Importance Sampling Microfacet-Based BSDFs using the Distribution of Visible Normals
Eric Heitz, Eugene d’Eon

To cite this version:

HAL Id: hal-00996995
https://inria.hal.science/hal-00996995v2
Submitted on 16 Jun 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Importance Sampling Microfacet-Based BSDFs using the Distribution of Visible Normals

E. Heitz\(^1\) and E. d’Eon\(^2\)

\(^1\)INRIA; CNRS; Univ. Grenoble Alpes
\(^2\)Jig Lab, Wellington, New Zealand

Abstract

We present a new approach to microfacet-based BSDF importance sampling. Previously proposed sampling schemes for popular analytic BSDFs typically begin by choosing a microfacet normal at random in a way that is independent of direction of incident light. To sample the full BSDF using these normals requires arbitrarily large sample weights leading to possible fireflies. Additionally, at grazing angles nearly half of the sampled normals face away from the incident ray and must be rejected, making the sampling scheme inefficient. Instead, we show how to use the distribution of visible normals directly to generate samples, where normals are weighted by their projection factor toward the incident direction. In this way, no backfacing normals are sampled and the sample weights contain only the shadowing factor of outgoing rays (and additionally a Fresnel term for conductors). Arbitrarily large sample weights are avoided and variance is reduced. Since the BSDF depends on the microsurface model, we describe our sampling algorithm for two models: the V-cavity and the Smith models. We demonstrate results for both isotropic and anisotropic rough conductors and dielectrics with Beckmann and GGX distributions.

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Keywords: Microfacet BSDF, Importance Sampling

Figure 1: A dielectric glass plate ($n = 1.5$) with anisotropic GGX roughness ($\alpha_x = 0.05, \alpha_y = 0.4$) on all faces (with the Smith masking function). For a similar sample budget and the same render time, our method (right) significantly reduces the variance and converges faster than the common technique used in previous work (left).
1. Introduction

Rendering realistic surfaces requires the use of complex material models. The most common way to define surface reflectance is the Bidirectional Scattering Distribution Functions (BSDF), which incorporates light transports at a surface’s interface such as reflection, described by a Bidirectional Reflectance Distribution Functions (BRDF), or transmission, described by a Bidirectional Transmittance Distribution Functions (BTDF). Microfacet theory is the most common theoretical framework for deriving analytic physically-based BSDFs for rough surfaces. For Monte Carlo rendering algorithms to make images efficiently using BSDFs, an efficient importance sampling must be provided. In this paper we present a new strategy for deriving importance sampling for analytic microfacet BSDFs. In contrast to previous approaches, we exploit the structure of the microfacet theory to directly sample from the distribution of visible normals, which produces higher quality and lower variance samples in comparison to previous approaches.

A key element of our approach is to move away from normal distributions directly and use, instead, the distribution of slopes for the microsurface. In slope space, we note two key invariance relationships that produce an optimal sampling strategy for any surface roughness or anisotropy, provided the optimal strategy for an isotropic surface with unit roughness is known.

In Section 2, we review the microfacet-based BSDF model and we explain why the common associated importance sampling scheme is not optimal. In Section 3, we show that importance sampling the BSDF with the distribution of visible normals makes a better use of the sampling space and produces sample weights in $[0, 1]$. This method does not create firefly artifacts and leads to a significative variance reduction (Figure 2). To make this practical, in Section 4, we show how to importance sample the distribution of visible normals analytically with the V-cavity microsurface model. In Section 5, we show how to importance sample the distribution of visible normals with the Smith microsurface model. We propose analytic sampling procedures for Beckmann and GGX distributions in the associated supplemental material. For other distributions, if an analytic sampling procedure is not available we propose to use precomputed data. Thanks to the shape invariance property of the distribution of visible slopes, we can do the precomputations for only one BSDF with roughness $\alpha = 1$ and we use the same data with arbitrary roughness ($\alpha \neq 1$) and anisotropy ($\alpha_x \neq \alpha_y$). In this way, we importance sample an entire class of anisotropic parametric BSDFs with a small amount of precomputed data.

2. Background and Previous Work

2.1. Microfacet Theory

Microfacet theory was originally developed in the early work of Torrance et al. [TS67, CT82]. Several alternative models or approximations are built on top of this theory [War92, vGSK98, AS00, KSK01]. This paper is built on the top of two major previous works:

- The extension of microfacet theory to transmittance by Stam [Sta01]. Closely related, Walter et al. [WMLT07] describe BSDF importance sample strategies for different distribution of normals.
- Heitz’ investigation of the properties of the masking-shadowing function in microfacet-based BRDFs [Hei14]. He shows how the masking function is related to the microsurface model and to the correct normalization of the BRDF.

In the following, we review microfacet theory as presented in these two previous works.

![Figure 2: An isotropic rough conductor with GGX distribution ($\alpha = 0.05$) and the Smith masking function (64 samples per pixel).](image)

![Figure 3: The physically-based BSDF model is constructed by applying the operators reflect and transmit on the distribution of visible normals. The ratio of reflection and transmission is given by the Fresnel term $F$.](image)
The Distribution of Normals $D$ describes the statistical distribution of the normals $\omega_m$ of which the microsurface is made. The area of the microsurface projected onto the geometric normal $\omega_g$ is the unit area of the geometric surface:

$$\int_D (\omega_g \cdot \omega_o) D(\omega_m) d\omega_m = 1.$$  \hfill (1)

Parametric models typically specify roughness parameters $\alpha_x$ and $\alpha_y$ in two orthogonal tangent directions ($\alpha_x = \alpha_y$ for isotropic BSDFs). When these parameters need to be included explicitly we write $D(\alpha_m, \alpha_x, \alpha_y)$ instead of $D(\omega_m)$.

