Bounded VNDF Sampling for the Smith–GGX BRDF

YUSUKE TOKUYOSHI*, Advanced Micro Devices, Inc., Japan
KENTA ETO*, Advanced Micro Devices, Inc., Japan

(a) Previous VNDF sampling (Rendering time: 4.24 s, RMSE: 0.149)
(b) Our bounded VNDF sampling (Rendering time: 4.62 s, RMSE: 0.125)
(c) Reference

Fig. 1. Path tracing using the previous VNDF sampling method [Dupuy and Benyoub 2023] (a) and our bounded VNDF sampling (b) for microfacet-based reflective surfaces (3840×2160 pixels, 64 samples per pixel, AMD Radeon™ RX 7900 XTX GPU). Our method efficiently reduces noise compared to the previous method while increasing the depth of paths, especially for rough surfaces.

Sampling according to a *visible normal distribution function* (VNDF) is often used to sample rays scattered by glossy surfaces, such as the Smith–GGX microfacet model. However, for rough reflections, existing VNDF sampling methods can generate undesirable reflection vectors occluded by the surface. Since these occluded reflection vectors must be rejected, VNDF sampling is inefficient for rough reflections. This paper introduces an unbiased method to reduce the number of rejected samples for Smith–GGX VNDF sampling. Our method limits the sampling range for a state-of-the-art VNDF sampling method that uses a spherical cap-based sampling range. By using our method, we can reduce the variance for highly rough and low-anisotropy surfaces. Since our method only modifies the spherical cap range in the existing sampling routine, it is simple and easy to implement.

CCS Concepts: • **Computing methodologies** → **Ray tracing**.

Additional Key Words and Phrases: importance sampling, microfacet BRDF, visible normal distribution

ACM Reference Format:
Yusuke Tokuyoshi and Kenta Eto. 2024. Bounded VNDF Sampling for the Smith–GGX BRDF. *Proc. ACM Comput. Graph. Interact. Tech.* 7, 1 (May 2024), 18 pages. https://doi.org/10.1145/3651291

*These two authors contributed equally to this work.

Authors' addresses: Yusuke Tokuyoshi, Advanced Micro Devices, Inc., Marunouchi Trust Tower Main Bldg. 10F, 1-8-3 Marunouchi, Tokyo, Japan, yusuke.tokuyoshi@amd.com; Kenta Eto, Advanced Micro Devices, Inc., Marunouchi Trust Tower Main Bldg. 10F, 1-8-3 Marunouchi, Tokyo, Japan, Kenta.Eto@amd.com.

1 INTRODUCTION

Hardware ray tracing and Monte Carlo integration (e.g., GPU path tracing) are used for recent interactive applications as well as offline renderers. In real-time applications such as video games, glossy reflections are often rendered by tracing randomly sampled reflection rays with hardware ray tracing [Deligiannis and Schmid 2019] or screen-space ray tracing [Stachowiak 2015]. To render high-quality glossy reflections with a limited number of samples, efficient importance sampling according to a bidirectional reflectance distribution function (BRDF) is vital for those applications.

The microfacet BRDF model [Cook and Torrance 1982] with the Smith microsurface model [1967] and the GGX distribution of microfacet normals [Trowbridge and Reitz 1975; Walter et al. 2007] is one of the de-facto standard glossy reflection models in computer graphics productions [Andersson et al. 2023; Burley 2012; Georgiev et al. 2019; Kulla and Conty 2017; McAuley et al. 2013]. For this reflection model, importance sampling according to a *visible normal distribution function* (VNDF) [Dupuy and Benyoub 2023; Heitz 2018; Heitz and d'Eon 2014] is widely used in modern renderers. It samples a microfacet normal visible to an incoming direction. By using VNDF sampling, we can reduce the variance for highly glossy surfaces.

However, VNDF sampling is not always efficient for rough reflections (Fig. 1a). For example, if the incoming direction is equal to the surface normal and the GGX roughness parameter is 1, the reflection vectors given by VNDF sampling are uniformly distributed on a unit sphere. For opaque surfaces, reflection vectors in the lower hemisphere must be rejected because they are occluded by the surface. Therefore, in this case, the existing VNDF sampling is twice as inefficient as naïve uniform sampling on the upper hemisphere. Although rejecting samples reduces the number of rays to be traced, random rejection can induce code path divergence and potentially decrease the thread occupancy on the GPU. Without using a sophisticated algorithm to avoid such divergence [Laine et al. 2013], rejecting rays due to VNDF sampling can reduce the GPU efficiency as well as increase the variance. Since reflections are often dominant compared to refractions in most scenes (especially for current video games), there is room for practical quality improvements in VNDF sampling.

To improve the rendering efficiency, we introduce an unbiased method to reduce the variance of VNDF-based reflection vector sampling (Fig. 1b). Our method limits the sampling range of visible normals to reduce the number of rejected samples whose reflection vectors are in the lower hemisphere. For the state-of-the-art VNDF sampling method using a spherical cap-based sampling range [Dupuy and Benyoub 2023], we present an analytical bound for this spherical cap. Using our sampling method for rough reflective surfaces, we can improve the image quality with a small overhead. Although our method cannot be applied for stochastic multi-scattering simulation in the microsurface [Bitterli and d'Eon 2022; Cui et al. 2023; Heitz et al. 2016; Wang et al. 2022], our method significantly reduces firefly noise for energy-compensated microfacet BRDFs [Kulla and Conty 2017; Turquin 2019] that are often used in productions to approximate multiple scattering.

Our contributions are as follows.

- We introduce a bounded VNDF sampling range to reduce the variance for rough reflections.
- To efficiently implement our method, we derive a simple analytic bound of the spherical cap-based sampling range for reflections (§ 3).
- We present the probability density function (PDF) for our method supporting backfacing shading normals (§ 4).
- We discuss the application to energy-compensated Smith–GGX BRDFs (§ 5) and demonstrate its effectiveness (§ 6).

2 BACKGROUND

2.1 Related Work

Microfacet BRDFs. Modern physically-based BRDF models use a microfacet theory [Atanasov et al. 2022; Heitz 2014] as the most common framework. This microfacet BRDF was introduced by Torrance, Sparrow [1967], and Cook [1982]. In this BRDF, light scattering during reflection on a rough surface is modeled using a distribution of microscopic specular facets, also called microfacets. The light scattering profile is then affected by the distribution of microfacet normals, called a normal distribution function (NDF), and by the masking and shadowing from the micro geometry. Two common NDFs used in computer graphics are the GGX distribution [Walter et al. 2007] (a.k.a., Trowbridge–Reitz [1975]) and the Beckmann distribution [1963]. The microfacet BRDF satisfies physically-based constraints [Heitz 2014] by using an appropriate microsurface model such as the Smith [1967] and V-cavity [Cook and Torrance 1982] models. The combination of the Smith microsurface model and the GGX NDF provides a closer match to real-world measured materials, and it is nowadays widely used in renderers and game engines [Andersson et al. 2023; Burley 2012; Georgiev et al. 2019; Kulla and Conty 2017; McAuley et al. 2013].

VNDF Sampling. To reduce the variance in Monte Carlo light transport algorithms, Heitz and d'Eon [2014] introduced a VNDF sampling method for the Smith and V-cavity microsurface models. To facilitate this VNDF sampling, their method transforms the microsurface into a space of unit roughness. In a follow-up paper, Heitz [2018] presented an exact solution for the Smith–GGX VNDF sampling routine. For Smith–Beckmann VNDF sampling, Jakob [2014] proposed a numerical approach to improve the quality of stratified sampling. Tokuyoshi [2021] discussed the PDF for VNDF sampling when a shading normal is backfacing to an incoming direction. Recently, Dupuy and Benyoub [2023] introduced a simpler formulation for Smith–GGX VNDF sampling by mapping the VNDF into a spherical cap. Their spherical cap VNDF sampling is faster than Heitz's method, while its sampling quality is equivalent. Our work improves the sampling quality of the spherical cap VNDF for the Smith–GGX microfacet BRDF. While existing VNDF sampling methods are applicable to both reflections and refractions, we introduce a new sampling method specialized for reflections.

Multiple Scattering BRDFs. Multi-scattering effects in the microsurface are often required in production rendering because the original microfacet BRDF has a dark material appearance for rough surfaces due to the assumption of single scattering. For multiple scattering of the Smith model, Monte Carlo random walk with VNDF sampling has been studied to obtain accurate results [Bitterli and d'Eon 2022; Cui et al. 2023; Heitz et al. 2016; Wang et al. 2022]. However, such random walk algorithms require many samples for convergence and induce code path divergence on the GPU. Therefore, instead of such stochastic algorithms, efficient approximations have been used for time-sensitive rendering applications in the industry. Kulla and Conty [2017] compensated for the loss of energy using lookup tables based on Kelemen and Szirmay-Kalos [2001]. Turquin [2019] scaled the single-scattering BRDF to compensate for the energy loss. Such an energy compensation is performed at real-time frame rates and therefore used in game engines [Hillaire and de Rousiers 2023] as well as offline renderers. However, the existing VNDF sampling can produce firefly noise for these energy-compensated BRDFs, especially on rough surfaces. That is because the density of VNDF-based reflection vectors in the upper hemisphere can be insufficient for the BRDFs. Our bounded VNDF sampling reduces the firefly noise for the energy-compensated BRDFs by increasing the sampling density in the upper hemisphere as roughness increases.

