Slope-Space Integrals for Specular Next Event Estimation

GUILLAUME LOUBET, École Polytechnique Fédérale de Lausanne (EPFL)
TIZIAN ZELTNER, École Polytechnique Fédérale de Lausanne (EPFL)
NICOLAS HOLZSCHUCH, Inria, Univ. Grenoble-Alpes, CNRS, LJK
WENZEL JAKOB, École Polytechnique Fédérale de Lausanne (EPFL)

Fig. 1. We address the problem of importance sampling light paths involving specular or near-specular reflection (a) or refraction (b). Our approach finds triangles that reflect or refract light from a point on a light source towards a particular shading point in the scene. Our technique can be used in addition to standard next event estimation and handles specular-diffuse-specular (“SDS”) sub-paths that are a well-known failure case of standard uni- and bidirectional path tracers. The images were rendered in 30m (a) and 5m (b) using an unidirectional path tracer together with the proposed sampling technique.

Monte Carlo light transport simulations often lack robustness in scenes containing specular or near-specular materials. Widely used uni- and bidirectional sampling strategies tend to find light paths involving such materials with insufficient probability, producing unusable images that are contaminated by significant variance.

This article addresses the problem of sampling a light path connecting two given scene points via a single specular reflection or refraction, extending the range of scenes that can be robustly handled by unbiased path sampling techniques. Our technique enables efficient rendering of challenging transport phenomena caused by such paths, such as underwater caustics or caustics involving glossy metallic objects.

We derive analytic expressions that predict the total radiance due to a single reflective or refractive triangle with a microfacet BSDF and we show that this reduces to the well known Lambert boundary integral for irradiance. We subsequently show how this can be leveraged to efficiently sample connections on meshes comprised of vast numbers of triangles.

Our derivation builds on the theory of off-center microfacets and involves integrals in the space of surface slopes. Our approach straightforwardly applies to the related problem of rendering glints with high-resolution normal maps describing specular microstructure. Our formulation alleviates problems raised by singularities in filtering integrals and enables a generalization of previous work to perfectly specular materials. We also extend previous work to the case of GGX distributions and introduce new techniques to improve accuracy and performance.

CCS Concepts: • Computing methodologies → Rendering; Ray tracing.

Additional Key Words and Phrases: specular transport, SDS paths, microfacet theory, glints

ACM Reference Format:

1 INTRODUCTION

Physically-based rendering of scenes containing specular materials like metal, glass, water, or plastic requires robust sampling strategies that can find light paths involving specular or near-specular interactions. Specular paths that connect the camera and light sources subtending a small solid angle often lead to visually striking effects in rendered images. Caustics induced by the glossy objects in Fig. 1a, and the underwater scene in Fig. 1b are examples of such specular paths. Uni- and bidirectional path tracing algorithms [Lafortune and Willems 1993; Veach and Guibas 1994] are widely used yet...
Our approach builds on a hierarchical data structure to efficiently identify the triangles involved in indirect specular transport at render time.

Our approach is closely related to previous work on sampling specular paths in refractive media [Walter et al. 2009], which we improve and generalize to the case of rough materials and reflection. In contrast to path guiding approaches [Müller et al. 2017; Reibold et al. 2018; Vorba et al. 2014] and algorithms based on walks on specular manifolds [Hanika et al. 2015a; Jakob 2013; Jakob and Marschner 2012], our approach achieves robust global exploration and ensures high-quality sampling of this type of specular sub-paths. Its main limitation is that it can only sample sub-paths with exactly one intermediate specular interaction, hence more complex specular chains must be filtered or sampled using other techniques.

Our method can also be applied to the related problem of rendering specular glints on geometry with high-frequency normal maps [Yan et al. 2014]. Our approach extends this work to the case of GGX distributions of microfacets and it reduces artifacts by accounting for surface curvature and varying half vectors inside texels. We also propose an improved pruning strategy for the case of rough materials. Our formulation alleviates problems raised by singularities in filtering integrals and enables a generalization to perfectly specular materials that runs about twice as fast as in the case of rough materials with tiny amounts of roughness, due to simpler expressions and a more drastic pruning of the search space.

An open source implementation of the methods presented in this paper is available at rgl.epfl.ch/publications/Loubet2020SlopeSpace.

2 RELATED WORK

Our approach builds on microfacet theory and is related to specialized sampling techniques for specular transport. We briefly review both areas in this section.

Specular path sampling. Scenarios with small light sources and narrowly peaked BRDFs are traditional failure cases of various rendering algorithms, which has spurred the development of a number of specialized approaches. Mitchell and Hanrahan [1992] proposed an algorithm for rendering caustics due to an implicitly defined surface, relying on interval arithmetic and a Newton-like iteration to find solutions. Walter et al. [2009] proposed a related technique for arbitrary triangle meshes with interpolated shading normals. Their work focuses on smooth dielectrics and introduces a spatiotemporal acceleration data structure to prune large parts of the search space. We generalize their approach to the case of reflection and refraction from rough conductors and dielectrics, and we propose a new pruning strategy that substantially accelerates sampling of specular paths.

Jakob and Marschner’s [2013; 2012] Manifold Exploration method samples specular light paths within the framework of Metropolis Light Transport [Veach and Guibas 1997]. They leverage information provided by the differential geometry of the manifold of specular paths to sample valid configurations in the neighborhood of a given path, leading to a “manifold walk” in path space. Their approach was later considerably improved for the case of rough materials [Hanika et al. 2015b; Kaplanyan et al. 2014].

Manifold Next Event Estimation (MNEE) proposed by Hanika et al. [2015a] imports these tools into the framework of classical (i.e. non-MCMC) Monte Carlo methods. Their algorithm—like ours—samples specular light paths connecting two points. Starting from a deterministically chosen invalid initialization, their method performs a manifold walk to bring the starting path into a valid physical configuration. MNEE handles specular and near-specular refraction events, but it can at most find a single specular path connecting two
given endpoints. This is sufficient to render simple smooth geometry (e.g. spheres), but becomes problematic when the geometry is more complex and many solutions exist. The recent work of Zeltner et al. [2020] introduces a specular path sampling routine based on manifold walks that addresses some of these limitations.