The Distribution of Visible Normals $D_{\omega}$ describes the statistical distribution of normals $\omega_m$ visible from the incident direction $\omega_i$ and is defined by

$$D_{\omega} (\omega_m) = \frac{G_1 (\omega_i, \omega_m) |\omega_i \cdot \omega_m| D(\omega_m)}{|\omega_i \cdot \omega_o|}, \quad \hfill (2)$$

where the term $|\omega_m \cdot \omega_o|$ is the projected area of the normal $\omega_m$ in direction $\omega_o$, $|\omega_i \cdot \omega_o|$ is the projected area of the geometric surface in direction $\omega_o$, and $G_1$ is the masking function. The relation between $D$ and $D_{\omega}$ is illustrated in Figure 3. Heitz [Hei14] shows that in a mathematically well-defined BSDF model, the masking function $G_1$ should be such that the distribution is normalized:

$$\int_D D_{\omega} (\omega_m) d\omega_m = 1.$$  

Our approach is to use the knowledge of the BSDF model to improve the sampling strategy. We can do it only for mathematically well-defined models. Two common microsurface models have masking functions $G_1$ that satisfy this criteria: the V-cavity and the Smith models.

The V-cavity Microsurface Model was introduced by Torrance et al. [TS67, CT82]. It assumes a continuous distribution of separate microsurfaces rather than just one microsurface [Hei14]. Each microsurface is composed of two normals $\omega_m = (x_m, y_m, z_m)$ and $\omega_m' = (-x_m, -y_m, z_m)$ (see Figure 11 in Appendix B) and the contribution of each surface is weighted by PDF $P(\omega_m \cdot \omega_o) D(\omega_m)$ in the BSDF. The masking function of a V-cavity microsurface is given by

$$G_1 (\omega_i, \omega_m) = \chi^+ \left( \frac{\omega_i \cdot \omega_m}{\omega_i \cdot \omega_o} \right) \min \left( 1, \frac{\omega_m \cdot \omega_o}{|\omega_i \cdot \omega_m|} \right), \quad \hfill (3)$$

where the heaviside function ensures that backfacing microfacets are discarded for the correct side of the microsurface. The traditional masking-shadowing function accounts for height correlation in the BRDF: $\omega_o$ is in the upper hemisphere ($\omega_o \cdot \omega_y > 0$) and masking and shadowing occur on the same side of the microfacet and are correlated. We generalize this to the BTDF: $\omega_o$ is in the lower side of the microsurface and masking and shadowing occur on opposite sides of the microfacet and are anticorrelated.

The Smith Microsurface Model. The Smith masking function was originally derived from a raytracing formulation [Smi67] for Beckmann distributions $D$ and generalized to other distributions by Brown [Bro80]. Recently, Heitz showed that the Smith masking function can be derived from the equation of the conservation of the projected area [Hei14]. The Smith microsurface model is the most accurate model for compactly describing geometrical-optics reflectance and transmission from explicit random height fields (such as an ocean surface), and is better than the V-cavity model for approximating reflectance of measured material [Hei14]. However, specific masking functions $G_1$ must be derived for each new distribution of normals. Analytic solutions are available for Beckmann [Smi67] and GGX [WMLT07]. Heitz showed that the same formulas can be used with the anisotropic extensions of $D$ [Hei14]. Different forms of the Smith masking-shadowing function are available, the most simple is the non-correlated form:

$$G_2 (\omega_i, \omega_o, \omega_m) = G_1 (\omega_i, \omega_o) G_1 (\omega_i, \omega_m).$$

\[\text{Table 1: Notation.}\]

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_g$</td>
<td>(0, 0, 1) for geometric normal</td>
</tr>
<tr>
<td>$\omega_m$</td>
<td>($x_m, y_m, z_m$) for microfacet normal</td>
</tr>
<tr>
<td>$\omega_o$</td>
<td>($x_o, y_o, z_o$) for incident direction $\omega_i$</td>
</tr>
<tr>
<td>$\omega_i$</td>
<td>($x_i, y_i, z_i$) for outgoing direction $\omega_i'$</td>
</tr>
<tr>
<td>$D(\omega_m)$</td>
<td>distribution of normals</td>
</tr>
<tr>
<td>$D_{\omega}$</td>
<td>distribution of visible normals</td>
</tr>
<tr>
<td>$G_1$</td>
<td>masking function</td>
</tr>
<tr>
<td>$g$</td>
<td>dot product $\omega_i \cdot \omega_o$</td>
</tr>
<tr>
<td>$m$</td>
<td>absolute value of the dot product $</td>
</tr>
<tr>
<td>$\chi^+$</td>
<td>Heaviside function: $1$ if $a &gt; 0$ and $0$ if $a \leq 0$</td>
</tr>
</tbody>
</table>

\[\text{© 2014 The Author(s)}\]

Computer Graphics Forum © 2014 The Eurographics Association and John Wiley & Sons Ltd.
The Scattering Model is given by the local illumination equation:

\[ L(\omega_b) = \int_{\Omega} L(\omega_o) |\omega_o \cdot \omega_b| f(\omega_o, \omega_b) d\omega_o, \tag{4} \]

where \( f \) is the BSDF and \( L(\omega_o) \) is the radiance in the outgoing direction. The construction of the BSDF model is detailed by Walter et al. [WMLT07] and is illustrated in Figure 3(b): the distribution of visible normals is transformed into a distribution of outgoing directions \( \omega_o \) by reflection and transmission. The BSDF \( f(\omega_i, \omega_o) = f_r(\omega_i, \omega_o) + f_t(\omega_i, \omega_o) \) is the sum of the BRDF

\[ f_r(\omega_i, \omega_o) \sim \frac{F(\omega_i, \omega_o, \omega_{in}) G_2(\omega_i, \omega_o, \omega_{in}) D(\omega_{in})}{4|\omega_o \cdot \omega_i| |\omega_i \cdot \omega_{in}|}, \]

and the BTDF

\[ f_t(\omega_i, \omega_o, \omega_{in}) = \frac{|\omega_i \cdot \omega_{in}|}{|\omega_i \cdot \omega_{in}|} \frac{|\omega_o \cdot \omega_{in}|}{|\omega_o \cdot \omega_{in}|} \left( n_b(\omega_i \cdot \omega_o) + n_o(\omega_i \cdot \omega_{in}) \right)^2, \]

where \( \omega_{in} \) and \( \omega_{in} \) are the half vectors of reflection and transmission respectively, and \( n_b \) and \( n_o \) are the indices of refraction of the incident and transmitted sides respectively (we refer the reader to [WMLT07] for more details).