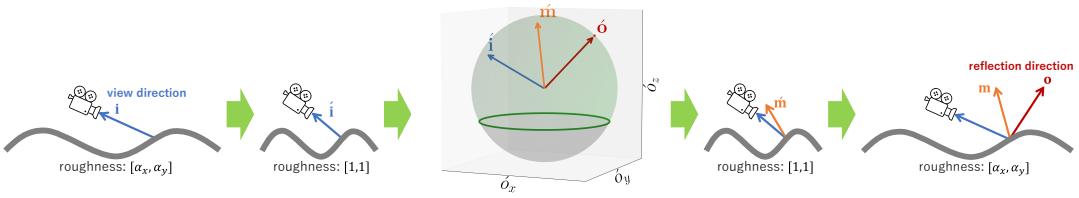


Fig. 2. VNDF sampling using a spherical cap [Dupuy and Benyoub 2023]. This method samples a reflection vector $\acute{\mathbf{o}}$ within the spherical cap (middle) to obtain the visible normal $\acute{\mathbf{m}}$ in the unit-roughness space.

Table 1. Notations used in this paper

Symbol		Description
$[\alpha_x, \alpha_y]$	$\in (0,\infty)^2$	Anisotropic GGX roughness parameter
$\mathbf{i} = [i_x, i_y, i_z]$	$\in \mathrm{S}^2$	Incoming direction in tangent space
$\mathbf{o} = [o_x, o_y, o_z]$	$\in \mathrm{S}^2$	Outgoing direction (i.e., reflection vector) in tangent space
$\mathbf{m} = [m_x, m_y, m_z]$	$\in \mathrm{S}^2$	Microfacet normal in tangent space (i.e., $\mathbf{m} = (\mathbf{i}+\mathbf{o})/\|\mathbf{i}+\mathbf{o}\|$)
$\acute{\mathbf{i}} = [\acute{i}_x, \acute{i}_y, \acute{i}_z]$	$\in \mathrm{S}^2$	Incoming direction in unit-roughness space
$\acute{\mathbf{o}} = [\acute{o}_x, \acute{o}_y, \acute{o}_z]$	$\in \mathrm{S}^2$	Reflection vector in unit-roughness space
$\acute{\mathbf{m}} = [\acute{m}_x, \acute{m}_y, \acute{m}_z]$	$\in \mathrm{S}^2$	Microfacet normal in unit-roughness space (i.e., $\acute{\mathbf{m}} = (\acute{\mathbf{i}}+\acute{\mathbf{o}})/\|\acute{\mathbf{i}}+\acute{\mathbf{o}}\|$)
$[\theta, \phi]$	$\in \mathbb{R}^2$	Polar coordinate of $\mathbf{o}$ (i.e., $\mathbf{o} = [\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta]$)
$p(\mathbf{m})$	$\in [0,\infty)$	PDF for previous VNDF sampling
$p_{\mathrm{our}}(\mathbf{m})$	$\in [0,\infty)$	PDF for our bounded VNDF sampling
$\chi^+(x)$	$\in \{0,1\}$	Heaviside function: 1 if $x > 0$ and 0 if $x \le 0$

2.2 VNDF Sampling Using a Spherical Cap

In this subsection, we briefly review the algorithm of spherical cap VNDF sampling [Dupuy and Benyoub 2023]. Fig. 2 illustrates this algorithm. For notations used in this paper, please see Table 1. The spherical cap VNDF sampling first stretches the tangent space into a space of unit roughness (i.e., $\alpha_x = \alpha_y = 1$) as in Heitz and d'Eon [2014]. In this unit-roughness space, the algorithm samples a reflection vector $\acute{\mathbf{o}}$ for an incoming direction $\acute{\mathbf{i}}$, and then computes the microfacet normal $\acute{\mathbf{m}}$ given by the halfvector of $\acute{\mathbf{i}}$ and $\acute{\mathbf{o}}$. For the Smith–GGX VNDF, $\acute{\mathbf{o}}$ is uniformly distributed on a spherical cap whose center is the surface normal and whose angle is $\arccos(-\acute{i}_z)$. Thus, we uniformly sample $\acute{\mathbf{o}}$ from this spherical cap to obtain a visible microfacet normal $\acute{\mathbf{m}}$ in the unit-roughness space. After getting a visible normal $\acute{\mathbf{m}}$, we unstretch the space to obtain the visible normal $\mathbf{m}$ in the original tangent space. To summarize, this VNDF sampling algorithm is as follows.

(1) Stretch the incoming direction $\mathbf{i}$ to $\acute{\mathbf{i}} = [\alpha_x i_x, \alpha_y i_y, i_z]/\|[\alpha_x i_x, \alpha_y i_y, i_z]\|$.
(2) Sample $\acute{\mathbf{o}}$ on the spherical cap: $\acute{o}_z \in (-\acute{i}_z, 1]$.
(3) Compute the halfvector: $\acute{\mathbf{m}} = (\acute{\mathbf{i}}+\acute{\mathbf{o}})/\|\acute{\mathbf{i}}+\acute{\mathbf{o}}\|$.
(4) Unstretch the microfacet normal: $\mathbf{m} = [\alpha_x \acute{m}_x, \alpha_y \acute{m}_y, \acute{m}_z]/\|[\alpha_x \acute{m}_x, \alpha_y \acute{m}_y, \acute{m}_z]\|$.

After sampling the visible normal $\mathbf{m}$ using the above algorithm, we compute the outgoing direction $\mathbf{o} = 2(\mathbf{i}\cdot\mathbf{m})\mathbf{m} - \mathbf{i}$ for specular reflections. Although this algorithm is simpler than Heitz [2018], it does not improve the sampling quality. In this paper, we reduce the variance for specular reflections by introducing a bounding spherical cap tighter than the previous spherical cap $(-\acute{i}_z, 1]$.

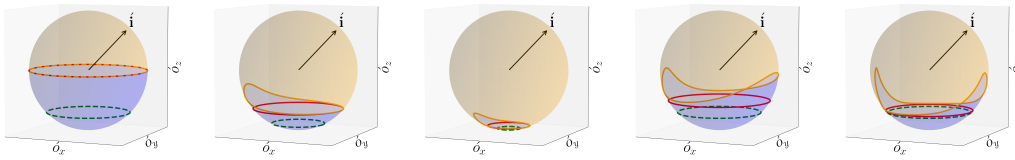


Fig. 3. Previous lower bound (green dashed line) and our lower bound (red line) for the spherical cap in VNDF sampling. The orange and blue regions correspond to reflection vectors in the upper and lower hemispheres, respectively. Since our method bounds the orange region more tightly than the previous method, it reduces the number of reflection vectors occluded by the surface.

3 OUR BOUNDED VNDF SAMPLING

The previous method sampled $\acute{o}_z$ within $(-\acute{i}_z, 1]$. In this paper, we introduce a tighter lower bound on $\acute{o}_z$ than the previous method for reflection vector sampling (Fig. 3). Since our method does not change the algorithm except for this lower bound, it is simple to implement in an existing sampling routine. Listing 1 shows the HLSL code of our VNDF sampling routine.

3.1 Transformation of Reflection Vector Bounds

To obtain the bounding spherical cap for our VNDF-based reflection vector sampling, we first project the bound of reflection vectors $\mathbf{o}$ to $\acute{\mathbf{o}}$ in the unit-roughness space. Then, we derive a simple bounding spherical cap for this projected bound. Since a reflection vector $\mathbf{o}$ should be in the upper hemisphere centered at the surface normal, the bound of reflection vectors is the circle on the tangent plane. In this paper, we write this circle as $[\cos\phi, \sin\phi, 0]$ by using the polar coordinate $[\theta, \phi]$ for $\mathbf{o} = [\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta]$, because $\cos\theta > 0$ for reflections. By transforming this circle from tangent-space $\mathbf{o}$ to unit-roughness space $\acute{\mathbf{o}}$, we obtain a lower bound of $\acute{o}_z$ at ϕ: $\min_{\theta\in[0,\pi/2]} \acute{o}_z(\theta, \phi) = \acute{o}_z(\pi/2, \phi)$, where

$$\acute{o}_z\left(\frac{\pi}{2}, \phi\right) = \left(\frac{(i_x + \cos\phi)^2 + (i_y + \sin\phi)^2 + i_z^2}{\frac{(i_x+\cos\phi)^2}{\alpha_x^2} + \frac{(i_y+\sin\phi)^2}{\alpha_y^2} + i_z^2} - 1\right)\acute{i}_z. \tag{1}$$

For the derivation, please refer to Appendix A. When $\acute{i}_z > 0$, we obtain ϕ that minimizes Eq. 1 from the following equation:

$$\operatorname*{argmin}_{\phi} \acute{o}_z\left(\frac{\pi}{2}, \phi\right) = \operatorname*{argmin}_{\phi} \frac{(i_x + \cos\phi)^2 + (i_y + \sin\phi)^2 + i_z^2}{\frac{(i_x+\cos\phi)^2}{\alpha_x^2} + \frac{(i_y+\sin\phi)^2}{\alpha_y^2} + i_z^2}. \tag{2}$$

Although Eq. 2 is generalized for anisotropic roughness, we find a simple solution for the following special cases:

$$[\cos\phi, \sin\phi] = \begin{cases} [i_x, i_y]/\|[i_x, i_y]\| & \text{if } \alpha_x = \alpha_y < 1 \\ [-i_x, -i_y]/\|[i_x, i_y]\| & \text{if } \alpha_x = \alpha_y > 1 \end{cases}. \tag{3}$$

In the next subsection, we show the derivation for these special case solutions. Then, based on the special case bounds, we introduce a practical loose bound for general anisotropic cases.