**Guided path tracing.** Path guiding approaches exploit information about previously generated light paths to improve later sampling steps. Various types of caches have been proposed, including 5-dimensional incident radiance fields [Herholz et al. 2016; Müller et al. 2017; Vorba et al. 2014] and sets of high variance paths [Reibold et al. 2018]. In the case of caustics, the radiance function is characterized by high-frequency variation in both its spatial and directional arguments, which means that guiding techniques require a high-resolution representation built from a large number of relevant light paths. In common cases involving small/distant light sources or complex specular geometry, such paths are sampled with an insufficient probability, and guiding becomes ineffective. High-resolution normal maps further exacerbate this type of problem. We compare our method to state-of-the-art guiding techniques to demonstrate these limitations.

**Glint rendering.** Specular microstructure, e.g. specified via high-resolution normal maps, tends to produce severe aliasing and variance in rendered images. This happens due to small sub-pixel highlights that standard techniques fail to sample efficiently. Filtering approaches [Bruneton and Neyret 2011; Dupuy et al. 2013; Olano and Baker 2010; Xu et al. 2017] resolve this lack of convergence but generally result in a loss of visual fidelity and realism. Yan et al. [2014] proposed an algorithm for evaluating the average distribution of normals within a given surface region, which enables efficient rendering of glinty appearance. Like Walter et al. [2009], they rely on spatio-directional hierarchies to prune large groups of texels when evaluating normal queries. We reformulate their approach in the space of microfacet slopes (Sec. 5), while additionally accounting for surface curvature and variations of view and light directions. We also introduce a simpler approach for of smooth reflectance models, as well as better pruning strategies for the rough case.

Yan et al. [2016] improved the performance of the previous approach by approximating normal distributions using 4D Gaussian mixtures with millions of elements. This trades higher storage footprint (several gigabytes of memory for a typical normal map) in exchange for a roughly hundred-fold performance improvement. Our spatial-angular slope space hierarchies could also benefit from this optimization. Many extensions and variants of glint rendering techniques have been proposed in prior work, e.g., to support multiple scattering [Chermain et al. 2019; Raymond et al. 2016], directional product integrals [Gamboa et al. 2018], procedural glints [Jakob et al. 2014; Kuznetsov et al. 2019], wave optics [Werner et al. 2017; Yan et al. 2018] and real-time rendering [Zirr and Kaplanyan 2016]. Some of these methods have been formulated in slope space [Chermain et al. 2019; Zirr and Kaplanyan 2016].

**Off-center microfacet models.** We build on microfacet models with off-center normal distributions, meaning that their average normal differs from the geometric normal. Such distributions were used by Olano and Baker [2010] to filter normal maps, and Dupuy, Heitz et al. [2013] introduced the missing masking and shadowing functions. Schussler et al. [2017] recently added vertical microfacets to off-center distributions to ensure energy conservation on normal-mapped surfaces. A symmetric off-center BRDF model can be found in the work of Dupuy [2015]. In the supplemental material of the article Dong et al. [2015], Walter et al. [2015] introduced an off-center microfacet distribution of normals called the Ellipsoid NDF along with evaluation, sampling and shadowing-masking terms. This model is exactly equivalent to off-center GGX distributions as discussed in our supplemental material.

**Integrals over normals and slopes.** The spaces of microfacet slopes and projected directions appear in numerous prior works [Dupuy et al. 2013; Hanika et al. 2015b; Heitz 2014; Kaplanyan et al. 2014; Olano and Baker 2010; Smith 1986] and have proven to be convenient spaces for addressing light transport and filtering problems. In particular, Kaplanyan et al. [2016] rely on slope-space integrals for filtering specular highlights on specular surfaces with high curvature. Their approach had a limited accuracy in the case of GGX distributions because of the lack of analytic integrals of this distribution in an arbitrary polygonal domain. We provide these missing formulas in section 3.

### 3 SLOPE-SPACE INTEGRALS FOR SPECULAR TRIANGLES

In this section, we derive expressions that approximate the total amount of transport between two points \( p_s \) and \( p_l \) by means of a triangle with shading normals defined at its vertices and a specular or near-specular BSDF, as shown in Fig. 2. Such expressions enable the importance sampling of specular triangles when rendering caustics (Section 4), or specular texels when rendering glints (Section 5). We also discuss importance sampling techniques for the positions of specular interactions within a triangle.

As spatially varying roughness within triangles would lead to intractable integration problems in the general case, we assume that BSDF roughness is constant in each triangle. For the sake of simplicity, also we assume that the endpoints \( p_s \) and \( p_l \) are located on scene surfaces. Our results easily generalize to other cases, e.g., when \( p_s \) is a directional environment light source, or when \( p_l \) is a camera.

#### 3.1 Integral of the radiance at \( p_v \)

Given a position \( p_v \) (e.g. a position on a light source), the radiance due to paths with two segments connecting \( p_v \) and \( p_l \) by means of

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( D )</td>
<td>Distribution of normals</td>
</tr>
<tr>
<td>( P )</td>
<td>Normalized distribution of slopes</td>
</tr>
<tr>
<td>( p_l, p_s, p_h )</td>
<td>3D points</td>
</tr>
<tr>
<td>( l, v, n, h )</td>
<td>3D vectors and normals</td>
</tr>
<tr>
<td>( n, h )</td>
<td>2D positions in slope space</td>
</tr>
<tr>
<td>( \Delta_l )</td>
<td>Triangle in world space</td>
</tr>
<tr>
<td>( \Delta_h )</td>
<td>Triangle in slope space</td>
</tr>
<tr>
<td>( f_s, D_s, P_s )</td>
<td>BSDF and distributions at the point ( p_s )</td>
</tr>
</tbody>
</table>

Table 1. List of important notations.
a specular triangle $\Delta_i$ is given by an integral:
\[
\int_{\Delta_i} f_s(\alpha_s, -\nu) \frac{|v \cdot n_v| |V(p_v \leftrightarrow p_s)| |v \cdot n_s|}{||p_v - p_s||^2} f_s(v, l) |l \cdot n_s|^L(p_s, -l) \, dp_s
\]  
(1)

where $p_s \in \Delta_i$, $l = \frac{p_s p_l}{p_s}$, $v = \frac{p_s p_v}{p_s}$, $f_s$ and $n_v$ are the BSDF and normal at $p_v$, $f_s$ and $n_s$ are the BSDF and normal at $p_s$, and $V(p_v \leftrightarrow p_s)$ is a visibility term whose value is either 0 or 1. The term $L(p_s, -l)$ denotes the incident radiance at $p_s$ from $p_l$. For simplicity, we omit the direction $\alpha_s$ in the following equations.