2.2. Importance Sampling the BSDF

Typically, importance sampling is used to solve Equation (4). The most common importance sampling strategy for parametric BSDFs uses the distribution of normals to generate the samples [War92, AS00, APS00, KSK01, Ash07, WMLT07]. In this paper, we call it the “previous method” and it is illustrated in Algorithm 1. In order to importance sample the cosine weighted BSDF \( |\omega_i \cdot \omega_o| f(\omega_i, \omega_o) \), previous algorithms start by generating normals \( \omega_{in} \) with the PDF \( |\omega_i \cdot \omega_{in}| D(\omega_{in}) \) rather than sampling outgoing directions \( \omega_o \) directly. Then, a light transport operator is applied on the sampled normal to generate the outgoing direction. The transport operators, reflect

\[ \omega_o = 2|\omega_o \cdot \omega_{in}| |\omega_{in} - \omega_i|, \]

and transmit

\[ \omega_o = \left( nc - \text{sign}(\omega_i \cdot \omega_{o}) \sqrt{1 + n(c^2 - 1)} \right) \omega_{in} - n\omega_i, \]

with \( c = (\omega_i \cdot \omega_{in}) \), and \( n = \frac{n_b}{n_o} \),

are chosen randomly and with a probability depending on the Fresnel term: \( F(\omega_i, \omega_{in}) \) is the reflection probability and \( 1 - F(\omega_i, \omega_{in}) \) is the transmission probability. The PDF resulting from this procedure is the PDF used to generate the normals multiplied by the Jacobian of the chosen operator:

\[ \left\| \frac{\partial \omega_{in}}{\partial \omega_i} \right\| = \frac{1}{4|\omega_i \cdot \omega_{in}|}, \]

Since the same Jacobians are present in the sampling PDF and in the BRDF/BTDF expressions, they simplify out in the sample weight expression. Also, since reflection and transmission are sampled randomly according to the Fresnel term, the value of Fresnel is present in the sampling PDF as well as in the BRDF/BTDF, and simplifies out in the sample weight expression. In both cases the final expression for the sample weight reduces to

\[ \text{weight}(\omega_o) = |\omega_o \cdot \omega_{in}| G_2(\omega_i, \omega_o, \omega_{in}) \]

Discussion This sampling algorithm is almost perfect at normal incidence but we can see in Figure 4 that it has two main problems at grazing angles. The first problem is that up to half of the generated normals \( \omega_{in} \) are backfacing \((\omega_i \cdot \omega_{in} < 0)\). The associated weights are 0 and so half of the sample space is wasted. The second problem at grazing angles is that the sample weights can be arbitrarily high with the Smith model leading to unwanted variance, especially for paths with multiple surface interactions (though the sample weights are at most 2 with the V-cavity model). This is responsible for the firefly artifacts in Figure 2. Walter et al. [WMLT07] notice this problem and propose using a BSDF with a slightly different roughness \( \alpha' \) for sampling the NDF at grazing angles: \( \alpha' = (1.2 - 0.2/\sqrt{|\omega_i \cdot \omega_{in}|(\alpha)}) \). This is denoted as “Walter’s trick” in Figure 4. Walter’s trick significantly reduces the maximal weight with Beckmann distributions but does not help much with GGX (in our experience, despite trying offset values greater than 1.2) and also does not alleviate the problem of wasted samples.

Algorithm 1 Sample \( \omega_{in} \) (previous)

\[ \omega_{in} \leftarrow \text{Sample} \left( |\omega_i \cdot \omega_{o}| D(\omega_{in}) \right) \]

if \( U < F(\omega_i, \omega_{in}) \) then

\[ \omega_o \leftarrow \text{reflect}(\omega_i, \omega_{in}) \]

else

\[ \omega_o \leftarrow \text{transmit}(\omega_i, \omega_{in}) \]

end if

Algorithm 2 Sample \( \omega_{in} \) (our)

\[ \omega_{in} \leftarrow \text{Sample} \left( D(\omega_{in}) \right) \quad \circ \text{Algorithm 3 or 4} \]

if \( U < F(\omega_i, \omega_{in}) \) then

\[ \omega_o \leftarrow \text{reflect}(\omega_i, \omega_{in}) \]

else

\[ \omega_o \leftarrow \text{transmit}(\omega_i, \omega_{in}) \]

end if
Table 2: Properties of the importance sampling techniques.

<table>
<thead>
<tr>
<th></th>
<th>previous</th>
<th>our</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDF(ω₀₀)</td>
<td>(ω₀₀ - ω₀)D(ω₀₀)</td>
<td>D₂₀(ω₀₀)</td>
</tr>
<tr>
<td>weight(ω₀)</td>
<td>(ω₀)(ω₀ - ω₀)D(ω₀₀)</td>
<td>G₁(ω₀, ω₀₀)</td>
</tr>
<tr>
<td>V-cavities G</td>
<td>weight(ω₀) ∈ [0, 2]</td>
<td>weight(ω₀) ∈ [0, 1]</td>
</tr>
<tr>
<td>Smith G</td>
<td>weight(ω₀) ∈ [0, ∞)</td>
<td>weight(ω₀) ∈ [0, 1]</td>
</tr>
</tbody>
</table>

Table 2 compares the previous importance sampling method with our method for the V-cavity and the Smith microsurface models. Since G₁ ≤ G₂ the weight in our method is always less than 1: firefly artifacts do not appear (see Figure 2). This can be observed in Figure 4. We also see that our method does not waste the sampling space like previous methods. This is because we never generate backfacing normals.