3.2 Bounding Spherical Cap

3.2.1 Isotropic Roughness. For isotropic roughness $\alpha = \alpha_x = \alpha_y$, we can simplify Eq. 2 into the following equation:

$$\operatorname*{argmin}_{\phi} \acute{o}_z\left(\frac{\pi}{2}, \phi\right) = \operatorname*{argmin}_{\phi} \frac{r(\phi) + i_z^2}{\frac{r(\phi)}{\alpha} + i_z^2} = \begin{cases} \operatorname{argmax}_{\phi} r(\phi) & \text{if } \alpha < 1 \\ \operatorname{argmin}_{\phi} r(\phi) & \text{if } \alpha > 1 \\ \mathbb{R} & \text{if } \alpha = 1 \end{cases}, \tag{4}$$

where $r(\phi) = (i_x + \cos\phi)^2 + (i_y + \sin\phi)^2$ represents the squared distance between a unit circle $[\cos\phi, \sin\phi]$ and a point $[-i_x, -i_y]$ on the tangent plane. Therefore, the right side of Eq. 4 for $\alpha \neq 1$ is the farthest and closest point on the unit circle from $[-i_x, -i_y]$, which gives Eq. 3. Although we can obtain the lower bound by substituting Eq. 3 into Eq. 1, we derive a simple form of the lower bound by using the furthest and closest distances $\max_\phi r(\phi)$ and $\min_\phi r(\phi)$. These squared distances are simply given by the radius of the unit circle and the norm of $[-i_x, -i_y]$ as follows: $\max_\phi r(\phi) = \left(1 + \sqrt{i_x^2 + i_y^2}\right)^2$ and $\min_\phi r(\phi) = \left(1 - \sqrt{i_x^2 + i_y^2}\right)^2$. Using these distances, we unify the equations for the two cases $\alpha < 1$ and $\alpha > 1$ into one case $\alpha \neq 1$ as follows:

$$\operatorname*{argmin}_{\phi} \acute{o}_z\left(\frac{\pi}{2}, \phi\right) = \begin{cases} \left\{\phi | r\left(\phi\right) = s^2\right\} & \text{if } \alpha \neq 1 \\ \mathbb{R} & \text{if } \alpha = 1 \end{cases}. \tag{5}$$

where $s = 1 + \operatorname{sgn}(1 - \alpha)\sqrt{i_x^2 + i_y^2}$. Therefore, by substituting $\alpha_x = \alpha_y = \alpha$ and $(i_x + \cos\phi)^2 + (i_y + \sin\phi)^2 = s^2$ into Eq. 1, we obtain our lower bound without trigonometric functions as follows:

$$\boxed{\min_{\phi} \acute{o}_z\left(\frac{\pi}{2}, \phi\right) = -k\acute{i}_z}, \quad \text{where } \boxed{k = \frac{(1 - \alpha^2)s^2}{s^2 + \alpha^2 i_z^2}}. \tag{6}$$

This lower bound $-k\acute{i}_z$ is the infimum for reflections if $\acute{i}_z > 0$. Hence, we sample $\acute{o}_z$ within $(-k\acute{i}_z, 1]$ for our VNDF sampling when $\acute{i}_z > 0$. When $\acute{i}_z \leq 0$ (i.e., backfacing shading normal), we use the previous range $(-\acute{i}_z, 1]$ as in Dupuy and Benyoub [2023].

3.2.2 Anisotropic Roughness. Unlike the special case in isotropic roughness, Eq. 2 is not trivial for anisotropic roughness $\alpha_x \neq \alpha_y$. The solution of Eq. 2 is not necessarily Eq. 3 in general, and we could not find the closed-form solution in $\alpha_x \neq \alpha_y$. Therefore, for $\acute{i}_z > 0$, we introduce a loose bound that is obtained by Eq. 6 with setting α conservatively as follows:

$$\alpha = \min(\alpha_x, \alpha_y, 1). \tag{7}$$

When $\alpha_x \leq 1$ or $\alpha_y \leq 1$, this conservative α produces a lower bound slightly looser than the infimum for $\acute{o}_z$. On the other hand, when $\alpha_x > 1$ and $\alpha_y > 1$, the lower bound is limited to zero: $-k\acute{i}_z = 0$, and thus it can be significantly looser than the infimum. However, this bound $-k\acute{i}_z = 0$ is still tighter than the previous lower bound $-\acute{i}_z$. In addition, the case of $\alpha_x > 1$ and $\alpha_y > 1$ is rare in practical scenes because the maximum roughness parameter is often limited to 1 in productions [Burley 2012]. Thus, our approach is practical. Please see the supplementary document for the interactive visualization of our bound.

Listing 1. Our bounded VNDF sampling for reflections. The difference from Dupuy and Benyoub [2023] is written in red.

```
float3 SampleGGXReflection(float3 i, float2 alpha, float2 rand) {
 float3 i_std = normalize(float3(i.xy * alpha, i.z));
 // Sample a spherical cap
 float phi = 2.0f * M_PI * rand.x;
 float a = saturate(min(alpha.x, alpha.y)); // Eq. 7
 float s = 1.0f + length(float2(i.x, i.y)); // Omit sgn for a<=1
 float a2 = a * a; float s2 = s * s;
 float k = (1.0f - a2) * s2 / (s2 + a2 * i.z * i.z); // Eq. 6
 float lower_bound = i.z > 0.0f ? -k * i_std.z : -i_std.z;
 float z = mad(lower_bound, rand.y, 1.0f - rand.y);
 float sin_theta = sqrt(saturate(1.0f - z * z));
 float3 o_std = {sin_theta * cos(phi), sin_theta * sin(phi), z};
 // Compute the microfacet normal m
 float3 m_std = i_std + o_std;
 float3 m = normalize(float3(m_std.xy * alpha, m_std.z));
 // Return the reflection vector o
 return 2.0f * dot(i, m) * m - i;
}
```

Listing 2. Our PDF for reflections: $p_{\text{our}}(\mathbf{m})\|\partial\mathbf{m}/\partial\mathbf{o}\|$ for $i_z \geq 0$. Our contribution is written in red. This implementation omits the Heaviside function $\chi^+(\acute{o}_z + k\acute{i}_z)$, assuming that our method is applied only to reflections. For $i_z < 0$, we calculate the previous PDF $p(\mathbf{m})\|\partial\mathbf{m}/\partial\mathbf{o}\|$ using a numerically stable form (please see Appendix C).

```
float GGXReflectionPDF(float3 i, float3 o, float2 alpha) {
 float3 m = normalize(i + o);
 float ndf = D(m, alpha);
 float2 ai = alpha * i.xy;
 float len2 = dot(ai, ai);
 float t = sqrt(len2 + i.z * i.z);
 if (i.z >= 0.0f) {
  float a = saturate(min(alpha.x, alpha.y)); // Eq. 7
  float s = 1.0f + length(float2(i.x, i.y)); // Omit sgn for a<=1
  float a2 = a * a; float s2 = s * s;
  float k = (1.0f - a2) * s2 / (s2 + a2 * i.z * i.z); // Eq. 6
  return ndf / (2.0f * (k * i.z + t)); // Eq. 10 * ||dm/do||
 }
 // Numerically stable form of the previous PDF for i.z < 0
 return ndf * (t - i.z) / (2.0f * len2); // = Eq. 8 * ||dm/do||
}
```

4 PDF FOR OUR BOUNDED VNDF SAMPLING

The VNDF is the PDF for visible microfacet normals. For the Smith–GGX model, this PDF is given by

$$p(\mathbf{m}) = \frac{2D\left(\mathbf{m}\right)\max\left(\mathbf{i}\cdot\mathbf{m},0\right)}{i_z + \sqrt{\alpha_x^2 i_x^2 + \alpha_y^2 i_y^2 + i_z^2}}, \tag{8}$$

where $D(\mathbf{m})$ is the GGX NDF:

$$D\left(\mathbf{m}\right) = \frac{\chi^+(m_z)}{\pi\alpha_x\alpha_y\left(\frac{m_x^2}{\alpha_x^2} + \frac{m_y^2}{\alpha_y^2} + m_z^2\right)^2}. \tag{9}$$

When $i_z \leq 0$ (i.e., backfacing shading normal), we use this PDF as in previous VNDF sampling [Dupuy and Benyoub 2023; Tokuyoshi 2021]. For $i_z > 0$, we introduce a new PDF for our sampling method. Our method shrinks the range of the spherical cap from $(-\acute{i}_z, 1]$ to $(-k\acute{i}_z, 1]$. Thus, by replacing the spherical cap in the previous PDF with our spherical cap, we obtain the PDF

for our sampling technique:

$$\boxed{p_{\text{our}}(\mathbf{m}) = \frac{2D\left(\mathbf{m}\right)\max\left(\mathbf{i}\cdot\mathbf{m},0\right)}{k i_z + \sqrt{\alpha_x^2 i_x^2 + \alpha_y^2 i_y^2 + i_z^2}}\chi^+\left(k\acute{i}_z + \acute{o}_z\right)}. \tag{10}$$

For derivation, please refer to Appendix B. Since our method limits the sampling range for $\acute{o}_z$ into $(-k\acute{i}_z, 1]$, we represent this limited range using the Heaviside function $\chi^+(k\acute{i}_z + \acute{o}_z)$. By multiplying the above PDF $p_{\text{our}}(\mathbf{m})$ by the Jacobian for the transformation between halfvectors and reflection vectors $\|\partial\mathbf{m}/\partial\mathbf{o}\| = 1/(4|\mathbf{i}\cdot\mathbf{m}|)$, we obtain the PDF for our reflection vector sampling $p_{\text{our}}(\mathbf{m})\|\partial\mathbf{m}/\partial\mathbf{o}\|$. Listing 2 shows the code for our PDF for reflections. In this implementation, we omit the Heaviside function $\chi^+(k\acute{i}_z + \acute{o}_z)$, because we use our sampling method only for opaque surfaces. For opaque surfaces, we get $\chi^+(k\acute{i}_z + \acute{o}_z) = 1$ if the reflection vector $\mathbf{o}$ is in the upper hemisphere. If the reflection vector $\mathbf{o}$ is in the lower hemisphere, the integrand in the lighting integral must be zero. Therefore, omitting the Heaviside function does not affect Monte Carlo integration. In implementation, the only difference between our method and the previous method is the presence of k.