Eq. 1 generally lacks an analytic solution that would be needed to make caustics and glints rendering practical. The difficulty in obtaining closed-form results is related to three challenges:

- View and light directions $v$ and $l$ vary non-linearly and affect most of the terms including the BSDF $f_s(v, l)$ and the incident radiance $L(p_s, -l)$.
- Variations in the shading normal effectively induce spatial variation in the BSDF of the specular surface.
- When $f_s$ is a smooth conductor or a smooth dielectric, this integral reduces to a discrete sum of light path contributions.

This is the approach studied by Walter et al. [2009] for dielectric materials. Even here, there is no explicit solution for valid light paths, and previous works [Hanika et al. 2015a; Walter et al. 2009] require iterative nonlinear solvers.

We will introduce simplifying assumptions to resolve these mathematical difficulties.

### 3.2 Far-field approximation

Far-field approximations neglect small variations in directions and/or positions involving a distant surface. They have been applied in the context of glint rendering [Jakob et al. 2014; Yan et al. 2014, 2016], where the simplified problem reduces to evaluating the averaged distribution of normals in a surface region.

We instead introduce a far-field approximation within each triangle $\Delta_i$, which is very accurate for most combinations of shading points, light samples and triangles in a typical scene, in particular for high-quality meshes with small triangles. In practice, we ensure that this approximation always holds by using on-the-fly adaptive triangle subdivisions as discussed in section 4.6.

Assuming that $p_v$ and $p_l$ are distant with respect to triangle $\Delta_i$, and that the variations of the BRDF at $p_v$ and the incident light at $p_s$ are negligible, most of the terms of Eq. 1 can be moved outside the integral. The problem then reduces to estimating the average BRDF over $\Delta_i$ for fixed directions $v$ and $l$:

\[
f_v(-\nu) \frac{|v \cdot n_v| |V(p_v \leftrightarrow p_s)| |v \cdot n_s|}{||p_v - p_s||^2} f_s(v, l) |l \cdot n_s|^L(p_s, -l) \, dp_s
\]  
(2)

\[
\approx f_v(-\nu) \frac{|v \cdot n_v| |V(p_v \leftrightarrow p_s)| |v \cdot n_s|}{||p_v - p_s||^2} |l \cdot n_s|^L(p_s, -l) \int_{\Delta_i} f_s(v, l) \, dp_s.
\]  
(3)

We note that $f_v$ remains spatially varying due to interpolated shading normals over $\Delta_i$. When shading frames are used, Eq. 3 involves some form of spherical convolution with spherical triangles, for which there is no known closed-form expressions in general.

### 3.3 Off-center distributions of slopes

This problem simplifies when $f_s$ is defined using a microfacet BSDFs with off-center microfacet distributions instead of shading frames. As in previous work, we define off-center distributions as translations of centered distributions in the space of microfacet slopes [Dupuy et al. 2013]. The conversion between normals and slopes defines a bijection:

\[
\tilde{m} = (\tilde{m}_x, \tilde{m}_y)^T = \left( -\frac{m_x}{m_z}, -\frac{m_y}{m_z} \right)^T \Leftrightarrow m = \left( -\frac{\tilde{m}_x}{\tilde{m}_z + 1}, \frac{\tilde{m}_y}{\tilde{m}_z + 1} \right)^T
\]  
(4)

where $\tilde{m}_x$ and $\tilde{m}_y$ are the slopes of the microfacet with normal $m = (m_x, m_y, m_z)^T$. As in previous work [Dupuy et al. 2013] we define and use normalized distributions in slope space:

\[
D(m) = \frac{P(m)}{|m| n|} n^T,
\]  
(5)

where $P$ is the normalized distribution of slopes and $n$ is the normal of the surface.

In the case of reflection, for instance, the integral term in Eq. 3 with such BSDF models yields:

\[
\int_{\Delta_i} f_s(v, l) \, dp_s = \int_{\Delta_i} \frac{D_s(h) F(v, h) G(v, l, h)}{4 |v \cdot n_s| |l \cdot n_s|^2} \, dp_s
\]  
(6)

\[
\approx \int_{\Delta_i} \frac{P_s(h) F(v, h) G(v, l, h)}{4 |v \cdot n_s| |l \cdot n_s|^2} \, dp_s
\]  
(7)

\[
= \int_{\Delta_i} \frac{F(v, h) G(v, l, h)}{4 |v \cdot n_s| |l \cdot n_s|^2} \int_{\Delta_i} P_s(h) \, dp_s
\]  
(8)

where $P_s$ is a spatially-varying slope distribution, $h$ denotes the half vector at $p_s$ aligned with $v + l$, $F$ is the Fresnel term and $G$ is the masking/shadowing term [Cook and Torrance 1982; Walter et al. 2007].

### 3.4 Linear variations of slopes

The integrand $P_s$ in Equation 8 is a spatially-varying distribution, which depends on the variations of the shading normals over the triangle $\Delta_i$, that is, the variations of the mean slope of the distribution. This integral does not have a closed-form expression in general. However, we found that closed-form expressions can be derived...
we always ensure that these approximations hold by subdividing whose support is the same spherical triangle. This is similar to the approach of Yan et al. [2014] who interpolate orthogonal projections slope space results in spherical distributions of interpolated normals the Cartesian components in our experiments. As illustrated in Fig. 4, instead of neglecting variations of half-vectors in specular triangles, we use a linear approximation of such variations in slope space. This approximation is very accurate when points \( p_v \) and \( p_l \) are distant enough from the triangle (b), and it becomes less accurate when one of the point is close to the triangle, as half vector variations become increasingly non-linear (c). Note that points \( p_v \) and \( p_l \) are typically much farther from the triangle than what is shown case (b) (i.e., the approximation is superior), especially when working with high-resolution meshes and normal maps.

if the mean slope varies linearly over \( \Delta_i \). Therefore, we linearly interpolate the shading normals in the space of slopes:

\[
\hat{n}(b_0, b_1, b_2) = b_0 \hat{n}_0 + b_1 \hat{n}_1 + b_2 \hat{n}_2
\]

where \( b_0, b_1, b_2 \) are the barycentric coordinates of \( p_s \) in \( \Delta_i \), \( \hat{n}_0, \hat{n}_1, \hat{n}_2 \) are the slopes of the shading normals at the vertices of \( \Delta_i \), and \( \hat{n} \) is the mean slope at \( p_s \). We found that this interpolation in slope space is visually indistinguishable from interpolation of the Cartesian components in our experiments. As illustrated in Fig. 3, interpolating three normals in the spherical domain and in slope space results in spherical distributions of interpolated normals whose support is the same spherical triangle. This is similar to the approach of Yan et al. [2014] who interpolate orthogonal projections of the normals onto the tangent plane for the same reasons.

