Bounded VNDF Sampling for Smith–GGX Reflections

Kenta Eto*
Advanced Micro Devices, Inc.
Tokyo, Japan
Kenta.Eto@amd.com

Yusuke Tokuyoshi*
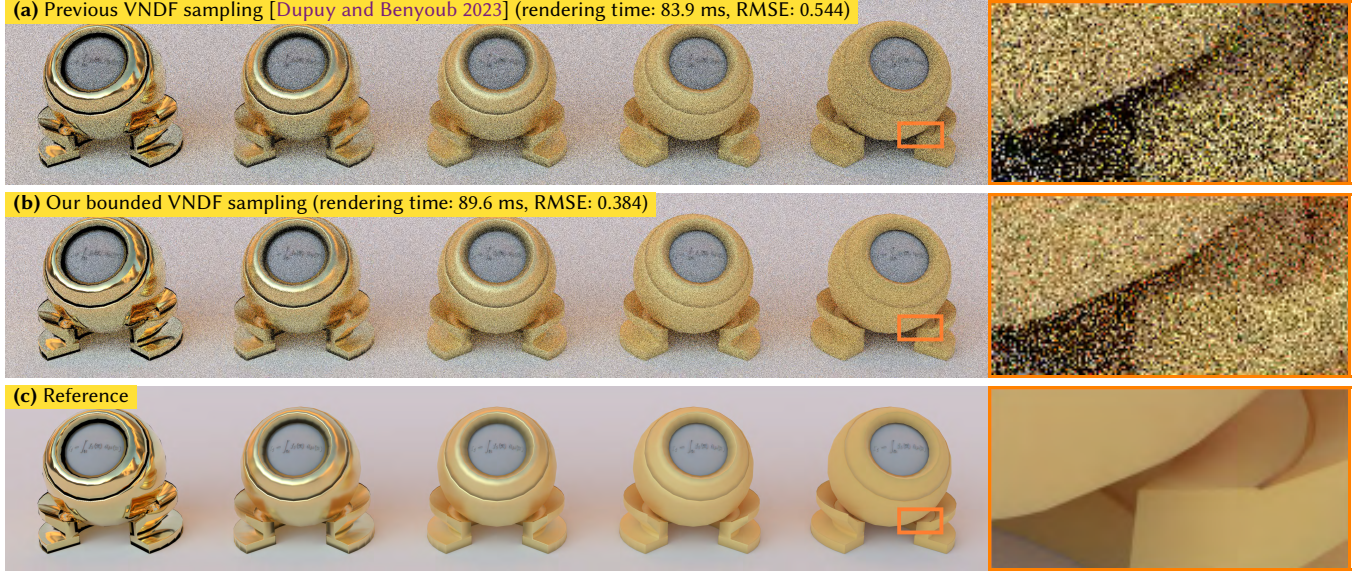Advanced Micro Devices, Inc.
Tokyo, Japan
yusuke.tokuyoshi@amd.com

Figure 1: Comparison between the previous VNDF sampling method [Dupuy and Benyoub 2023] (a) and our bounded VNDF sampling (b) for microfacet-based reflective surfaces (4096×768 pixels, 8 samples per pixel, AMD Radeon™ RX 7900 XTX GPU). Roughness parameter $[\alpha_x, \alpha_y]$ for each orb is $[0.01, 0.01]$, $[0.1, 0.1]$, $[0.3, 0.3]$, $[0.6, 0.6]$, and $[1, 1]$ from left to right. Our method efficiently reduces noise compared to the previous method while increasing the depth of paths, especially for rough surfaces.

ABSTRACT

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.

*These two authors contributed equally to this work.

CCS CONCEPTS

• **Computing methodologies → Ray tracing**.

KEYWORDS

importance sampling, microfacet BRDF, visible normal distribution

1 INTRODUCTION

Importance sampling according to a *visible normal distribution function* (VNDF) [Dupuy and Benyoub 2023; Heitz 2018; Heitz and d'Eon 2014] for the Smith–GGX microfacet BRDF model [Walter et al. 2007] is widely used for both offline and real-time rendering in computer graphics productions. It samples a microfacet normal visible to an incoming direction. By using VNDF sampling, we can reduce the variance for highly specular 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
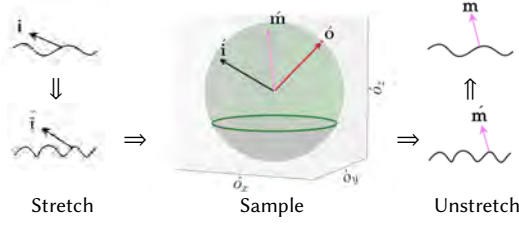
Figure 2: VNDF sampling with a spherical cap [Dupuy and Benyoub 2023]. This method samples a reflection vector within the spherical cap (middle) to obtain the visible normal in the stretched space.