5 APPLICATION TO ENERGY-COMPENSATED BRDFS

Our method is applicable to energy-compensated Smith–GGX BRDFs as well as the single-scattering BRDF model. In this section, we first show that the previous VNDF sampling can produce firefly noise for energy-compensated BRDFs. We then explain how our method reduces this noise.

Single-Scattering BRDF. The single-scattering microfacet BRDF $f_{\text{ss}}(\mathbf{i},\mathbf{o}) \in [0,\infty)$ [Cook and Torrance 1982] is given by the following equation:

$$f_{\text{ss}}(\mathbf{i},\mathbf{o}) = \frac{F(\mathbf{i}\cdot\mathbf{m})G_2(\mathbf{i},\mathbf{o},\mathbf{m})D(\mathbf{m})}{4|\mathbf{i}\cdot\mathbf{n}||\mathbf{o}\cdot\mathbf{n}|}, \tag{11}$$

where $\mathbf{n} = [0,0,1]$ is the surface normal in tangent space, $F(\mathbf{i}\cdot\mathbf{m}) \in [0,1]$ is the Fresnel term, and $G_2(\mathbf{i},\mathbf{o},\mathbf{m}) \in [0,1]$ is the Smith masking-shadowing function [1967]. For this single-scattering BRDF, the weight $w_{\text{ss}}(\mathbf{o})$ for a ray direction $\mathbf{o}$ sampled according to the VNDF (Eq. 8) is equal to or less than 1 [Heitz and d'Eon 2014] having:

$$w_{\text{ss}}(\mathbf{o}) = \frac{f_{\text{ss}}(\mathbf{i},\mathbf{o})|\mathbf{o}\cdot\mathbf{n}|}{p(\mathbf{m})\|\partial\mathbf{m}/\partial\mathbf{o}\|} = \frac{F(\mathbf{i}\cdot\mathbf{m})G_2(\mathbf{i},\mathbf{o},\mathbf{m})}{G_1(\mathbf{i},\mathbf{m})} \leq 1, \tag{12}$$

where $G_1(\mathbf{i},\mathbf{m}) = p(\mathbf{m})/D(\mathbf{m}) \geq G_2(\mathbf{i},\mathbf{o},\mathbf{m})$ which is equal to the Smith masking function [1967] when $i_z \geq 0$. Thus, VNDF sampling does not increase the ray weight during path tracing and thus reduces the variance for the single-scattering BRDF model. However, for rough surfaces, such a single-scattering model introduces the loss of energy (Fig. 4a), because it ignores multiple scattering in the microsurface.

Energy-Compensated BRDF. To compensate for the energy loss for the microfacet BRDF, an approximate multi-scattering term $f_{\text{ms}}(\mathbf{i},\mathbf{o}) \in [0,\infty)$ [Kulla and Conty 2017; Turquin 2019] is often added to the BRDF for production rendering. In this case, the weight for a sample direction is

$$w_{\text{ms}}(\mathbf{o}) = \frac{(f_{\text{ss}}(\mathbf{i},\mathbf{o}) + f_{\text{ms}}(\mathbf{i},\mathbf{o}))|\mathbf{o}\cdot\mathbf{n}|}{p(\mathbf{m})\|\partial\mathbf{m}/\partial\mathbf{o}\|}. \tag{13}$$

Turquin [2019] approximated the multi-scattering term $f_{\text{ms}}(\mathbf{i},\mathbf{o})$ using the following equation:

$$f_{\text{ms}}(\mathbf{i},\mathbf{o}) \approx F_{\text{ms}}\frac{1-E(\mathbf{i})}{E(\mathbf{i})}f_{\text{ss}}(\mathbf{i},\mathbf{o}), \tag{14}$$

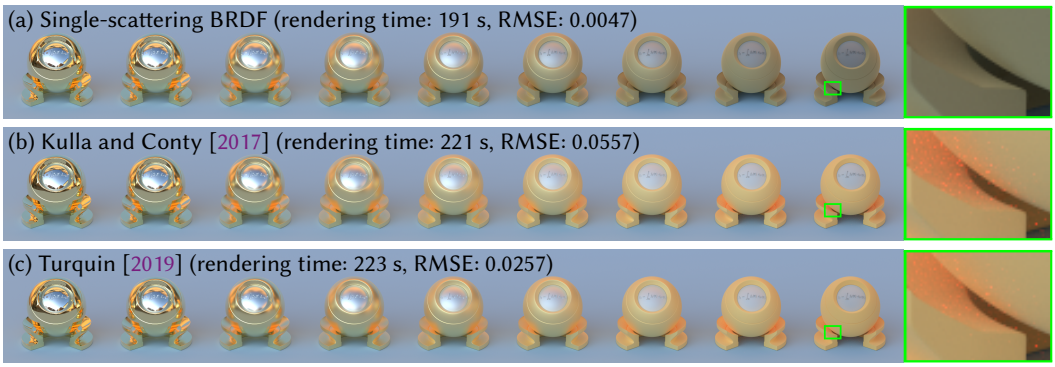
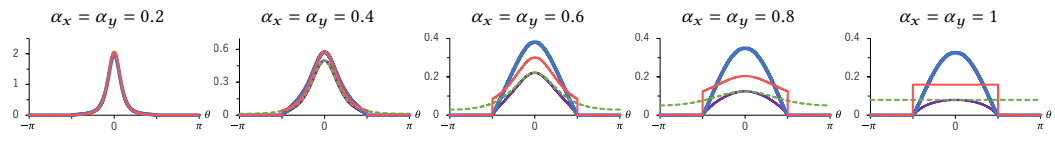
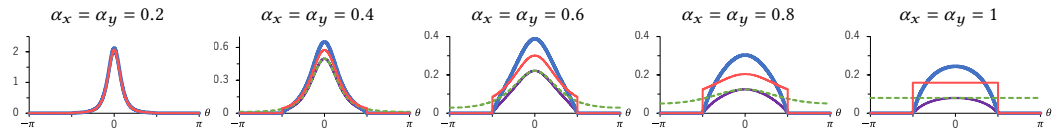


Fig. 4. Single-scattering Smith–GGX BRDF (a) and energy-compensated Smith–GGX BRDFs (b, c) with different roughness for each Orb object. All images (4096×512 pixels, 64×48 pixels for closeup) are rendered using the previous VNDF sampling [Dupuy and Benyoub 2023] with 16384 samples per pixel. Despite using such a large number of samples, there is noticeable firefly noise on rough surfaces for energy-compensated BRDFs.



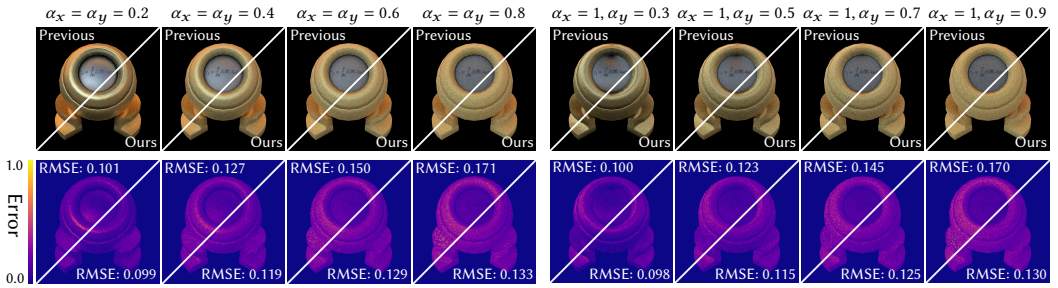
(a) Smith–GGX BRDF with Kulla and Conty's energy compensation [2017]



(b) Smith–GGX BRDF with Turquin's energy compensation [2019]

Fig. 5. Plots of the single-scattering microfacet BRDF lobe $f_{ss}(\mathbf{i},\mathbf{o})|\mathbf{o}\cdot\mathbf{n}|$ (purple line), the energy-compensated BRDF lobe $(f_{ss}(\mathbf{i},\mathbf{o}) + f_{ms}(\mathbf{i},\mathbf{o}))|\mathbf{o}\cdot\mathbf{n}|$ (blue line), the previous PDF $p(\mathbf{m})\|\partial\mathbf{m}/\partial\mathbf{o}\|$ (green dashed line), and our PDF $p_{our}(\mathbf{m})\|\partial\mathbf{m}/\partial\mathbf{o}\|$ (red line) for $\mathbf{i} = [0, 0, 1]$. The Fresnel term is set to $F(\mathbf{i}\cdot\mathbf{m}) = 1$. The horizontal axis is the angle θ of the reflection vector $\mathbf{o}$. While the previous PDF is equal to or higher than the single-scattering lobe, it can be lower than the energy-compensated BRDFs. On the other hand, our PDF is higher than the previous PDF in $-\pi/2 < \theta < \pi/2$ (i.e., upper hemisphere) and closer to the energy-compensated BRDF lobe. Thus, our method reduces the variance (i.e., firefly noise) for the energy-compensated BRDFs.