Moreover, we can relax the far-field assumption in Equation 8 and take into account first-order variations of half vector inside the triangle. This is similar to the approximations used in the work of Kaplanyan et al. [2016]—they use first-order approximations in projected pixel footprints onto the surface, while we use first-order approximations inside triangles \( \Delta_i \). We first compute the exact half vectors at the corners of a triangle and their corresponding slopes, and we replace the constant half-vector \( \hat{h} \) by a function \( \hat{h}_{\text{linear}}(b_0, b_1, b_2) \) that defines approximate slopes for half vectors in the triangle using linear interpolation:

\[
\hat{h}_{\text{linear}}(b_0, b_1, b_2) = b_0 \hat{h}_0 + b_1 \hat{h}_1 + b_2 \hat{h}_2.
\]

Examples of such linear approximations of the variations of half vectors can be found in Fig. 4. These approximations are accurate unless \( p_v \) or \( p_l \) lie within a small distance of the triangle. In practice, we always ensure that these approximations hold by subdividing triangles on the fly, as discussed in section 4.6.

Finally, we can write Equation 8 in the space of slopes. There is a bijection between positions \( p_s \) on the triangle \( \Delta_i \) and mean slopes \( \hat{n} \), as well as a bijection between positions \( p_v \) and the approximate slopes \( \hat{h}_{\text{linear}} \) of the half-vector. Since these transformations are linear, the integral over \( \Delta_i \) can be rewritten as an integral over a triangle in slope space. We now use the fact that the slopes \( \hat{n} \) define a translation of a centered slope distribution \( P \), and turn Eq. 8 into a slope-space integral over the triangle \( \Delta_i = (\hat{h}_0 - \hat{n}_0, \hat{h}_1 - \hat{n}_1, \hat{h}_2 - \hat{n}_2) \):

\[
\int_{\Delta_i} P(\hat{h}_{\text{linear}}) \, d\hat{p}_s = \int_{\Delta_i} P(\hat{s} - \hat{n}_{\text{linear}}) \left( \frac{\partial \hat{p}_s}{\partial \hat{s}} \right) d\hat{s}
\]

The Jacobian term can be computed from the linear transformations from world space to barycentric coordinates and from barycentric coordinates to slope space, it remains constant for each particular triangle \( \Delta_i \). The above equation shows that the combined approximations of distant endpoints, off-center microfacets, and barycentric normal interpolation in slope space transform the total radiance due to an individual triangle into a slope-space integral over a triangular domain, as shown in Fig. 5a.

Note that these derivations are similar in spirit to the frameworks of Yan et al. [2014] (in the space of projected normals) and Kaplanayan et al. [2016] (based on differential geometry in surface filtering kernels).

### 3.5 Integrating slope distributions on arbitrary polygons

The integration of Beckmann slope distributions (i.e., 2D Gaussians) was previously studied by Yan et al. [2014]. Due to the lack of a closed-form expression for this integral, Yan et al. rely on a spline approximation.

We show that such integrals can be easily computed in the case of GGX distributions. First, the integration of arbitrary GGX distributions in polygonal domains (Fig. 5a) can be reduced to the case of the isotropic GGX distribution with unit roughness \( P_{\text{std}} \) through a simple linear transformation of the space of slopes, as shown in Fig. 5b and discussed in our supplemental material. The corresponding GGX distribution of normals \( D_{\text{std}} \) is a uniform distribution on the hemisphere:

\[
\int_{\Delta} P_{\text{std}}(\hat{m}) \, d\hat{m} = \frac{1}{\pi} \int_{\Delta} [\hat{m} \cdot \hat{n}] \, d\hat{m}. \tag{13}
\]

We have seen that a triangular domain in the space of slopes maps to a spherical triangle (Fig. 3). This gives a geometric interpretation for Eq. 13: GGX integrals in polygonal domains are equivalent to integrating cosine-weighted spherical triangles, that is, vertically-projected areas of spherical triangles as shown in Fig. 5c. Such integrals have a simple analytic expression:

\[
\frac{1}{\pi} \int_{\Delta} [\hat{m} \cdot \hat{n}] \, d\hat{m} = \frac{1}{2\pi} \sum_{i=0}^{3} \frac{1}{\cos(\hat{n}_i \cdot \hat{n}_j)} \left\| \frac{\hat{n}_i \times \hat{n}_j}{\|\hat{n}_i \times \hat{n}_j\|} \cdot \hat{n} \right\| . \tag{14}
\]

where \( \hat{n}_i \) are directions towards the vertices of \( \Delta \). This is a classical result from the work of Lambert [1760] which had various applications in Computer Graphics, in particular for rendering polygonal area lights [Heitz et al. 2016; Nishita and Nakamae 1985]. A complete derivation can be found in the work of Heitz [2017] and more details...
While the previously used slope integral could be used as an efficient variation of the cosine term within triangles is often very small, in the case of an isotropic distribution with unit roughness (b) through a simple linear transformation. Then, the integral corresponds to the vertically-projected area of a spherical triangle (c), which has a simple analytic expression. We also importance sample the distribution restricted to triangles (d). These mathematical tools are at the core of our algorithms for rendering caustics and high-resolution normal maps.