Sections 4 and 5 are dedicated to implement method Sample D₂₀(ω₀₀) for the V-cavity and the Smith microsurface models, respectively.

4. Sampling D₂₀(ω₀₀) with the V-cavity Model

Importance sampling D₂₀(ω₀₀) with the V-cavity model is summarized in Algorithm 3. We start by sampling from the PDF (ω₀₀ - ω₀)D(ω₀₀) as in the previous method. This gives us a random microsurface with dual normals ω₀₀ and ω₀₀′ (see Figure 11). Then, we choose randomly between ω₀₀ or ω₀₀′ using selection probabilities proportional to (ω₀₀ - ω₀) and (ω₀₀ - ω₀′), respectively, which correspond to their projected areas toward ω₀. This random choice, weighted by the projected area, is what transforms the PDF (ω₀₀ - ω₀)D(ω₀₀) into D₂₀(ω₀₀). In practice, we implement this by a random swap of ω₀₀ by ω₀₀′ with probability (ω₀₀ - ω₀) / (ω₀₀ - ω₀′).

Algorithm 3 Sampling D₂₀(ω₀₀) (V-cavity model)

```
ω₀₀ ← Sample (ω₀₀ - ω₀)D(ω₀₀)  \( \triangleright 1. \) choose microsurface

if \( U < \frac{(ω₀₀ - ω₀)(∞ - ω₀)}{(ω₀₀ - ω₀')(∞ - ω₀')} \)  \( \triangleright 2. \) choose normal

ω₀₀ ← ω₀₀′

end if
```

Figure 5 illustrates how the algorithm works. When the incident ray approaches the geometric surface, it starts by choosing a V-cavity microsurface. Since the V-cavity model assumes separate microsurfaces, they have all the same visibility and can be chosen with the view-independent PDF (ω₀₀ - ω₀)D(ω₀₀). Then, on the chosen V-cavity microsurface, the ray must choose between ω₀₀ and ω₀₀′ with a probability weighted by their respective visibilities. We incorporate this second choice in the final PDF with a random swap between the two normals. Thus, our algorithm requires an additional random number \( U \). The formal proof for this algorithm is given in Appendix B.
Algorithm 4. The distributions of visible normals and normal \( \omega_i = (\alpha_i, \beta_i, \gamma_i) \) where the normalization factor in slope space is:

\[
\frac{1}{c} \int \langle \omega_i, \omega_i \rangle D(\omega_i) d\omega_i
\]

where \( c = \int \langle \omega_i, \omega_i \rangle D(\omega_i) d\omega_i \) is the normalization factor of the PDF. This normalization factor does not appear in Equation (2) but is implicitly contained in the masking function \( G_1 \) in the following, we use Heitz’s [Hei14] formulation of Equation (5) in slope space. We recall that \( \tilde{m} = (x_h, y_h) = (-\frac{x}{m}, -\frac{y}{m}) \) are the slopes associated to the normal \( \omega_m = (x, y, z) \) and reciprocally \( \omega_m = \left( \frac{-x}{\sqrt{x^2 + y^2 + z^2}}, \frac{-y}{\sqrt{x^2 + y^2 + z^2}} \right) \).

If the microsurface is a heightfield, the distribution of normals is linked to the distribution of slopes \( P_{\omega}^{22} (x_h, y_h) \) by

\[
\langle \omega_m, \omega_m \rangle D(\omega_m) = \frac{\partial \tilde{m}}{\partial \omega_m} P_{\tilde{m}}^{22} (\tilde{m} = (x_h, y_h)),
\]

where \( \frac{\partial \tilde{m}}{\partial \omega_m} \) is the Jacobian of the slope to normal transformation. The distributions of visible normals and of visible slopes are linked in the same way by

\[
D_{\omega_m} (\omega_m) = \frac{\partial \omega_m}{\partial \tilde{m}} P_{\tilde{m}}^{22} (\tilde{m} = (x_h, y_h)).
\]

In slope space, Equation (5) is written:

\[
P_{\omega_m}^{22} (x_h, y_h) = \frac{1}{c} \int \chi' (x_h x_m + y_h y_m + z_h) P_{\omega_m}^{22} (x_h, y_h),
\]

where the normalization factor in slope space is:

\[
c = \int \chi' (x_h x_m + y_h y_m + z_h) P_{\omega_m}^{22} (x_h, y_h) d\omega_m.
\]

\[
\text{If we sample } P_{\omega_m}^{22} \text{ and transform the output slope into a normal } \omega_m, \text{ the resulting PDF is } D_{\omega_m}. \text{ This is the first idea of Algorithm 4.}
\]

5.2. Stretch Invariance

Stretching a microsurface is illustrated in Figure 6. The slopes of the microsurface are divided by the stretching coefficients: \( x_h, y_h \) becomes \( \left( \frac{x_h}{\alpha}, \frac{y_h}{\alpha} \right) \). The stretching operator \( S_{\lambda, \nu} \) that transforms the normalized incident vector \( \omega_i = (\alpha_i, \beta_i, \gamma_i) \) into another normalized incident vector \( \omega_o = (\alpha_i, \beta_i, \gamma_i) \) is defined as

\[
S_{\lambda, \nu} \omega_i = \frac{(\lambda \alpha_i, \lambda \beta_i, \lambda \gamma_i)}{\sqrt{(\lambda \alpha_i)^2 + (\lambda \beta_i)^2 + (\lambda \gamma_i)^2}}.
\]