where $F_{ms} \in [0, 1]$ is the multi-scattering Fresnel term [Jakob et al. 2014], and Turquin [2019] further approximated it with $F_{ms} \approx F(1)$ for conductors. $E(\mathbf{i}) = \int_{S^2} f_{ss}(\mathbf{i},\boldsymbol{\omega})|\boldsymbol{\omega}\cdot\mathbf{n}|/F(\mathbf{i}\cdot\mathbf{m})\mathrm{d}\boldsymbol{\omega}$ is the directional-hemispherical reflectance without the Fresnel term, and it is given by a lookup table [McAuley et al. 2013]. Although the VNDF is approximately proportional to Turquin's reflection model, it can produce high variance that is visible as firefly noise in rendered images (Fig. 4c). This is because the sample weight given by Eq. 13 can be larger than 1. Since the resulting ray weight is the product of sample direction weights at each bounce during path tracing, it can increase exponentially. As shown in Fig. 5, the mismatch between the energy-compensated BRDF and VNDF becomes larger as roughness increases. Thus, firefly noise can be noticeable, especially



(a) Isotropic roughness (b) Anisotropic roughness

Fig. 6. Quality comparison (1024×1024 pixels, 16 spp) between previous VNDF sampling [Dupuy and Benyoub 2023] and our bounded VNDF sampling for isotropic roughness (a) and anisotropic roughness (b). Our sampling reduces RMSE, especially for higher roughness and lower anisotropy.

on rough surfaces. The same problem occurs for Kulla and Conty's energy compensation [2017] as shown in Fig. 4b (please refer to the supplementary document for the detail of Kulla and Conty's compensation).

Firefly Reduction Using Our Method. For reflection vectors in the upper hemisphere, our bounded VNDF sampling generates higher density samples than the previous VNDF sampling (i.e., $p_{\mathrm{our}}(\mathbf{m}) > p(\mathbf{m})$ if $o_z > 0$). Thus, our bounded VNDF sampling produces a smaller sample weight $w_{\mathrm{ms}}^{\mathrm{our}}(\mathbf{o})$ than the previous method (Eq. 13) as follows:

$$w_{\mathrm{ms}}^{\mathrm{our}}(\mathbf{o}) = \frac{(f_{\mathrm{ss}}(\mathbf{i},\mathbf{o}) + f_{\mathrm{ms}}(\mathbf{i},\mathbf{o}))|\mathbf{o}\cdot\mathbf{n}|}{p_{\mathrm{our}}(\mathbf{m})\|\partial\mathbf{m}/\partial\mathbf{o}\|} < w_{\mathrm{ms}}(\mathbf{o}) \quad \text{if } o_z > 0. \tag{15}$$

The difference between $w_{\mathrm{ms}}(\mathbf{o})$ and $w_{\mathrm{ms}}^{\mathrm{our}}(\mathbf{o})$ becomes larger as roughness increases. Therefore, our method reduces the firefly noise caused by multi-scattering approximations for rough surfaces. Although our method cannot completely avoid the increase of the ray weight unlike the single-scattering case, fireflies are significantly reduced in our experiments (Sect. 6).

6 EXPERIMENTAL RESULTS

Here we show images rendered using a HIP RT [AMD 2023]-based path tracer performed on an AMD Radeon™ RX 7900 XTX GPU. All scenes have the Smith–GGX microfacet BRDF with Turquin's energy compensation [2019]. The reference images are rendered with 1 M samples per pixel (spp) for Figs. 4 and 10, and 65536 spp for others (please see the supplementary images). The image quality is evaluated with the root-mean-squared error (RMSE) metric. Although VNDF sampling of 1 spp is often used in combination with a denoiser and upscaler for video games, this section uses a large number of samples without such post-processing to show the difference in noise.

Isotropic Roughness. Fig. 6a shows the quality comparison between our method and the previous method [Dupuy and Benyoub 2023] using different isotropic roughness parameters. In this figure, each Orb object is not tessellated unlike Fig. 4 and has significant backfacing shading normals. Our method supports such backfacing shading normals as in the previous method. Compared to the previous method, our method reduces errors, especially for high-roughness surfaces ($\alpha_x = \alpha_y = 0.8$ in Fig. 6a). On the other hand, the quality difference is insignificant for low roughness ($\alpha_x = \alpha_y = 0.2$ in Fig. 6a). That is because our bounding spherical cap depends on the roughness parameter unlike

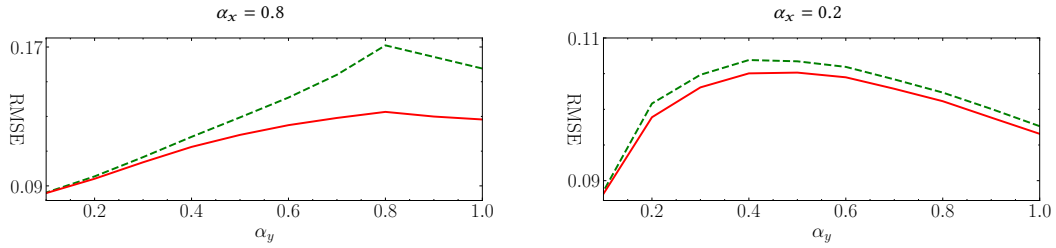


Fig. 7. Plots of RMSE for the previous method [Dupuy and Benyoub 2023] (green dashed line) and our method (red line) using different anisotropic roughness for the Orb model used in Fig. 6

the previous method, and the difference from the previous spherical cap is more significant for a higher roughness parameter.

Anisotropic Roughness. Figs. 6b and 7 show the quality comparison for anisotropic roughness. Since our bounding spherical cap is tighter for lower anisotropy, our method reduces errors, especially for low anisotropy. For highly anisotropic roughness, our bounded VNDF sampling can still generate many reflection vectors in the lower hemisphere, as in the previous method. Therefore, our method has room for improvement in high anisotropy. This improvement is left for future work.

Multiple importance sampling. For practical path tracing, VNDF sampling and next event estimation are often combined using multiple importance sampling (MIS) [Veach and Guibas 1995]. Since MIS compensates for the lack of sampling density, especially in direct illumination, it can mitigate the disadvantage of the previous VNDF sampling for rough surfaces. However, for indirect illumination, MIS with the previous VNDF sampling can still produce noticeable noise on rough surfaces as shown in Fig. 8. Our method reduces the variance more than the previous VNDF sampling, even with MIS (Fig. 9). This variance reduction is effective for an arbitrary number of samples. Therefore, our method is applicable to a wide range of practical applications, from offline rendering to real-time rendering (which uses a 1-spp image as input to a denoiser and upscaler).

Energy-Compensated BRDFs. Fig. 10 shows the effectiveness of our bounded VNDF sampling for the Smith–GGX BRDF with different energy-compensation terms [Kulla and Conty 2017; Turquin 2019]. Compared to the previous VNDF sampling, our method significantly reduces firefly noise caused by interreflections due to rough surfaces. For VNDF sampling, Turquin's model produces less noise than Kulla and Conty's model as discussed by Turquin [2019]. Our bounded VNDF is moreover close to Turquin's BRDF lobe as shown in Fig. 5. Therefore, our method efficiently reduces the firefly noise for energy-compensated BRDFs, especially Turquin's model.

Overhead. Our method has a computational overhead for calculating the lower bound (Eq. 6), but it is only five lines of code in Listing 1 and Listing 2. Table 2 shows this computational overhead on the CPU. For white furnace test [Heitz 2014] on the GPU (which evaluates the BRDF by sampling reflection vectors but not tracing rays), we did not obtain a clear difference in rendering time with and without our lower bound. While the overhead of our sampling routine is negligibly small, the method increases the depth of light paths by reducing the number of rejected sample rays. Thus, it can increase the rendering time for path tracing (Table 3), though the increased time is small compared to the quality improvement. On the other hand, reducing the random rejection reduces

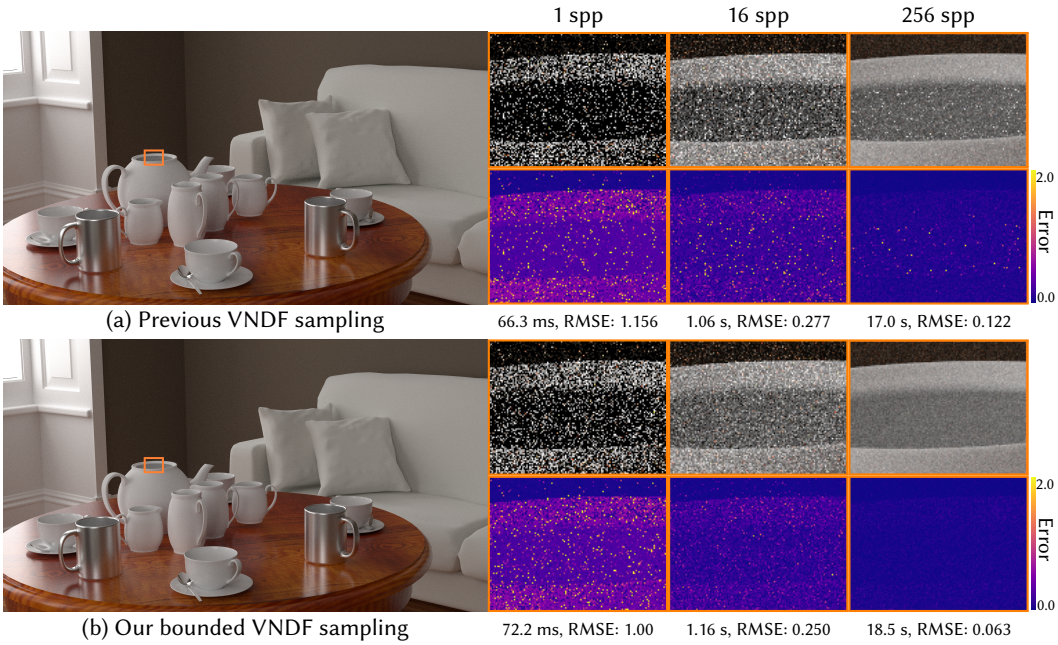
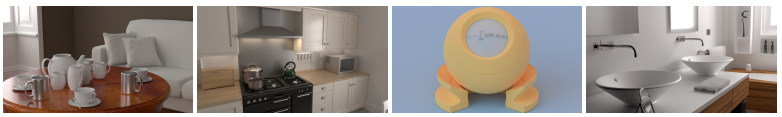


Fig. 8. Path tracing using multiple importance sampling (MIS) of VNDF sampling and next event estimation for the Fireplace Room scene (3840×2160 pixels). Although MIS can reduce the variance for low-frequency BRDFs, (a) combination with the previous VNDF sampling [Dupuy and Benyoub 2023] still produces noticeable noise on rough surfaces. (b) Combination with our bounded VNDF sampling significantly reduces this noise.