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, for this case, existing VNDF sampling is twice as inefficient as naïve uniform sampling on the upper hemisphere. 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.

In this paper, 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 method for rough reflective surfaces, we can improve the image quality with small overhead.

Our contributions are as follows.

- We introduce a bounded VNDF sampling range to reduce the variance for rough reflections.
- To implement our method efficiently, we derive a bound of the spherical cap-based sampling range for reflections.
- We present the probability density function (PDF) for our method supporting backfacing shading normals.

2 BACKGROUND ON VNDF SAMPLING

Heitz [2018] introduced a practical VNDF sampling method for the Smith–GGX model [Walter et al. 2007]. He stretched the tangent space to transform the GGX NDF to a uniform hemispherical distribution and then sampled a visible normal in the hemisphere in this stretched space. Dupuy and Benyoub [2023] introduced a simpler VNDF sampling method (shown in Fig. 2) than Heitz's. They first sampled a reflection vector $\acute{\mathbf{o}} = [\acute{o}_x, \acute{o}_y, \acute{o}_z] \in S^2$ for an incoming direction $\acute{\mathbf{i}} = [\acute{i}_x, \acute{i}_y, \acute{i}_z] \in S^2$ in the stretched space, and then computed a stretched-space visible normal (i.e., halfvector) $\acute{\mathbf{m}} = [\acute{m}_x, \acute{m}_y, \acute{m}_z] \in S^2$ from $\acute{\mathbf{i}}$ and $\acute{\mathbf{o}}$. For the Smith–GGX model, $\acute{\mathbf{o}}$ is uniformly distributed on a spherical cap whose center is the surface normal and whose angle is $\arccos(-\acute{i}_z)$. Let $\mathbf{i} = [i_x, i_y, i_z] \in S^2$ be the tangent-space incoming direction and $[\alpha_x, \alpha_y] \in (0, \infty)^2$ be roughness parameters (i.e., stretching factors for tangent space), then their VNDF sampling is the following algorithm:

(1) Stretch the incoming direction $\mathbf{i}$ to $\acute{\mathbf{i}} = \frac{[\alpha_x i_x, \alpha_y i_y, i_z]}{\|[\alpha_x i_x, \alpha_y i_y, i_z]\|}$.

(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: $\mathbf{m} = \frac{[\alpha_x \acute{m}_x, \alpha_y \acute{m}_y, \acute{m}_z]}{\|[\alpha_x \acute{m}_x, \alpha_y \acute{m}_y, \acute{m}_z]\|}$.

After sampling a 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 Dupuy and Benyoub's method is simpler and faster than Heitz's method, 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 $(-\acute{i}_z, 1]$.

3 OUR METHOD

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 $\mathbf{o}$ into the stretched-space reflection vectors $\acute{\mathbf{o}}$. 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 stretched-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

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

For the derivation, please refer to the supplemental document. When $\acute{i}_z > 0$, we obtain ϕ that minimizes Eq. 1 by the following:

$$\underset{\phi}{\arg\min}\, \acute{o}_z\left(\frac{\pi}{2}, \phi\right) = \underset{\phi}{\arg\min}\, \frac{(i_x + \cos\phi)^2 + (i_y + \sin\phi)^2 + i_z^2}{\frac{(i_x+\cos\phi)^2}{\alpha_x^2} + \frac{(i_y+\sin\phi)^2}{\alpha_y^2} + i_z^2}. \quad (2)$$

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:

$$\underset{\phi}{\arg\min}\, \acute{o}_z\left(\frac{\pi}{2}, \phi\right) = \begin{cases} \arg\max_\phi r(\phi) & \text{if } \alpha < 1 \\ \arg\min_\phi r(\phi) & \text{if } \alpha > 1 \\ \mathbb{R} & \text{if } \alpha = 1 \end{cases} \quad (3)$$

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. The maximum and minimum of this squared