Heitz [Hei14] calls \( P_{\omega_m}^{22} \) shape-invariant if it has the same shape after stretching, i.e. if for any \( \lambda, \nu > 0 \)

\[
P_{\omega_m}^{22} (x_h, y_h) = \frac{1}{\lambda \nu} P_{\omega_m}^{22} \left( \frac{x_h}{\lambda}, \frac{y_h}{\nu}, \alpha, \beta, \gamma \right),
\]

where \( P_{\omega_m}^{22} \) is written \( P_{\omega_m}^{22} (x_h, y_h, \alpha, \beta, \gamma) \) to specify the roughness parameters explicitly. This is usually the case when the distribution comes from a distribution of slopes. Heitz showed that if \( P_{\omega_m}^{22} \) is shape-invariant, then the masking function \( G_1 \) is stretch-invariant. We introduce a stronger property: if the distribution of slopes \( P_{\omega_m}^{22} \) is shape-invariant, then the distribution of visible slopes \( P_{\omega_m}^{22} \) is also shape-invariant. This is written

\[
P_{\omega_m}^{22} (x_h, y_h) = \frac{1}{\lambda \alpha} P_{\omega_m}^{22} \left( \frac{x_h}{\lambda \alpha}, \frac{y_h}{\alpha}, \alpha, \beta, \gamma \right),
\]

and is illustrated in Figure 7. The formal proof of this result is provided in Appendix A. A practical consequence of Equation (9) is that the distribution of visible slopes can always be expressed with an isotropic roughness \( \alpha \) (we choose arbitrarily \( \alpha = 1 \)):

\[
P_{\omega_m}^{22} (x_h, y_h, \alpha, \beta, \gamma) = \frac{1}{\alpha} P_{\omega_m}^{22} \left( \frac{x_h}{\alpha}, \frac{y_h}{\alpha}, \alpha, \beta, \gamma \right).
\]

The factor \( 1/\alpha \) in the right-hand side is the Jacobian of the transformation from \( \left( \frac{x_h}{\alpha}, \frac{y_h}{\alpha} \right) \) to \( (x_h, y_h) \).

\[\rightarrow\] If a sampling algorithm is available for \( ( \frac{x_h}{\alpha}, \frac{y_h}{\alpha} ) \) with PDF \( P_{\omega_m}^{22} (\omega_m) \left( \frac{x_h}{\alpha}, \frac{y_h}{\alpha}, 1, 1 \right) \), we can use it to sample \( (x_h, y_h) \) with PDF \( P_{\omega_m}^{22} (x_h, y_h, \alpha, \beta, \gamma) \). This is the second idea of Algorithm 4.

We note that a similar observation was made by Becker and Max [BM93]: they show that the distribution of visible normals is invariant to height-stretching, which is equivalent to scaling the slope amplitude \( ||\tilde{m}|| \), where the slope coordinates \( x_h \) and \( y_h \) are scaled by the same factor. This allows for isotropic roughness changes only. Instead of height-stretching we use slope-stretching, where \( x_h \) and \( y_h \) can be scaled separately, which permits fully anisotropic roughness changes.

\[\dagger\] Note that the normals are not modified in the same way because they are not vectors but covectors. The normals after stretching are given by the new slopes: \( \omega_m = \frac{(-x - \frac{x_h}{\alpha}, -y - \frac{y_h}{\alpha}, 1)}{\sqrt{(\frac{x}{\alpha})^2 + (\frac{y}{\alpha})^2 + 1}} \).
5.3. Sampling $P_{\alpha\theta}^{-}(x_{\alpha\theta}, 1, 1)$

After stretching the configuration, the distribution of visible slopes is $P_{\alpha\theta}^{22}(x_{\alpha\theta}, y_{\alpha\theta}, 1, 1)$. In order to sample $(x_{\alpha\theta}, y_{\alpha\theta})$ with PDF $P_{\alpha\theta}^{22}(x_{\alpha\theta}, y_{\alpha\theta}, 1, 1)$, we first assume that the frame is such as $\theta_{0} = (x_{0}, 0, z_{0})$. This is possible since we are now using an isotropic distribution of normals ($\alpha_{x}, \alpha_{y} = 1$). The expression of Equation (7) simplifies to

$$P_{\alpha\theta}^{22}(x_{\alpha\theta}, y_{\alpha\theta}, 1, 1) = \frac{1}{c} \chi^{1}(x_{\alpha\theta} + z_{\theta}) (x_{\alpha\theta} + z_{\theta}) P_{\alpha\theta}^{22}(x_{\alpha\theta}, y_{\alpha\theta}).$$

This expression emphasizes that the projection factor $\chi^{1}(x_{\alpha\theta} + z_{\theta}) (x_{\alpha\theta} + z_{\theta})$ is not related to the slope component $y_{\alpha\theta}$. We use this property to sample $P_{\alpha\theta}^{22}$ in two steps a. and b.

5.3.1. a. Sampling $P_{\alpha\theta}^{22-}(x_{\alpha\theta}, 1, 1)$

We start by sampling the marginalized random variable $x_{\alpha\theta}$ given by the PDF

$$P_{\alpha\theta}^{22-}(x_{\alpha\theta}, 1, 1) = \frac{1}{c} \int_{-\infty}^{+\infty} \chi^{1}(x_{\alpha\theta} + z_{\theta}) (x_{\alpha\theta} + z_{\theta}) P_{\alpha\theta}^{22}(x_{\alpha\theta}, y_{\alpha\theta}, 1, 1) dy_{\alpha\theta}$$

$$= \frac{1}{c} \chi^{1}(x_{\alpha\theta} + z_{\theta}) (x_{\alpha\theta} + z_{\theta}) \int_{-\infty}^{+\infty} P_{\alpha\theta}^{22}(x_{\alpha\theta}, y_{\alpha\theta}, 1, 1) dy_{\alpha\theta}$$

$$= \frac{1}{c} \chi^{1}(x_{\alpha\theta} + z_{\theta}) (x_{\alpha\theta} + z_{\theta}) P^{22-}(x_{\alpha\theta}, 1, 1).$$