Table 2. VNDF sampling time on an AMD Ryzen™ 9 7950X CPU.

	Dupuy and Benyoub [2023]	Our method	Difference
Sampling routine (Listing 1)	8.32 ns	10.43 ns	2.11 ns
PDF (Listing 2)	5.00 ns	6.57 ns	1.57 ns

Table 3. Increase in rendering time due to the path depth increased by our method. When limiting the maximum path depth to a few bounces (a–c), the rate of increase is significantly smaller than in a high limit (d).



	Fireplace Room	Kitchen	Orb	Salle de bain
(a) Up to 1 bounce	1.33 %	0.23 %	0.86 %	0.39 %
(b) Up to 2 bounces	2.91 %	0.31 %	2.93 %	0.85 %
(c) Up to 4 bounces	4.72 %	0.46 %	6.19 %	1.23 %
(d) Up to 1024 bounces	8.92 %	1.18 %	10.4 %	3.56 %

code path divergence, especially at shallow depth steps in path tracing. Therefore, the increased time is shorter at shallow depth steps than at deep depth steps.

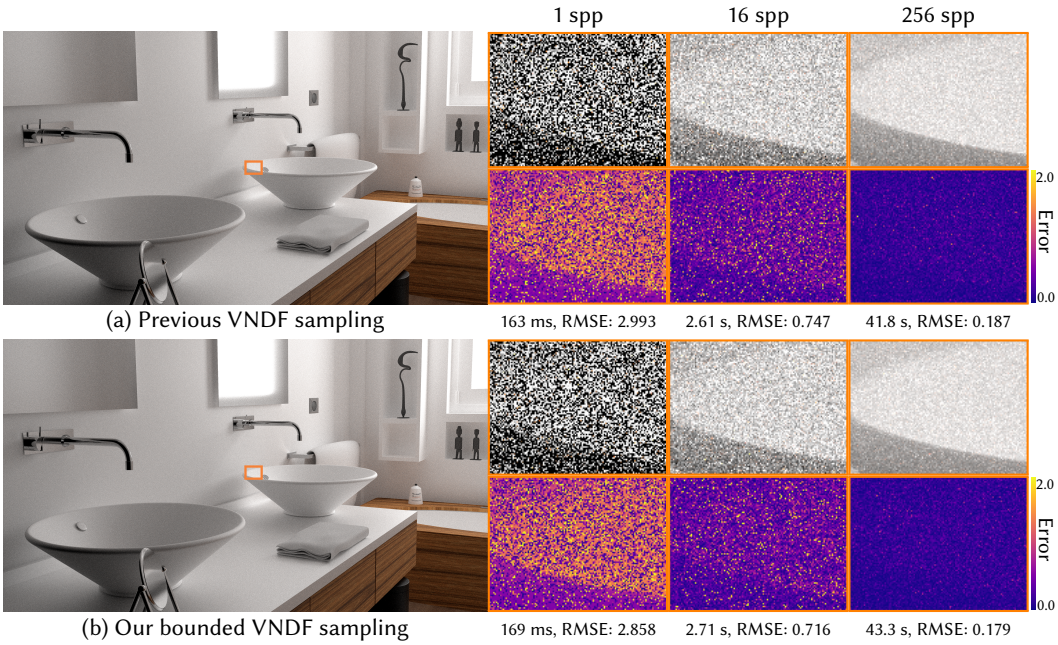


Fig. 9. Path tracing using multiple importance sampling of VNDF sampling and next event estimation for the Salle de bain scene (3840×2160 pixels). Unlike Fig. 8, (a) the previous VNDF sampling [Dupuy and Benyoub 2023] does not produce noticeable firefly noise for this scene. Even in this case, (b) our method reduces the RMSE by about 4.5% with a computational cost of 3.6%.

7 LIMITATIONS

Reflection Only. Since our method is specialized for reflections, it cannot be used to sample refraction vectors as shown in Fig. 11a. Our main target application is opaque surfaces such as metallic materials.

Highly Specular Surfaces. Since the previous VNDF sampling is sufficiently efficient for highly specular surfaces, the quality improvement by our method is slight for this case. On the other hand, our method is significantly efficient for rough reflections unlike the previous method. Therefore, our VNDF sampling is more robust than the previous method for scenes with various roughness.

False Positives. Although we introduce a lower bound for the spherical cap-based sampling range, the spherical cap is still wider than the region representing reflection vectors (see Fig. 3). Thus, our method still generates samples to be rejected (i.e., false positives). Such false positives are significant, especially for high anisotropy. Therefore, our method does not significantly reduce the variance for highly anisotropic surfaces.

Random Walk Multiple Scattering. Our method is designed for sampling reflection vectors in the upper hemisphere. Therefore, it cannot be used for microscale multiple scattering that requires reflection vectors in the lower hemisphere for Monte Carlo random walk [Bitterli and d'Eon 2022; Cui et al. 2023; Heitz et al. 2016; Wang et al. 2022] (Fig. 11b). On the other hand, our method is suitable for energy compensation-based multi-scattering approximations [Kulla and Conty 2017; Turquin 2019] that are often used in time-sensitive applications (e.g., video games).

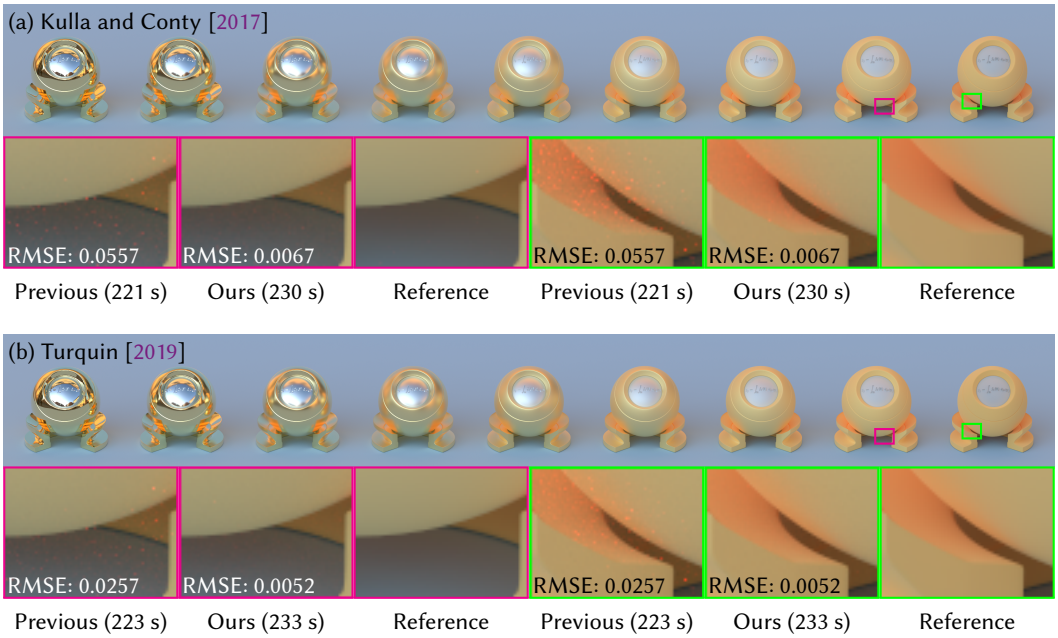
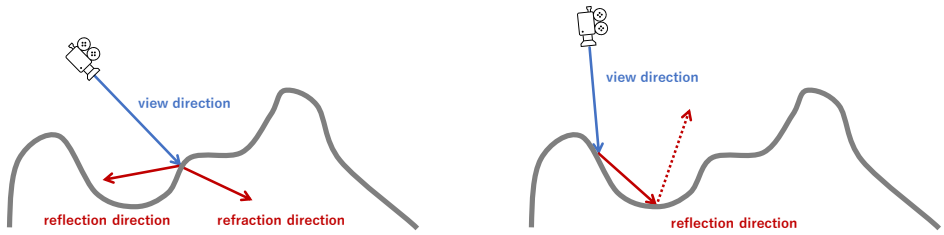


Fig. 10. Our bounded VNDF sampling for the Orb scene with different energy-compensated Smith-GGX BRDFs (4096×512 pixels, 16384 spp, 64×48 pixels for closeup). While the previous VNDF sampling [Dupuy and Benyoub 2023] produces noticeable firefly noise even for such a large number of samples, our method significantly reduces this noise. Although our method is effective for both Kulla and Conty's compensation (a) and Turquin's compensation (b), it is more suitable for Turquin's model.