3.6 Sampling slopes within triangles

While the previously used slope integral could be used as an efficient radiance estimator, doing so would introduce bias in the rendering algorithm. To use our technique for unbiased rendering, we must therefore sample a position within the triangle and evaluate the associated visibility terms and BSDFs exactly.

Uniform sampling in triangles would be an exceedingly poor sampling strategy because the energy is often concentrated in small subregions of the triangle. And half vectors align with interpolated normals. A better solution is to importance sample positions proportionally to the integrand. This can be done in slope space, and it reduces to importance sampling a slope in the distribution \( P \) restricted to the triangle \( \Delta_i \) as shown in Fig. 5d. Since there is a bijection from \( \Delta_i \) and the geometric triangle \( \Delta_i \), the sampled slope corresponds to a particular light path with known density.

We have seen that an equivalence between GGX integrals and cosine-weighted spherical triangles can be leveraged for deriving closed-form integrals in polygonal domains (Sec. 3.5). For the same reasons, importance sampling a GGX distribution restricted to a triangle in the space of slopes is equivalent to uniform sampling of vertically projected spherical triangles, as shown in Fig. 5c. While exact procedures exist for projected spherical caps [Peters and Dachsbacher 2019; Ureña and Georgiev 2018] there is no known analytic solution that achieves stratified sampling for projected spherical triangles. Ureña et al. [2000] proposed an approximate importance sampling procedure based on adaptive subdivisions of the sampling domain, while Arvo [2001] relies on the numerical inversion of a cumulative distribution function. We rely on a similar approach and provide additional details in our supplemental material. Note that this problem is distinct from the problem of sampling of spherical triangles, which has a known solution [Arvo 1995]. In practice, variation of the cosine term within triangles is often very small, in which case this sampling technique provides a useful alternative approach.

3.7 Singular distributions of slopes

The derivations in this section are valid whenever integration in slope space is feasible, which requires a bijective mapping between positions in world and slope space. This means that the triangle in slope space must not be degenerate (e.g. collapsed to a line or point). In the far-field approximation, this would be a problem, since flat or cylinder-like geometry would result in exactly such degenerate cases. During this project, we initially derived specialized integrals and sampling procedures to handle degeneracies, which introduced considerable additional complexity.

However, since the half vector typically varies across the triangle, and since we correct the slope distribution accordingly, the slope-space triangles \( \Delta_i \) are generally not singular. Some singularities could remain for scenes with both an orthographic camera and a directional emitter such as an environment map, or when the variation of slopes due to interpolated vertex normals exactly compensates the variation of slopes of the half vector, which we have not observed in our experiments.

3.8 The case of smooth conductors and dielectrics

So far, our discussion has been focused on rough conductors and dielectrics. Applying our approximations to the case of smooth materials with Dirac delta slope distributions reduces the slope-space integrals to either zero or one specular light path per triangle. The light path corresponds to the solution of the equation

\[
\tilde{n}(b_0, b_1, b_2) = \tilde{n}_{\text{linear}}(b_0, b_1, b_2),
\]

which can be solved by inverting a 2×2 matrix. The triangle contains a solution if the resulting barycentric coordinates define a point inside \( \Delta_i \). Eq. 11 reduces to

\[
\int_{\Delta_i} P_s(\tilde{n}) \, dh = \int_{\Delta_i} \delta_n(\tilde{n}) \left\| \frac{dp_s}{ds} \right\| \, ds = \left\| \frac{dp_s}{ds} \right\|. \tag{16}
\]

which uses the fact the \( \Delta_i \) is a non-singular distribution of slopes defining a bijection from surface to slopes. Although this approach is numerically stable, it is biased and it can lead to artifacts when our far-field approximation does not hold. As in previous work [Walter et al. 2009] and as discussed in section 4.6, we selectively subdivide triangles in such difficult cases. The subdivision criterion allows for controlling the error and its role is similar to error thresholds in numerical root finding techniques based on Newton iterations [Hanika et al. 2015a; Jakob and Marschner 2012; Walter et al. 2009].

4 APPLICATION TO CAUSTIC RENDERING

The derivations of section 3 provide an algorithm to connect two given scene points via an intermediate specular vertex. We now show how this algorithm can be used as an additional Next Event Estimation (NEE) strategy in path tracing-style algorithms—we refer to this approach as Specular Next Event Estimation (SNEE).

4.1 Overview

We add SNEE to the core loop of a standard unidirectional path tracer. Like classical NEE, which samples direct connections to light
Walter et al. [2009] rely on a spindle test to efficiently prune BVH
where the scene, which means that essentially all scene surfaces receive
triangles efficiently based on their position with respect to a spindle, and there is no equivalent test for pruning
problematic: the set of valid specular reflection configurations is not
restricted to a spindle, and there is no equivalent test for pruning
sub-paths. Each node of the BVH stores an axis-aligned bounding
prune large groups of triangles during the search for connection
triangles. In practice, the vast majority of triangles do not contribute
towards transport between two given
4.2 Hierarchical pruning of triangles
is detected.
V and s and the corresponding path throughputs and probability densities
are computed, and the path is rejected if occlusion
is detected.

4.3 Sampling rate for the precomputation
One limitation of this approach is that it does not strictly guarantee that the precomputed bounding boxes contain all the challenging
caustics, meaning that some specular light paths are not sampled with SNEE. Caustics that do not lie in the precomputed BVH are sampled using standard techniques to achieve unbiased results. However,
Insufficient sampling rates result in residual
outliers in rendered images, or a loss of energy if these are filtered.

4.4 Selective activation of SNEE
Since SNEE is more computationally demanding than other sampling
techniques, it is essential that it is only used when truly

sources, SNEE produces sub-paths that connect to light sources,
and the corresponding path throughputs and probability densities
can be computed to combine all available sampling strategies using multiple importance sampling.
SNEE requires a point \( p_s \) on an emitter that can be generated using any available strategy. We then sample an interaction \( p_s \) on the (near-)specular surface. The contribution of the resulting path is given by
\[
f_s((v-p_s)\cdot n_s) \cdot \frac{\|v-n_s\|}{\|p_v-p_s\|} = \frac{f_s(v,l) \cdot n_s \cdot L(p_s,-l)}{p_s \cdot p_{AL} \cdot p_1},
\] (17)
where \( p_s \) is the area density of \( p_s \) within the triangle, \( p_{AL} \) is the discrete probability of choosing triangle \( i \) (i.e., proportional to its predicted contribution), and \( p_1 \) is the probability for \( p_1 \) with respect to the solid angle measure at \( p_s \). The visibility terms \( V(p_v \leftrightarrow p_s) \) and \( V(p_s \leftrightarrow p_1) \) are computed, and the path is rejected if occlusion is detected.