(10)

where $P^{22-}(x_{\alpha\theta}, 1, 1)$ is the marginalized PDF of $x_{\alpha\theta}$. We invert

$$U_{1} = \int_{-\infty}^{x_{\alpha\theta}} P_{\alpha\theta}^{22-}(x_{\alpha\theta}, 1, 1) dx_{\alpha\theta}$$

$$\Rightarrow x_{\alpha\theta} = C_{\alpha\theta}^{-1}(U_{1}, 1, 1),$$

(11)

where $C_{\alpha\theta}$ is the cumulative distribution function (CDF) associated to the marginalized PDF $P_{\alpha\theta}^{22-}$.

This equation can be solved analytically for Beckmann and GGX distributions (see our supplemental material). For other distributions, if an analytic solution is not available, we propose to precompute the solutions and store them in a 2D table $x_{\alpha\theta} = T_{\alpha\theta}(\theta_{0}, U_{1})$. Indeed, $x_{\alpha\theta}$ depends only on two variables: $\theta_{0}$ (because $x = \sin \theta_{0}$ and $z_{\theta} = \cos \theta_{0}$) and $U_{1}$.

5.3.2. b. Sampling $P_{\alpha\theta}^{22}(y_{\alpha\theta}|x_{\alpha\theta}, 1, 1)$

Then, we sample the conditional random variable $y_{\alpha\theta}|x_{\alpha\theta}$ given by the PDF

$$P_{\alpha\theta}^{22|1}(y_{\alpha\theta}|x_{\alpha\theta}, 1, 1) = \frac{1}{x_{\alpha\theta}} \chi^{1}(x_{\alpha\theta} + z_{\theta}) (x_{\alpha\theta} + z_{\theta}) P_{\alpha\theta}^{22}(x_{\alpha\theta}, y_{\alpha\theta}, 1, 1)$$

$$= \frac{1}{x_{\alpha\theta}} \chi^{1}(x_{\alpha\theta} + z_{\theta}) (x_{\alpha\theta} + z_{\theta}) P^{22}(x_{\alpha\theta}, y_{\alpha\theta}, 1, 1) dy_{\alpha\theta}$$

(12)

This shows that the sampling of $y_{\alpha\theta}$ does not depend on the incident direction $\theta_{0}$. Hence, we invert

$$U_{2} = \int_{-\infty}^{y_{\alpha\theta}} P_{\alpha\theta}^{22}(y_{\alpha\theta}|x_{\alpha\theta}, 1, 1) dy_{\alpha\theta}$$

$$\Rightarrow y_{\alpha\theta} = C_{\alpha\theta}^{2|1}(U_{2}|x_{\alpha\theta}, 1, 1),$$

(13)

where $C_{\alpha\theta}^{2|1}(y_{\alpha\theta}|x_{\alpha\theta}, 1, 1)$ the CDF associated to the view-independent conditional PDF $P_{\alpha\theta}^{22}$. This equation can be solved analytically for Beckmann and GGX distributions (see our supplemental material). For other distributions, if an analytic solution is not available, we propose to precompute the solutions and store them in a 2D table $y_{\alpha\theta} = T_{\alpha\theta}(x_{\alpha\theta}, U_{2})$. Indeed, $y_{\alpha\theta}|x_{\alpha\theta}$ depends only on two variables: $x_{\alpha\theta}$ and $U_{2}$. 

© 2014 The Author(s)

Computer Graphics Forum © 2014 The Eurographics Association and John Wiley & Sons Ltd.
5.4. Sampling $D_{\omega m}(\alpha m)$

In Algorithm 4, we sample $(x_\omega, y_\omega)$ with PDF $P^{22}_{\omega m}(x_\omega, y_\omega, 1, 1)$ by sampling $x_\omega$ with marginalized PDF $P^{22}_{\omega m}(x_\omega, 1, 1)$ and $y_\omega$ with conditional PDF $P^{22}_{\omega m}(y_\omega|x_\omega, 1, 1)$. Then, we transform this slope vector back to the canonical frame where $\omega = (\cos \phi_1 \sin \theta, \sin \phi_1 \sin \theta, \cos \theta)$ by applying a rotation of angle $\phi$ and we unstretch the slopes by multiplying by respectively $\alpha_x$ and $\alpha_y$. The sampled PDF at this point is $P^{22}_{\omega m}(x_\omega, y_\omega, \alpha_x, \alpha_y)$. Finally, we transform the slopes into a normal. The sampled PDF is $D_{\omega m}(0\omega m, \alpha x, \alpha y)$.

**Algorithm 4** Sample $D_{\omega m}(\alpha m)$ (Smith profile)

1. stretch
   $x_\omega \leftarrow C^{22}_{\omega m}^{-1}(U_1, 1, 1)$
2. a. sample $P^{22}_{\omega m}(x_\omega, 1, 1)$
   $y_\omega \leftarrow C^{22}_{\omega m}^{-1}(U_2|x_\omega, 1, 1)$
3. rotate
   $x_\omega \leftarrow \frac{\cos \phi_1}{\sin \phi_1} (x_\omega)
   y_\omega \leftarrow \frac{\sin \phi_1}{\cos \phi_1} (y_\omega)$
4. unstretch
   $\omega \leftarrow \frac{x_\omega}{\sqrt{x_\omega^2 + y_\omega^2}}$ and $\omega \leftarrow \frac{y_\omega}{\sqrt{x_\omega^2 + y_\omega^2}}$

**Figure 8:** Illustration of Algorithm 4.

Analytic Computation of $c^{2-1}_{\omega m}$ and $c^{22-1}_{\omega m}$

The cornerstone of Algorithm 4 is the inversion of the CDFs $C^{22-1}_{\omega m}$ and $C^{22}_{\omega m}$. We found this to be possible analytically for Beckmann and GGX distributions, for which we provide the detailed derivations of the inversion procedures and their C++ implementations in the associated supplemental material.

