2-}(x_{\tilde{m}})\, dx_{\tilde{m}} \\
&\quad - G_1^{\text{dist}} \int_{-\infty}^{+\infty} \cot\theta_o \, P^{2-}(x_{\tilde{m}})\, dx_{\tilde{m}} \\
&= G_1^{\text{dist}} \left(\int_{-\infty}^{+\infty} x_{\tilde{m}} P^{2-}(x_{\tilde{m}})\, dx_{\tilde{m}} - \int_{-\infty}^{\cot\theta_o} x_{\tilde{m}} P^{2-}(x_{\tilde{m}})\, dx_{\tilde{m}} \right) \\
&\quad + G_1^{\text{dist}} \left(\int_{-\infty}^{\cot\theta_o} \cot\theta_o \, P^{2-}(x_{\tilde{m}})\, dx_{\tilde{m}} - \int_{-\infty}^{+\infty} \cot\theta_o \, P^{2-}(x_{\tilde{m}})\, dx_{\tilde{m}} \right) \\
&= G_1^{\text{dist}} \int_{\cot\theta_o}^{+\infty} x_{\tilde{m}} P^{2-}(x_{\tilde{m}})\, dx_{\tilde{m}} - G_1^{\text{dist}} \int_{\cot\theta_o}^{+\infty} \cot\theta_o \, P^{2-}(x_{\tilde{m}})\, dx_{\tilde{m}} \\
&= G_1^{\text{dist}} \int_{\cot\theta_o}^{\infty} (x_{\tilde{m}} - \cot\theta_o)\, P^{2-}(x_{\tilde{m}})\, dx_{\tilde{m}}. \qquad (121)
\end{aligned}$$

By dividing by G_1^{dist} on each side, we get

$$\frac{(1 - G_1^{\text{dist}})}{G_1^{\text{dist}}} = \frac{1}{\cot\theta_o} \int_{\cot\theta_o}^{\infty} (x_{\tilde{m}} - \cot\theta_o)\, P^{2-}(x_{\tilde{m}})\, dx_{\tilde{m}}, \qquad (122)$$

which leads to the final form

$$\boxed{G_1^{\text{dist}}(\omega_o) = \frac{1}{1 + \Lambda(\omega_o)},} \qquad (123)$$

where function Λ is defined by

$$\boxed{\Lambda(\omega_o) = \frac{1}{\cot\theta_o} \int_{\cot\theta_o}^{\infty} (x_{\tilde{m}} - \cot\theta_o)\, P^{2-}(x_{\tilde{m}})\, dx_{\tilde{m}}.} \qquad (124)$$

Our derivation, based on the projected area, has lead us to the generalized form of the Smith masking term [Brown 1980; Walter et al. 2007].

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 [Ross et al. 2005; Heitz et al. 2013; Dupuy et al. 2013] presented in Equation (99):

$$G_2(\omega_o,\omega_i,\omega_m) = \frac{\chi^+(\omega_o\cdot\omega_m)\,\chi^+(\omega_i\cdot\omega_m)}{1+\Lambda(\omega_o)+\Lambda(\omega_i)}. \tag{125}$$

The microsurface is defined by the distribution of normals $D(\omega_m)$, and the associated distribution of slopes is $P^{22}(\tilde{m})$ as presented in Appendix A. We introduce $P^1(h)$, the height distribution of the microsurface. Note that the slopes of the microsurface are simply the gradients of the heights: $\tilde{m} = \nabla h$. Smith's derivation [Smith 1967; Walter et al. 2007] gives the probability that a point at height h with non-backfacing normal ω_m is visible from direction ω_o:

$$G_1(\omega_o,\omega_m,h) = G_1^{\text{local}}(\omega_o,\omega_m)\,G_1^{\text{dist}}(\omega_o,h), \tag{126}$$

where the local and distant masking functions, presented in Section 2.4, are given by

$$G_1^{\text{local}}(\omega_o,\omega_m) = \chi^+(\omega_o\cdot\omega_m), \tag{127}$$

$$G_1^{\text{dist}}(\omega_o,h) = \left(\int_{-\infty}^{h} P^1(h')\,dh'\right)^{\Lambda(\omega_o)}. \tag{128}$$