(a) Refraction vector sampling (b) Random walk multiple scattering

Fig. 11. Our method cannot be applied for sampling refraction vectors (a) and random walk multiple scattering (b) in the microsurface. These scattering events have to sample microfacet normals whose reflection vectors can be in the lower hemisphere, while our method reduces such samples.

Rendering Time. Although our sampling routine has a negligible computational overhead, it increases effective rays to be traced instead of rejecting rays. Therefore, our method can increase the rendering time while significantly improving the cost-effectiveness.

8 CONCLUSION

In this paper, we have presented an unbiased Smith–GGX VNDF sampling method specialized for reflections. Our method reduces the number of reflection vectors occluded by the surface by limiting the sampling range of the spherical cap VNDF. For this limited sampling range in the

case of isotropic roughness, we derived the infimum of the range for effective reflection vectors in the upper hemisphere. For anisotropic roughness, we also introduced a simple analytical lower bound. Using our method, we reduced the variance for highly rough and low anisotropy surfaces while increasing the depth of light paths. This paper has also shown that our method significantly reduces undesirable firefly noise for energy-compensated Smith–GGX BRDFs, which are often used in production renderers and game engines.

For future work, we would like to improve the efficiency for highly anisotropic roughness by deriving a tighter bound than the conservative bound presented. A sampling method that completely avoids occluded reflection vectors is also left for future work.

A PROJECTED BOUND IN THE UNIT-ROUGHNESS SPACE

To obtain the lower bound for $\acute{o}_z$, we project the tangent-space reflection vector $\mathbf{o}$ to the unit-roughness space reflection vector $\acute{\mathbf{o}}$. For this projection, we first express $\acute{o}_z$ using $\mathbf{i}$ and $\mathbf{m}$, and then we rewrite the form into the function of $\mathbf{o}$ based on the relationship $\mathbf{o} = 2(\mathbf{i}\cdot\mathbf{m})\mathbf{m} - \mathbf{i}$. The reflection vector in the unit-roughness space is $\acute{\mathbf{o}} = 2(\acute{\mathbf{i}}\cdot\acute{\mathbf{m}})\acute{\mathbf{m}} - \acute{\mathbf{i}}$, where the incoming direction $\acute{\mathbf{i}}$ and the microfacet normal $\acute{\mathbf{m}}$ in the unit-roughness space are given by

$$\acute{\mathbf{i}} = \frac{[\alpha_x i_x, \alpha_y i_y, i_z]}{\left\|[\alpha_x i_x, \alpha_y i_y, i_z]\right\|}, \quad \acute{\mathbf{m}} = \frac{[m_x/\alpha_x, m_y/\alpha_y, m_z]}{\left\|[m_x/\alpha_x, m_y/\alpha_y, m_z]\right\|}. \tag{16}$$

Therefore, the projected $\acute{o}_z$ is obtained as follows:

$$\begin{aligned}\acute{o}_z = 2(\acute{\mathbf{i}}\cdot\acute{\mathbf{m}})\acute{m}_z - \acute{i}_z &= \frac{2m_z\left(\left[\alpha_x i_x, \alpha_y i_y, i_z\right]\cdot\left[m_x/\alpha_x, m_y/\alpha_y, m_z\right]\right)}{\left\|\left[\alpha_x i_x, \alpha_y i_y, i_z\right]\right\|\left\|\left[m_x/\alpha_x, m_y/\alpha_y, m_z\right]\right\|^2} - \acute{i}_z \\ &= \frac{2m_z\,(\mathbf{i}\cdot\mathbf{m})}{\left\|\left[\alpha_x i_x, \alpha_y i_y, i_z\right]\right\|\left\|\left[m_x/\alpha_x, m_y/\alpha_y, m_z\right]\right\|^2} - \acute{i}_z.\end{aligned} \tag{17}$$

From $\mathbf{o} = 2(\mathbf{i}\cdot\mathbf{m})\mathbf{m} - \mathbf{i}$, we get

$$\mathbf{i}\cdot\mathbf{m} = \|\mathbf{i}+\mathbf{o}\|/2, \quad [m_x, m_y, m_z] = [i_x+o_x, i_y+o_y, i_z+o_z]/\|\mathbf{i}+\mathbf{o}\|. \tag{18}$$

Thus, by substituting Eq. 18 into Eq. 17, we yield $\acute{o}_z$ as a function of $\mathbf{o}$:

$$\acute{o}_z(\mathbf{o}) = \frac{(i_z+o_z)\|\mathbf{i}+\mathbf{o}\|^2}{\left\|\left[\alpha_x i_x, \alpha_y i_y, i_z\right]\right\|\left\|\left[\frac{i_x+o_x}{\alpha_x}, \frac{i_y+o_y}{\alpha_y}, i_z+o_z\right]\right\|^2} - \acute{i}_z. \tag{19}$$

Let $[\theta, \phi]$ be the polar coordinate of the reflection vector $\mathbf{o}$ (i.e., $o_x = \sin\theta\cos\phi$, $o_y = \sin\theta\sin\phi$, and $o_z = \cos\theta$), then we can write $\acute{o_z}$ as a function of $[\theta, \phi]$. Since $o_z = \cos\theta > 0$ for reflections, we project the lower bound of o_z (i.e., $\theta = \pi/2$) into $\acute{o}_z(\theta, \phi)$ as follows:

$$\begin{aligned}\acute{o}_z\left(\frac{\pi}{2}, \phi\right) &= \frac{i_z\|\mathbf{i}+[\cos\phi, \sin\phi, 0]\|^2}{\left\|\left[\alpha_x i_x, \alpha_y i_y, i_z\right]\right\|\left\|\left[\frac{i_x+\cos\phi}{\alpha_x}, \frac{i_y+\sin\phi}{\alpha_y}, i_z\right]\right\|^2} - \acute{i}_z \\ &= \boxed{\left(\frac{(i_x+\cos\phi)^2 + (i_y+\sin\phi)^2 + i_z^2}{\frac{(i_x+\cos\phi)^2}{\alpha_x^2} + \frac{(i_y+\sin\phi)^2}{\alpha_y^2} + i_z^2} - 1\right)\acute{i}_z}.\end{aligned} \tag{20}$$

B DERIVATION OF OUR PDF

For spherical cap VNDF sampling, Dupuy and Benyoub [2023] derived the Smith–GGX VNDF (Eq. 8) based on microsurface transformation [Atanasov et al. 2022] as follows:

$$p(\mathbf{m}) = q\left(\acute{\mathbf{o}}\right)\left\|\frac{\partial \acute{\mathbf{o}}}{\partial \acute{\mathbf{m}}}\right\|\frac{|\det \mathbf{M}|}{\|\mathbf{M}\mathbf{m}\|^3}, \quad \text{where } \mathbf{M} = \begin{bmatrix} 1/\alpha_x & 0 & 0 \\ 0 & 1/\alpha_y & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{21}$$

$q\left(\acute{\mathbf{o}}\right)$ is the uniform distribution on the spherical cap:

$$q\left(\acute{\mathbf{o}}\right) = \frac{\chi^+\left(\acute{i}_z + \acute{o}_z\right)}{2\pi\left(\acute{i}_z + 1\right)}. \tag{22}$$

$\|\partial\acute{\mathbf{o}}/\partial\acute{\mathbf{m}}\| = 4|\acute{\mathbf{i}}\cdot\acute{\mathbf{m}}|$ is the Jacobian for the transformation between $\acute{\mathbf{o}}$ and $\acute{\mathbf{m}}$. $|\det\mathbf{M}|/\|\mathbf{M}\mathbf{m}\|^3$ is the Jacobian derived in Atanasov et al. [2022]. Thus, we yield

$$\left\|\frac{\partial \acute{\mathbf{o}}}{\partial \acute{\mathbf{m}}}\right\|\frac{|\det \mathbf{M}|}{\|\mathbf{M}\mathbf{m}\|^3} = \frac{4|\acute{\mathbf{i}}\cdot\acute{\mathbf{m}}|}{\alpha_x\alpha_y\left(\frac{m_x^2}{\alpha_x^2}+\frac{m_y^2}{\alpha_y^2}+m_z^2\right)^{\frac{3}{2}}} = \frac{4|\mathbf{i}\cdot\mathbf{m}|}{\alpha_x\alpha_y\left(\frac{m_x^2}{\alpha_x^2}+\frac{m_y^2}{\alpha_y^2}+m_z^2\right)^{2}\sqrt{\alpha_x^2 i_x^2+\alpha_y^2 i_y^2+i_z^2}}. \tag{23}$$

In this paper, we replace $q\left(\acute{\mathbf{o}}\right)$ with the uniform distribution on our spherical cap:

$$q_{\text{our}}\left(\acute{\mathbf{o}}\right) = \frac{\chi^+\left(k\acute{i}_z + \acute{o}_z\right)}{2\pi\left(k\acute{i}_z + 1\right)}. \tag{24}$$

Hence, the PDF for our bounded VNDF sampling is

$$p_{\text{our}}(\mathbf{m}) = q_{\text{our}}\left(\acute{\mathbf{o}}\right)\left\|\frac{\partial \acute{\mathbf{o}}}{\partial \acute{\mathbf{m}}}\right\|\frac{|\det \mathbf{M}|}{\|\mathbf{M}\mathbf{m}\|^3} = \boxed{\frac{2D\left(\mathbf{m}\right)\max\left(\mathbf{i}\cdot\mathbf{m},0\right)}{ki_z+\sqrt{\alpha_x^2 i_x^2+\alpha_y^2 i_y^2+i_z^2}}\chi^+\left(k\acute{i}_z+\acute{o}_z\right)}. \tag{25}$$