$[\alpha_x, \alpha_y] = [1, 1]$ $[\alpha_x, \alpha_y] = [0.5, 0.5]$ $[\alpha_x, \alpha_y] = [0.2, 0.2]$ $[\alpha_x, \alpha_y] = [1, 0.5]$ $[\alpha_x, \alpha_y] = [1, 0.2]$
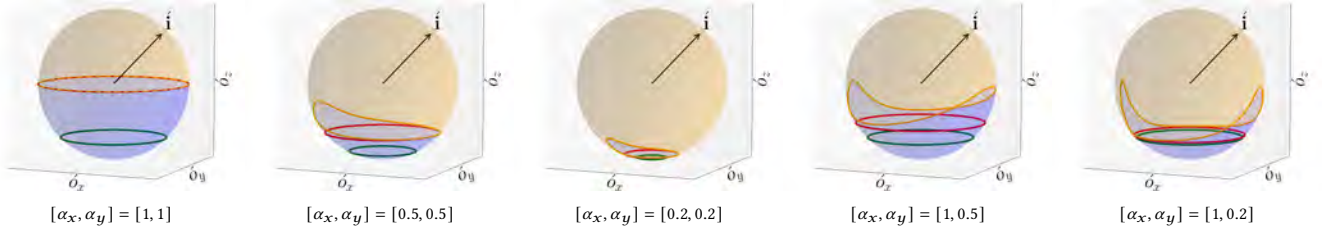
Figure 3: Previous lower bound (green) and our lower bound (red) for the spherical cap in VNDF sampling. The orange and blue regions correspond to reflection vectors in the upper and lower hemispheres, respectively. Since our method bounds the orange region more tightly than the previous method, it reduces the number of reflection vectors occluded by the surface.

distance $r(\phi)$ are $\left(1 \pm \sqrt{i_x^2 + i_y^2}\right)^2$. Thus, we rewrite Eq. 3 into

$$\underset{\phi}{\arg\min}\, \acute{o}_z\left(\frac{\pi}{2}, \phi\right) = \begin{cases} \{\phi \,|\, r(\phi) = s^2\} & \text{if } \alpha \neq 1 \\ \mathbb{R} & \text{if } \alpha = 1 \end{cases}. \quad (4)$$

where $s = 1 + \text{sgn}(1 - \alpha)\sqrt{i_x^2 + i_y^2}$. 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 for arbitrary $\acute{o}_z$ as follows:

$$\boxed{\min_{\phi} \acute{o}_z\left(\frac{\pi}{2}, \phi\right) = -k\acute{i}_z}, \quad \text{where} \quad \boxed{k = \frac{(1 - \alpha^2)s^2}{s^2 + \alpha^2 i_z^2}}. \quad (5)$$

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 case of isotropic roughness, Eq. 2 is not trivial for anisotropic roughness $\alpha_x \neq \alpha_y$. Therefore, for $\acute{i}_z > 0$, we introduce a loose bound that is obtained by Eq. 5 with setting α conservatively as follows:

$$\alpha = \min(\alpha_x, \alpha_y, 1). \quad (6)$$

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 supplemental document for the interactive visualization of our bound.

3.3 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

$$p(\mathbf{m}) = \frac{2D(\mathbf{m})\max(\mathbf{i} \cdot \mathbf{m}, 0)}{i_z + \sqrt{\alpha_x^2 i_x^2 + \alpha_y^2 i_y^2 + i_z^2}}, \quad (7)$$

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

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. 6
  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. 5
  float b = i.z > 0 ? k * i_std.z : i_std.z;
  float z = mad(1.0f - rand.y, 1.0f + b, -b);
  float sinTheta = sqrt(saturate(1.0f - z * z));
  float3 o_std = {sinTheta * cos(phi), sinTheta * 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: $p_{\text{our}}(\mathbf{m})\|\text{dm/do}\|$ 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 $p(\mathbf{m})\|\text{dm/do}\|$ using a numerically stable form (please see the supplemental document for derivation).

```
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. 6
    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. 5
    return ndf / (2.0f * (k * i.z + t)); // Eq. 8 * ||dm/do||
  }
  // Numerically stable form of the previous PDF for i.z < 0
  return ndf * (t - i.z) / (2.0f * len2); // = Eq. 7 * ||dm/do||
}
```

the spherical cap from $(-\acute{i}_z, 1]$ to $(-k\acute{i}_z, 1]$. Thus, by replacing the spherical cap in the previous PDF to our spherical cap, we obtain the PDF for our sampling technique:

$$\boxed{p_{\text{our}}(\mathbf{m}) = \frac{2D(\mathbf{m})\max(\mathbf{i} \cdot \mathbf{m}, 0)}{k i_z + \sqrt{\alpha_x^2 i_x^2 + \alpha_y^2 i_y^2 + i_z^2}} \chi^+\left(k\acute{i}_z + \acute{o}_z\right)}. \quad (8)$$

For derivation, please refer to the supplemental document. Since our method limits the sampling range for $ó_z$ into $(-k\dot{i}_z, 1]$, we represent this limited range using the Heaviside function $\chi^+(k\dot{i}_z + ó_z)$: 1 if $ó_z > -k\dot{i}_z$ and 0 if $ó_z \le -k\dot{i}_z$. By multiplying the above PDF $p_{\text{our}}(\mathbf{m})$ by the Jacobian for the transformation between halfvectors and reflection vectors $\|d\mathbf{m}/d\mathbf{o}\| = 1/(4|\mathbf{i} \cdot \mathbf{m}|)$, we obtain the PDF for our reflection vector sampling $p_{\text{our}}(\mathbf{m})\|d\mathbf{m}/d\mathbf{o}\|$. Listing 2 shows the code for our PDF for reflections. In this implementation, we omit the Heaviside function $\chi^+(k\dot{i}_z + ó_z)$, because we use our sampling method only for opaque surfaces. For opaque surfaces, we get $\chi^+(k\dot{i}_z + ó_z) = 1$ 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.