4.2 Hierarchical pruning of triangles
A naive implementation of SNEE would first estimate the contributions of all specular triangles towards transport between two given points \( p_v \) and \( p_1 \) and then importance sample one of the involved triangles. In practice, the vast majority of triangles do not contribute significantly towards this estimate. Walter et al. [2009] addressed this problem by precomputing a bounding volume hierarchy to prune large groups of triangles during the search for connection sub-paths. Each node of the BVH stores an axis-aligned bounding box and cone that bound the positions and shading normals of geometry in the associated subtree. To achieve good performance, a suitable pruning criterion given two scene points \( p_v \) and \( p_1 \) is key.
Walter et al. [2009] rely on a spindle test to efficiently prune BVH nodes using the property that a single refraction can only bend the light by a limited amount, as well as other tests that entail the computation of a bounding cone of half vectors for each visited node during a traversal of the BVH.

This pruning strategy was specially designed for the refractive case, and its generalization to the case of specular reflection remains problematic: the set of valid specular reflection configurations is not restricted to a spindle, and there is no equivalent test for pruning triangles efficiently based on their position with respect to \( p_v \) and \( p_1 \). Reflective objects tend to scatter light towards the entirety of the scene, which means that essentially all scene surfaces receive

low-energy caustics. Although they barely contribute to the pixel value, such caustics still produce considerable variance when standard BSDF and emitter sampling techniques are used. At the same time, this means that pruning becomes largely ineffective, as range queries will tend to find many candidate triangles with a valid specular configuration. Finally, the method also lacks awareness about global visibility, i.e., it finds triangles with the right orientation even if occlusion prevent them from being visible from \( p_v \), \( p_1 \), or both. We previously experimented with combinations of SNEE with the pruning hierarchy of Walter et al. [2009] and found that the drawbacks listed above make such an approach unnecessarily costly for the case of specular reflection.

We address all these problems by precomputing a different type of hierarchical data structure that enables fast queries for candidate triangles, whose contribution we subsequently evaluate using SNEE.
The data structure is generated in a brief pre-rendering step, which entails tracing rays from the light sources towards uniformly sampled positions in each specular triangle. We reflect and refract these rays and compute an axis-aligned bounding box containing all intersections with the non-specular scene geometry (Fig. 6) together with the associated amount of energy. In the case of refractive geometry, we separately bound reflected and refracted rays. Next, the per-triangle bounds are assembled into a weighted bounding volume hierarchy (BVH), where each node records the maximum expected irradiance from a specular triangle. Tree construction proceeds top-down, using the builder provided in PBRT [Pharr et al. 2016].

Following this step, finding specular triangles that participate in caustic light transport at a surface position reduces to a hierarchical range query. Additionally, the irradiance values stored within BVH nodes enable an efficient stochastic sampling procedure that focuses computation on salient parts of the domain in the case where many triangles contribute.

A limitation of this approach is that it does not strictly guarantee that the precomputed bounding boxes contain all the challenging caustics, meaning that some specular light paths are not sampled with SNEE. Caustics that do not lie in the precomputed BVH are sampled using standard techniques to achieve unbiased results. However, it is crucial that the precomputed per-triangle bounds contain all the light paths that produce outliers using standard techniques.

A sufficient number of rays thus have to be traced in the pre-computation phase to account for difficult visibility, non-singular light sources and material roughness. In our implementation, we use 500 samples per triangle for smooth materials and up to 2k samples for rough materials to ensure a dense sampling of microfacet distributions. In practice, we found that this approach is robust—in particular, no filtering or post-processing of any kind was used in any of our results. Insufficient sampling rates result in residual outliers in rendered images, or a loss of energy if these are filtered.
needed. The per-triangle bounds should be as compact as possible, and they should be restricted to those paths that add significant variance when using standard sampling techniques. Construction of the BVH therefore ignores paths, whose luminance using standard BSDF/emitter sampling techniques falls below a user-specified threshold. This parameter is intuitive: it is a bound of the maximum desired value of outlier samples. This strategy enables targeted application of SNEE to truly challenging specular geometry and light sources. When the illumination is represented using an environment map, this strategy can for instance restrict SNEE to directions that map to the sun or other bright light sources encoded in the environment map image. Standard sampling techniques are used for all the specular sub-paths that are not sampled with SNEE to ensure unbiasedness of the resulting estimator.

4.5 Stochastic pruning at render time

We use our pruning data structure to query candidate specular triangles for a shading point at render time. Then, we importance sample one triangle based on its predicted contribution as well as a specular interaction inside the triangle (Sec. 3). Our method achieves nearly-perfect importance sampling of specular light paths if all the candidate triangles are queried. This is also the approach of Walter et al. [2009]. However, this requires estimating the contribution of all the candidate triangles independently of their contribution to the image. This is generally suboptimal, in particular in scenes with low-energy caustics as discussed in Sec. 4.2. This can be improved by stochastically pruning nodes during the traversal of the BVH. We define the probability of exploring a node using

$$pdf_{\text{node}} = \min \left( 1, \frac{4_{\text{max}}}{l} \right),$$  

Fig. 7. Equal time rendering (2 minutes) of a unidirectional path tracer with our SNEE compared to other unbiased algorithms. Top: a challenging scene with smooth and rough conductors. Bottom: a slightly modified version of the swimming pool scene. We reduced the resolution of the water geometry from 1.8M to 120k triangles, and we use a thin dielectric approximation for the refraction through the windows. SNEE robustly finds SDS paths that produce outliers with standard BSDF and emitter sampling techniques. Metropolis Light Transport and Manifold Exploration (MLT+ME) efficiently explore some regions of the path space but images suffer from typical large-scale artifacts and incorrect brightness. We also compare SNEE with the improved version of the Practical Path Guiding (PPG) algorithm [Müller 2019; Müller et al. 2017]. This algorithm performs well in the case of refraction but offers little benefits compared to standard path tracing for reflection, due to the higher frequency of the light transport in the spatial and directional domain for such caustics. SNEE produces low-variance estimates for sub-paths with one specular interaction but it does not sample longer specular chains, which explains the remaining outliers, e.g., inter-reflection between the vases (top) or between the water and the windows (bottom). The rendering time includes training passes for PPG and the precomputation passes for SNEE, which took 40s (top) and 20s (bottom).
where \( I_{\text{max}} \) is the precomputed maximum expected irradiance of the node, and \( T \) is a user-specified threshold. This effectively enables adaptive stochastic pruning while preventing weighted path contributions from being excessively large.