Precomputation of $c^{2-1}_{\omega m}$ and $c^{22-1}_{\omega m}$

If an analytic inversion procedure of the CDFs $C^{22}_{\omega m}$ and $C^{22}_{\omega m}$ is not available, we propose to use precomputed data $T_{\omega m}[\theta, \Gamma]$ and $T_{\omega m}[x_\omega, y_\omega, \Gamma]$ parametrized in slope space. One advantage of the slope space is that the equation of $P^{22}_{\omega m}$ becomes separable: $X \{x_\omega, y_\omega\} \{x_\omega, y_\omega\} \{y_\omega\}$ does not depend on $y_\omega$ and thus can be pulled out one of the two integrals. This property is used in Equations (10) and (12) and makes the precomputations easier than they would have been with spherical coordinates. The problem of using spherical coordinates was discussed by Becker and Max [BM93]: they store tables parametrized in $(\theta, \phi)$ and they either assume that $\theta$ and $\phi$ can be sampled independently, which is approximate and leads to biased sampling, or they need to store a 3D table to sample $\theta$ for a given $\phi$ (or the reciprocal). This would require an enormous 3D table and would be impractical (>1GB). However, in slope space, Equation (12) shows that $P^{22}_{\omega m}(y_\omega|x_\omega, 1, 1) = P^{22}_{\omega m}(y_\omega|x_\omega, 1, 1)$ and the inversion does not depend on $\theta_1$, which allows us to use a 2D precomputed table indexed by $x_\omega$ and $U_2$. Thus, we need to precompute only two 2D tables $T_{\omega m}[\theta, \Gamma]$ and $T_{\omega m}[x_\omega, \Gamma]$. More importantly, thanks to the invariance property in slope space, we can use the same data with different roughnesses and anisotropy. In our implementation, $T_{\omega m}$ and $T_{\omega m}$ are both of size $1024^2$. Popular heavy-tailed distributions of slopes like GGX have very long tails but also a very sharp peak at the same time, and so require a 1024$^2$ table to capture them well. The total memory footprint for every parametric anisotropic BSDF with the same shape is then 32MB.

Limitation

Note that this technique works only when the distribution of normals $D$ is built from a shape-invariant distribution of slopes $P^{22}_{\omega m}$. For instance, this is the case for the Beckmann [SM67] and GGX [WMLT07], but not for the Phong distribution (which lacks a Smith masking function [WMLT07]).

6. Results and Validation

In Figure 1 we show a dielectric glass plate with anisotropic GGX roughness (Smith masking) on all faces. The illuminant is constant uniform distant white radiance from all directions (a white environment map). In all renders we use forward path tracing with unbounded path lengths. Multiple importance sampling (MIS) mixes BSDF and next-event-estimation (light sampling) at each path vertex. Figure 1 uses 64 independent samples per pixel. Our new BSDF sampling scheme runs in roughly the same time as prior work and provides a significant reduction in variance. Figure 9 shows a configuration where fireflies are still present with the previous method even with 4000 samples per pixel. Due to the default Gaussian filter in Mitsuba, a single bright sample may contribute significantly to several pixels (appearing as a 2x2 bright pixel block).

Figure 10 compares the convergence of importance sampling with the previous method to our method in a configuration shown in Figure 2, i.e. a rough conductor lit by an environment map. The plot shows that both methods converge to the same value, which was expected because the previous method and our method are both mathematically well-defined. However, the convergence rate is not the same.

In the two first rows we used the Smith model. We see that the convergence rate is very sensitive to the fact that the sampling weight can be arbitrary high with the previous method. This produces firefly artifacts that take a long time...
to disappear. In contrary, our method does not produce these artifacts since all BSDF sample weights are bounded to 1. In the two last rows we compare the methods used with the V-cavity model. While the previous method does not suffer from very high sampling weights (they are bounded to 2), the convergence of our method is faster. This is because we do not waste the sampling space by generating backfacing normals with weight 0.

More results and more convergence comparisons are available in the associated supplemental material.

Previous (Walter’s trick)  Our

4000spp, 11.2min  3844spp, 11.5min

Figure 9: A dielectric glass plate \((n = 1.5)\) with isotropic Beckmann roughness \((\alpha = 0.1)\) on all faces (with the Smith masking function). Fireflies are persistent with the previous method even with a large number of samples (4000 per pixel).

7. Conclusion

We have presented a new importance sampling scheme for microfacet-based BSDFs. The scope of our scheme is broad, encompassing many popular isotropic and anisotropic analytic BRDF and BSDF models and allows a variety of normal distribution functions. The performance of our scheme is comparable to previous methods and is suitable for both offline and GPU rendering. We have shown that importance sampling the BSDF using the distribution of visible normals rather than the distribution of normals is practical and makes convergence faster, significantly reducing artifacts at grazing incidence angles, especially for significant roughness levels. With the V-cavity microsurface model, this is implemented with a simple one conditional swap of the normal. With the much more accurate Smith microsurface model, we provide analytic sampling algorithms for Beckmann and GGX distributions in our supplemental material. Otherwise, we have shown that a small set of precomputed data can be used to sample BSDFs with any roughness and anisotropy, due to key shape invariance properties of the distribution of visible slopes.