4 RESULTS AND FUTURE WORK

Here we show images rendered using a GPU path tracer performed on an AMD Radeon™ RX 7900 XTX GPU. All materials are the Smith–GGX microfacet model with practical energy compensation using lookup tables [Turquin 2019] which is often used in productions. The image quality is evaluated with the root-mean-squared error (RMSE) metric.

Figs. 1 and 4 show the quality comparison between our method and the previous method [Dupuy and Benyoub 2023] using different isotropic roughness parameters. Our method reduces errors, especially for high-roughness surfaces ($\alpha_x = \alpha_y = 0.8$ in Fig. 4). On the other hand, the quality difference is insignificant for low roughness ($\alpha_x = \alpha_y = 0.2$ in Fig. 4). This is because our bounding spherical cap depends on the roughness parameter unlike the previous method, and the difference from the previous spherical cap is more significant for a higher roughness parameter. Although our method increases the depth of paths by reducing the number of rejected sample rays, the increase in rendering time shown in Fig. 1 is small compared to the quality improvement. Figs. 5 and 6 show the quality comparison for anisotropic roughness. Since our bounding spherical cap is tighter for lower anisotropy, our method is efficient, especially for low anisotropy.

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.

ACKNOWLEDGMENTS

REFERENCES

Brent Burley. 2012. Physically-Based Shading at Disney. In *SIGGRAPH '12 Course: Practical Physically-based Shading in Film and Game Production*. 10:1–10:7. https://doi.org/10.1145/2343483.2343493

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

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
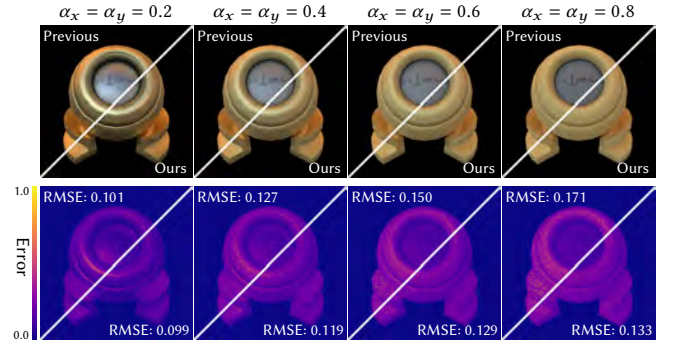

Figure 4: Quality comparison between previous VNDF sampling [Dupuy and Benyoub 2023] and our bounded VNDF sampling using different isotropic roughness (1024×1024 pixels, 16 samples per pixel). Our sampling reduces RMSE, especially for higher roughness.
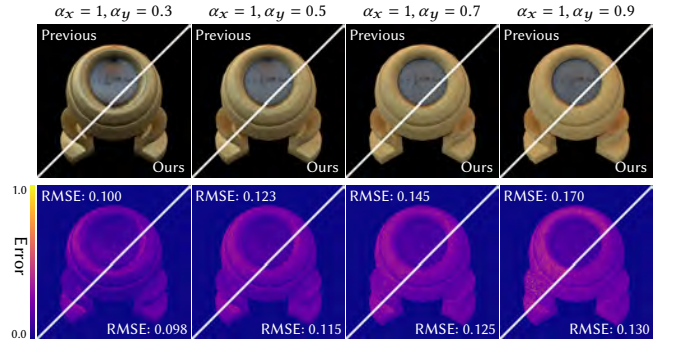

Figure 5: Quality comparison between previous VNDF sampling [Dupuy and Benyoub 2023] and our bounded VNDF sampling using different anisotropic roughness (1024×1024 pixels, 16 samples per pixel). Our sampling reduces RMSE, especially for lower anisotropy.
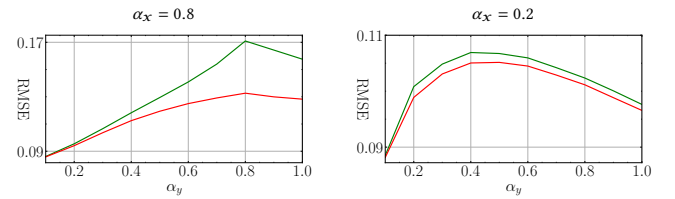

Figure 6: Plots of RMSE for the previous method [Dupuy and Benyoub 2023] (green) and our method (red) using different anisotropic roughness for the orb model used in Fig. 5

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

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

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