### 4.6 Triangle subdivision

Our derivations in section 3 rely on far-field assumptions that are only accurate when \( p_v \) and \( p_l \) are sufficiently distant from the triangle (relative to the scale of the triangle itself). In cases where one of the points (typically \( p_v \)) is close to the triangle, we recursively subdivide the triangle until the distance between the points and the sub-triangles is \( N \) times larger than the size of the sub-triangles. We use values of \( N \) between 5 and 20 depending on the roughness of the material (a higher precision is needed for small values of \( \alpha \). We also define a maximum depth of 20 for the subdivisions and ignore the sub-triangles when this maximum is reached. In practice, for high resolution meshes and distant lights, subdivision only occurs in a small number of pixels, e.g., when a specular object directly placed onto a floor that receives caustics. In general, the subdivision does not have a strong effect on overall runtime performance.

### 4.7 Implementation and results

We implemented our algorithm in Mitsuba 2 [Nimier-David et al. 2019]. All images in this article were rendered on an Intel i9-7960X workstation.

**Equal time comparisons.** Fig. 7 compares SNEE to other unbiased rendering algorithms on difficult scenes with specular geometry and small light sources. SNEE robustly samples SDS paths that produce outliers in both uni- and bidirectional path tracing. SNEE also outperforms the guided path tracer introduced by Müller et al. [2017] for reflected caustics. Robustly guiding light paths for such caustics requires very high resolution spatial and directional data structures that are not robustly trained due to the tiny probability of sampling important light paths. However, this path guiding approach tends to perform better for reflective caustics due to a lower cost per path compared to SNEE, at the cost of a larger memory footprint. Metropolis Light Transport and Manifold Exploration (MLT+ME) achieve good local exploration but the images show incorrect brightness and rendering artifacts at these low sample counts.

**Comparison with MNEE.** Fig. 8 shows comparisons between Manifold Next Event Estimation (MNEE) and our approach. Both techniques provide an additional connection strategy in path tracers, but they behave very differently. MNEE only supports a unique and deterministic initialization for the manifold walk for each pair of endpoints, otherwise the probability of finding a particular path lacks a closed-form expression. As a consequence, MNEE can only sample one highlight for each pair of endpoints and it offers limited benefits for rendering challenging caustics. MNEE is also restricted to refraction since it is unclear what would be a good initialization for reflection. In contrast, SNEE finds all the important light paths for reflection and refraction, but it is restricted to sub-paths with exactly one specular interaction.

**Comparison with Walter et al. [2009].** We provide comparisons with the pruning strategy of Walter et al. [2009] in Fig. 9. Our approach is significantly faster due to the combined effects of stochastic sampling and our improved approach to hierarchical pruning.
\[ \alpha_{ggx} = 0.001 \quad \alpha_{ggx} = 0.01 \quad \alpha_{ggx} = 0.05 \]

(a)

PT – 55 spp, 1s
PT – 165 spp, 3s
PT – 495 spp, 9s
PT – 265 spp, 6s
PT – 265 spp, 6s
PT – 530 spp, 12s

(b)

PT+SNEE – 10 spp, 1s
PT+SNEE – 10 spp, 3s
PT+SNEE – 10 spp, 9s
PT+SNEE – 10 spp, 6s
PT+SNEE – 10 spp, 6s
PT+SNEE – 10 spp, 12s

Fig. 10. Various scenes rendered with equal sample counts for PT+SNEE (b, our), and equal time comparison against standard path tracing (a). These results show how the cost of SNEE increases with the roughness of the material and the complexity of the caustics. Images in (a) and (b) only contain contributions from light paths that can be sampled by SNEE.

The performance difference is particularly large in the case of reflective caustics. Our approach achieves similar levels of variance at equal sample counts despite the stochastic pruning of specular triangles.

4.8 Discussion

The cost of SNEE queries depends on the number of triangles that contribute non-negligible energy. Therefore, SNEE scales well with the number of triangles in the scene if the number of connection paths remains reasonable. For instance, SNEE can efficiently sample the important light paths in underwater caustics like in Fig. 1b even if the ocean is modeled using a large-scale high-resolution mesh, since only a few specular paths contribute to each shading point. However, importance sampling indirect reflection from complex bumpy surfaces like the ones in Fig. 1a is less efficient, since many bumps redirect light towards a single point on the table. The cost of SNEE scales linearly with the number of light sources that produce caustics, since one acceleration structure must be computed for each of these lights.

5 SLOPE-SPACE INTEGRALS FOR HIGH-RESOLUTION NORMAL MAPS

Specular surfaces with high-resolution normal maps often result in highly realistic glinty appearance, but rendering of such surfaces using standard pixel sampling is impractical when the light source subtends a small solid angle. Glints are extremely bright sub-resolution highlights, whose size is typically much smaller than normal map texels— which themselves project to small fractions of a pixel.

Yan et al. [2014] addressed this problem by evaluating averaged normal distributions based on specified filter kernels on the specular surface. They derived their filtering equations using normals following orthogonal projection on the tangent plane, and they define the distribution of microfacets of the BRDF as a 2D Gaussian distribution in this space. Unfortunately, texel contributions are given by integrals that do not have closed-form expressions. In this section, we reformulate such filtering integrals for normal maps in slope space, and we show that the problem of importance sampling highlights in a normal map is very similar to the problem of importance sampling highlights on triangle meshes (Sec. 3). Our reformulation of previous work in the space of microfacet slopes brings several benefits that we discuss in this section. We note that transitioning from projected normals to surface slopes does not require fundamental changes to the hierarchical data structure used to implement NDF queries [Yan et al. 2014].