![Importance Sampling - Smith GGX (\(\alpha = 0.05\))](image1)

![Multiple Importance Sampling - Smith GGX (\(\alpha = 0.05\))](image2)

![Importance Sampling - V-cavity GGX (\(\alpha = 0.05\))](image3)

![Multiple Importance Sampling - V-cavity GGX (\(\alpha = 0.05\))](image4)

Figure 10: Comparisons of our importance sampling technique to previous methods (with and without MIS). The images are computed with 8 samples per pixel. The plots shows the convergence of several running estimates of the same pixel of the pedestal for both techniques. The pixel variance is seen directly in the spread of the various estimates.

References


[CT82] Cook R. L., Torrance K. E.: A reflectance model for
Appendix A: Shape Invariance of $P_{\omega m}^{22}$

We demonstrate the result of Equation (9): 

\[
P_{\omega m}^{22}(\omega_m, \omega_a, \omega_b) = \int \int (\omega_m \cdot \omega_a + \omega_m \cdot \omega_b + \omega_a \cdot \omega_b) \frac{P(\omega_m, \omega_a, \omega_b)}{V_{\omega_m} V_{\omega_a} V_{\omega_b}} d\omega_a d\omega_b
\]

\[
= \int \int (\omega_m \cdot \omega_a + \omega_m \cdot \omega_b + \omega_a \cdot \omega_b) \frac{P(\omega_m, \omega_a, \omega_b)}{V_{\omega_m} V_{\omega_a} V_{\omega_b}} \frac{V_{\omega_a} V_{\omega_b}}{V_{\omega_a} V_{\omega_b}} d\omega_a d\omega_b
\]

\[
= \frac{1}{V_{\omega_m}} \int \int \frac{P(\omega_m, \omega_a, \omega_b)}{V_{\omega_a} V_{\omega_b}} d\omega_a d\omega_b
\]

where we start to write Equation (7). In the second step we develop by applying Equation (8) and by multiplying the numerator and the denominator by

\[
\frac{1}{V_{\omega_a} V_{\omega_b}}
\]

In the third step, at the denominator we make the change of variable:

\[
\frac{1}{k_x k_y} d\omega_{x,y} = d\omega_x \frac{1}{k_x}
\]

Appendix B: Proof for Algorithm 3

To verify that Algorithm 3 samples PDF $D_{\omega_m}(\omega_m)$, we will expand $G_1$ from Equation (3) in the definition of $D_{\omega_m}(\omega_m)$ from Equation (2). The min in Equation (3) is used to differentiate between 2 cases illustrated in Figure 11.

Case (a) If $\omega_{m'}$ is back-facing, then the shadowing occurring on $\omega_{m}$ is $G_1(\omega_{m}, \omega_{m'}) = 2 \langle 0_{\omega_{m}}, 0_{\omega_{m'}} \rangle \frac{|\omega_{m} \cdot 0_{\omega_{m}}|}{|\omega_{m'} \cdot 0_{\omega_{m'}}|}$, and by using the fact that $\langle 0_{\omega_{m}}, 0_{\omega_{m'}} \rangle = 0$ and $|\omega_{m} \cdot 0_{\omega_{m}}| > 0$ we get

\[
D_{\omega_m}(\omega_m) = 2 \frac{\langle 0_{\omega_m}, 0_{\omega_m} \rangle}{|\omega_{m'} \cdot 0_{\omega_{m'}}|} \frac{|\omega_{m} \cdot 0_{\omega_{m}}|}{|\omega_{m'} \cdot 0_{\omega_{m'}}|} D(\omega_m)
\]

\[
= 2 \frac{\langle 0_{\omega_m}, 0_{\omega_m} \rangle}{|\omega_{m'} \cdot 0_{\omega_{m'}}|} \frac{|\omega_{m} \cdot 0_{\omega_{m}}|}{|\omega_{m'} \cdot 0_{\omega_{m'}}|} D(\omega_m)
\]

\[
= 2 \frac{\langle 0_{\omega_m}, 0_{\omega_m} \rangle}{|\omega_{m'} \cdot 0_{\omega_{m'}}|} \frac{|\omega_{m} \cdot 0_{\omega_{m}}|}{|\omega_{m'} \cdot 0_{\omega_{m'}}|} D(\omega_m)
\]

Case (b) If $\omega_{m}$ and $\omega_{m'}$ are both not back-facing for $\omega_{m}$, then there is no masking ($G_1 = 1$) and Equation (2) becomes

\[
D_{\omega_m}(\omega_m) = \frac{\langle 0_{\omega_m}, 0_{\omega_m} \rangle}{|\omega_{m} \cdot 0_{\omega_{m}}|} D(\omega_m)
\]

\[
= 2 \frac{\langle 0_{\omega_m}, 0_{\omega_m} \rangle}{|\omega_{m} \cdot 0_{\omega_{m}}|} \frac{|\omega_{m} \cdot 0_{\omega_{m}}|}{|\omega_{m} \cdot 0_{\omega_{m}}|} D(\omega_m)
\]

because $|\omega_{m} \cdot 0_{\omega_{m}}| = \langle 0_{\omega_m}, 0_{\omega_m} \rangle$.

Conclusion

In either case the distribution of visible normals is given by the same expression. The factor $\langle 0_{\omega_{m}}, 0_{\omega_{m'}} \rangle$ represents the repetition of visibility between $\omega_{m}$ and $\omega_{m'}$. In Algorithm 3, the conditional replacement of $\omega_{m}$ by $\omega_{m'}$ multiplies the original PDF $D_{\omega_m}(\omega_m)$ by the factor $\langle 0_{\omega_{m}}, 0_{\omega_{m'}} \rangle$, and the factor 2 is present in the sampled PDF because $\omega_{m}$ can be chosen for 2 microsurfaces (the one associated to himself and the one to $\omega_{m'}$) and reciprocally for $\omega_{m'}$.

Figure 11: Masking configurations with the V-cavity model.

Either one microfacet is back-facing and completely masked (a) or the two microfacets are completely visible (b).