C NUMERICALLY STABLE FORM OF THE PREVIOUS PDF

When $i_z < 0$ (i.e., backfacing shading normal), our method uses the previous PDF (Eq. 8). However, if $i_z < 0$ and $\alpha_x^2 i_x^2 + \alpha_y^2 i_y^2$ is small enough compared to i_z^2, the denominator can produce catastrophic cancellation due to floating point arithmetic. This numerical error can induce zero divide for the PDF. To improve the numerical stability, we equivalently rewrite the previous PDF into the following equation:

$$\begin{aligned} p(\mathbf{m}) &= \frac{2D\left(\mathbf{m}\right)\max\left(\mathbf{i}\cdot\mathbf{m},0\right)}{i_z+\sqrt{\alpha_x^2 i_x^2+\alpha_y^2 i_y^2+i_z^2}} = \frac{2D\left(\mathbf{m}\right)\max\left(\mathbf{i}\cdot\mathbf{m},0\right)\left(i_z-\sqrt{\alpha_x^2 i_x^2+\alpha_y^2 i_y^2+i_z^2}\right)}{\left(i_z+\sqrt{\alpha_x^2 i_x^2+\alpha_y^2 i_y^2+i_z^2}\right)\left(i_z-\sqrt{\alpha_x^2 i_x^2+\alpha_y^2 i_y^2+i_z^2}\right)} \\ &= \boxed{\frac{2D\left(\mathbf{m}\right)\max\left(\mathbf{i}\cdot\mathbf{m},0\right)\left(\sqrt{\alpha_x^2 i_x^2+\alpha_y^2 i_y^2+i_z^2}-i_z\right)}{\alpha_x^2 i_x^2+\alpha_y^2 i_y^2}}. \end{aligned} \tag{26}$$

Since $\sqrt{\alpha_x^2 i_x^2 + \alpha_y^2 i_y^2 + i_z^2} - i_z$ does not produce catastrophic cancellation for $i_z < 0$, we use this form in our implementation (Listing 2).

REFERENCES

AMD. 2023. HIP Ray Tracing. https://gpuopen.com/hiprt/

Zap Andersson, Paul Edmondson, Julien Guertault, Adrien Herubel, Alan King, Peter Kutz, Andréa Machizaud, Jamie Portsmouth, and Frédéric Servant. 2023. OpenPBR Surface specification. https://academysoftwarefoundation.github.io/OpenPBR/

Asen Atanasov, Vladimir Koylazov, Rossen Dimov, and Alexander Wilkie. 2022. Microsurface Transformations. *Comput. Graph. Forum* 41, 4 (2022), 105–116. https://doi.org/10.1111/cgf.14590

Petr Beckmann and André Spizzichino. 1963. *Scattering of Electromagnetic Waves from Rough Surfaces.* MacMillan.

Benedikt Bitterli and Eugene d'Eon. 2022. A Position-Free Path Integral for Homogeneous Slabs and Multiple Scattering on Smith Microfacets. *Comput. Graph. Forum* 41, 4 (2022), 93–104. https://doi.org/10.1111/cgf.14589

Brent Burley. 2012. Physically-Based Shading at Disney. In *SIGGRAPH '12 Course: Practical Physically-based Shading in Film and Game Production.* Article 10, 7 pages. https://doi.org/10.1145/2343483.2343493

Robert L. Cook and Kenneth E. Torrance. 1982. A Reflectance Model for Computer Graphics. *ACM Trans. Graph.* 1, 1 (1982), 7–24. https://doi.org/10.1145/357290.357293

Yuang Cui, Gaole Pan, Jian Yang, Lei Zhang, Ling-Qi Yan, and Beibei Wang. 2023. Multiple-Bounce Smith Microfacet BRDFs Using the Invariance Principle. In *SIGGRAPH Asia '23 Conf. Papers.* Article 39, 10 pages. https://doi.org/10.1145/3610548.3618198

Johannes Deligiannis and Jan Schmid. 2019. It Just Works: Ray Traced Reflections in 'Battlefield V'. In *GDC '19.*

Jonathan Dupuy and Anis Benyoub. 2023. Sampling Visible GGX Normals with Spherical Caps. *Comput. Graph. Forum* 42, 8 (2023), e14867. https://doi.org/10.1111/cgf.14867

Iliyan Georgiev, Jamie Portsmouth, Zap Andersson, Adrien Herubel, Alan King, Shinji Ogaki, and Frédéric Servant. 2019. Autodesk Standard Surface. https://autodesk.github.io/standard-surface/

Eric Heitz. 2014. Understanding the Masking-Shadowing Function in Microfacet-Based BRDFs. *JCGT* 3, 2 (2014), 48–107.

Eric Heitz. 2018. Sampling the GGX Distribution of Visible Normals. *JCGT* 7, 4 (2018), 1–13. https://jcgt.org/published/0007/04/01/

Eric Heitz and Eugene d'Eon. 2014. Importance Sampling Microfacet-Based BSDFs using the Distribution of Visible Normals. *Comput. Graph. Forum* 33, 4 (2014), 103–112. https://doi.org/10.1111/cgf.12417

Eric Heitz, Johannes Hanika, Eugene d'Eon, and Carsten Dachsbacher. 2016. Multiple-Scattering Microfacet BSDFs with the Smith Model. *ACM Trans. Graph.* 35, 4, Article 58 (2016), 14 pages. https://doi.org/10.1145/2897824.2925943

Sébastien Hillaire and Charles de Rousiers. 2023. Authoring Materials That Matters - Substrate in Unreal Engine 5. In *SIGGRAPH '23 Course: Advances in Real-Time Rendering in Games, Part II.*

Wenzel Jakob. 2014. *An Improved Visible Normal Sampling Routine for the Beckmann Distribution.* Technical Report. ETH Zürich. https://www.mitsuba-renderer.org/~wenzel/files/visnormal.pdf

Wenzel Jakob, Eugene d'Eon, Otto Jakob, and Steve Marschner. 2014. A Comprehensive Framework for Rendering Layered Materials. *ACM Trans. Graph.* 33, 4, Article 118 (2014), 14 pages. https://doi.org/10.1145/2601097.2601139

Csaba Kelemen and Lászlő Szirmay-Kalos. 2001. A Microfacet Based Coupled Specular-Matte BRDF Model with Importance Sampling. In *EG'01 Short Presentations.* 25–34. https://doi.org/10.2312/egs.20011003

Christopher Kulla and Alejandro Conty. 2017. Revisiting Physically Based Shading at Imageworks. In *SIGGRAPH '17 Course: Physically Based Shading in Theory and Practice.* Article 7, 8 pages. https://doi.org/10.1145/3084873.3084893

Samuli Laine, Tero Karras, and Timo Aila. 2013. Megakernels Considered Harmful: Wavefront Path Tracing on GPUs. In *HPG '13.* 137–143. https://doi.org/10.1145/2492045.2492060

Stephen McAuley, Stephen Hill, Adam Martinez, Ryusuke Villemin, Matt Pettineo, Dimitar Lazarov, David Neubelt, Brian Karis, Christophe Hery, Naty Hoffman, and Hakan Zap Andersson. 2013. Physically Based Shading in Theory and Practice. In *SIGGRAPH '13 Courses.* Article 22, 8 pages. https://doi.org/10.1145/2504435.2504457

Bruce G. Smith. 1967. Geometrical shadowing of a random rough surface. *IEEE Trans. Antennas Propag.* 15, 5 (1967), 668–671. https://doi.org/10.1109/TAP.1967.1138991

Tomasz Stachowiak. 2015. Stochastic Screen-Space Reflections. In *SIGGRAPH '15 Course: Advances in Real Time Rendering.* https://doi.org/10.1145/2776880.2787701

Yusuke Tokuyoshi. 2021. Unbiased VNDF Sampling for Backfacing Shading Normals. In *SIGGRAPH '21 Talks.* Article 8, 2 pages. https://doi.org/10.1145/3450623.3464655

Kenneth E. Torrance and Ephraim M Sparrow. 1967. Theory for Off-Specular Reflection From Roughened Surfaces. *J. Opt. Soc. Am.* 57, 9 (1967), 1105–1114. https://doi.org/10.1364/JOSA.57.001105

T. S. Trowbridge and Karl P. Reitz. 1975. Average Irregularity Representation of a Rough Surface for Ray Reflection. *J. Opt. Soc. Am.* 65, 5 (1975), 531–536. https://doi.org/10.1364/JOSA.65.000531

Emmanuel Turquin. 2019. *Practical Multiple Scattering Compensation for Microfacet Models.* Technical Report. Industrial Light & Magic. https://blog.selfshadow.com/publications/turquin/ms_comp_final.pdf

Eric Veach and Leonidas J. Guibas. 1995. Optimally Combining Sampling Techniques for Monte Carlo Rendering. In *SIGGRAPH '95*. 419–428. https://doi.org/10.1145/218380.218498

Bruce Walter, Stephen R. Marschner, Hongsong Li, and Kenneth E. Torrance. 2007. Microfacet Models for Refraction Through Rough Surfaces. In *EGSR '07*. 195–206. https://doi.org/10.2312/EGWR/EGSR07/195-206

Beibei Wang, Wenhua Jin, Jiahui Fan, Jian Yang, Nicolas Holzschuch, and Ling-Qi Yan. 2022. Position-Free Multiple-Bounce Computations for Smith Microfacet BSDFs. *ACM Trans. Graph.* 41, 4, Article 134 (2022), 14 pages. https://doi.org/10.1145/3528223.3530112