5.1 Filtering integrals for normal maps

The filtering integral for normal maps is given by

\[ \int_S W(p_s) f_{\alpha}(v, l) |l \cdot n| \, dp_s. \]
where $W$ is a normalized filter kernel on the surface $S$, and $f_d(v, l)$ denotes BRDF at the point $p_i$. The vectors $v$ and $l$ are the directions towards the camera ($p_x$) and a sampled position on a light source ($p_l$). A suitable filter $W$ can be determined using ray differentials [Ighey 1999] or covariance tracing [Belcour et al. 2017]. As in the work of Yan et al. [2014], we assume that the roughness of the base material is not spatially varying, and that the normals of the map are interpolated within texture-space triangles. We can split Eq. 19 into sub-integrals over texture-space triangles:

$$
\sum_{\Delta_i \in T} \int_{\Delta_i} W(t) f_d(v, l) |l \cdot n| \left\| \frac{\partial p_x}{\partial t} \right\| dt,
$$

(20)

where $t$ is a position in the texture space $T$, $\Delta_i$ is a triangle in the normal map and $\left\| \frac{\partial p_x}{\partial t} \right\|$ is the Jacobian determinant of the transformation between areas in world space and texture space. Since we use barycentric interpolation of normals in triangles $\Delta_i$, there are bijections between texture coordinates and slopes that allow turning the integral in Eq. 20 into a slope-space integral:

$$
\int_{\Delta_i} W(\tilde{s}) f_d(v, l) |l \cdot n| \left\| \frac{\partial p_x}{\partial s} \right\| ds
$$

(21)

This integral resembles the main result of section 3; it reduces to the evaluation of the contribution due to a single light path in the case of smooth conductors (Sec. 3.8), or it can be importance sampled when the material is rough (Sec. 3.6).

Compared to the work of Yan et al. [2014], our approach benefits from well-known masking-shadowing terms for GGX distributions and closed-form expressions for integrals over triangular domains. However, our framework cannot account for the spatially varying weight $W(\tilde{s})$ in each triangle in the normal map. We found that this is not a problem in practice since the triangles are much smaller than the filtering kernel (we are filtering high-resolution normal maps) and variations of $W(\tilde{s})$ inside triangles can be neglected when importance sampling highlights.

5.2 Accounting for varying half vectors and curvature

We can apply the techniques described in sections 3.2 and 3.4 to improve the accuracy of the filtering integral, while at the same time avoiding potential issues due to singular slope distributions (Sec. 3.7). We correct the distributions in each normal map triangle $\Delta_i$ using the derivatives of the shading normal at the center of the filtering kernel:

$$
\tilde{n}_i(t) = \frac{\partial \tilde{n}}{\partial u}(t_u - c_u) + \frac{\partial \tilde{n}}{\partial v}(t_v - c_v),
$$

(23)

where $c_u$ and $c_v$ are the texture-space coordinates of the center of the filter kernel. We also estimate the half vector variations in texture space:

$$
\tilde{h}(t) = \tilde{h}_c + \frac{\partial \tilde{h}}{\partial u}(t_u - c_u) + \frac{\partial \tilde{h}}{\partial v}(t_v - c_v),
$$

(24)

where $\tilde{h}_c$ correspond to the slopes of the half vector at the center of the filter kernel. The contribution of the normal map triangle $\Delta_i$ can be computed by integrating the distribution of slopes between the points

$$
\tilde{h}(t_j) - \tilde{n}_i(t_j) - \tilde{n}_j, \quad j \in \{0, 1, 2\}
$$

where $t_j$ are the texture coordinates of the vertex $j$ in $\Delta_i$ and $\tilde{n}_j$ are the corresponding slopes from the normal map.

The supplemental material showcases several animated results created using our slope-space representation, while accounting for surface curvature. These results demonstrate the good stability of the highlights even in challenging cases with disabled normal interpolation or highly-curved geometry, which produce slope distributions that are almost singular.

5.3 Pruning tests for texels

As in previous work, we rely on a quad-tree hierarchy to prune large regions of the normal map when integrating the response over a particular filter kernel. Each tree node corresponds to a rectangular region of the normal map and specifies a 2D box that bounds the slopes of underlying texels. We modified pruning tests to account for surface curvature and half vector variations inside the filter footprint. In the case of rough materials, we only consider clamped distributions of slopes during the queries. We account for the missing energy by sampling positions in the filter kernel.
and by keeping the contributions that correspond to the tail of the distribution. As shown in Fig. 11, this improves the performance of the queries compared to conservative truncation of the distribution, while ensuring that the entire distribution is correctly sampled.

5.4 Generalization to smooth materials

Our new and simpler solution that is specific to perfectly smooth materials (section 3.8) leads to an algorithm that runs about twice as fast as glints with small amounts of roughness, while producing similar appearance as shown in Fig. 12. This speedup is a consequence of more aggressive pruning that becomes possible as bounding boxes do not have to be expanded to account for the intrinsic roughness of the material. The evaluation of the filtered BRDF is also faster since evaluating a highlight in a texel for the intrinsic roughness of the material. The evaluation of the possible as bounding boxes do not have to be expanded to account for singularities in previous formulations of glint filtering, leading to a generalization to ideally specular materials that runs faster than the case of materials with tiny amounts of roughness. The main limitation of our method is its restriction to specular chains with a single vertex, hence other strategies remain necessary to handle specular transport with multiple interactions. Addressing this significantly more challenging case is an important topic for future work.

6 CONCLUSION

We introduced a new technique to sample specular reflection or refraction events connecting two given scene points, typically a shading point and a light source. Our method, named Specular Next Event Estimation, can supplement standard sampling techniques in path tracing-like algorithms and greatly improves the convergence of supported path classes, extending the scope of unidirectional rendering to scenes containing caustics and specular-diffuse-specular light paths. Building on a closed-form solution of a slope-space integral, our method supports surfaces ranging from ideally specular to near-specular and even highly rough materials, although standard techniques may be preferable in the latter case. Our first-order approximation of half vector variations largely eliminates problems raised by singularities in previous formulations of glint filtering, leading to a generalization to ideally specular materials that runs faster than the case of materials with tiny amounts of roughness.

REFERENCES