The height-averaged form is given by

$$\begin{aligned} G_1(\omega_o,\omega_m) &= G_1^{\text{local}}(\omega_o,\omega_m)\int_{-\infty}^{+\infty} G_1^{\text{dist}}(\omega_o,h)\,P^1(h)\,dh \\ &= \chi^+(\omega_o\cdot\omega_m)\int_{-\infty}^{+\infty}\left(\int_{-\infty}^{h} P^1(h')\,dh'\right)^{\Lambda(\omega_o)} P^1(h)\,dh \\ &= \frac{\chi^+(\omega_o\cdot\omega_m)}{1+\Lambda(\omega_o)}, \end{aligned} \tag{129}$$

which is the Smith masking function from Equation (43). 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 h is visible from both directions is just the product of the probabilities:

$$\begin{aligned} G_2(\omega_o,\omega_i,\omega_m,h) &= G_1(\omega_o,\omega_m,h)\,G_1(\omega_i,\omega_m,h) \\ &= G_1^{\text{local}}(\omega_o,\omega_m)\,G_1^{\text{dist}}(\omega_o,h)\,G_1^{\text{local}}(\omega_i,\omega_m)\,G_1^{\text{dist}}(\omega_i,h) \\ &= \chi^+(\omega_o\cdot\omega_m)\left(\int_{-\infty}^{h} P^1(h')\,dh'\right)^{\Lambda(\omega_o)} \chi^+(\omega_i\cdot\omega_m)\left(\int_{-\infty}^{h} P^1(h')\,dh'\right)^{\Lambda(\omega_i)} \\ &= \chi^+(\omega_o\cdot\omega_m)\,\chi^+(\omega_i\cdot\omega_m)\left(\int_{-\infty}^{h} P^1(h')\,dh'\right)^{\Lambda(\omega_o)+\Lambda(\omega_i)}, \end{aligned} \tag{130}$$

and the height-averaged form is given by

$$\begin{aligned}
G_2(\omega_o,\omega_i,\omega_m) &= \int_{-\infty}^{+\infty} G_2(\omega_o,\omega_i,\omega_m)\,P^1(h)\,dh \\
&= \int_{-\infty}^{+\infty} \chi^+(\omega_o\cdot\omega_m)\,\chi^+(\omega_i\cdot\omega_m)\left(\int_{-\infty}^{h} P^1(h')\,dh'\right)^{\Lambda(\omega_o)+\Lambda(\omega_i)} P^1(h)\,dh \\
&= \chi^+(\omega_o\cdot\omega_m)\,\chi^+(\omega_i\cdot\omega_m)\int_{-\infty}^{+\infty}\left(\int_{-\infty}^{h} P^1(h')\,dh'\right)^{\Lambda(\omega_o)+\Lambda(\omega_i)} P^1(h)\,dh \\
&= \frac{\chi^+(\omega_o\cdot\omega_m)\,\chi^+(\omega_i\cdot\omega_m)}{1+\Lambda(\omega_o)+\Lambda(\omega_i)},
\end{aligned} \tag{131}$$

which is the height-correlated masking-shadowing function presented in Equation (99).

C. MATLAB Code for the Weak White Furnace Test

In this section, we provide code to numerically compute the integral in Equation (36):

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

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

```
function [integral] = 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;
end
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;
  end
  % integrate
  integral = integral + sin(theta) * D * G / abs(4 * V(3));
end
end

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

```
function [integral] = 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;
end
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 = D / (pi * alpha_x * alpha_y * H(3)^4);
  else
    continue;
  end
  % integrate
  integral = integral + sin(theta) * D * G / abs(4 * V(3));
end
end

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

```
function [integral] = 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 = D / (pi * alpha^2 * H(3)^4);
  else
    D = 0;
  end
  % integrate
  integral = integral + sin(theta) * D * G / abs(4 * V(3));
end
end

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

```
function [integral] = 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 = D / (pi * alpha_x * alpha_y * H(3)^4);
  else
    D = 0;
  end
  % integrate
  integral = integral + sin(theta) * D * G / abs(4 * V(3));
end
end

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

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.

D. Author Contact Information

Eric Heitz
INRIA ; CNRS ; Univ. Grenoble Alpes
eric.heitz@inria.fr

Eric Heitz, Understanding the Masking-Shadowing Function in Microfacet-Based BRDFs, *Journal of Computer Graphics Techniques (JCGT)*, vol. 3, no. 2, 48–107, 2014
http://jcgt.org/published/0003/02/03/

Received:	2014-01-24		
Recommended:	2014-05-14	Corresponding Editors:	Stephen Hill and Naty Hoffman
Published:	2014-06-30	Editor-in-Chief:	Morgan McGuire