Microfacet-based Normal Mapping for Robust Monte Carlo Path Tracing

VINCENT SCHÜSSLER, Karlsruhe Institute of Technology
ERIC HEITZ, Unity Technologies
JOHANNES HANIKA, Karlsruhe Institute of Technology
CARSTEN DACHSBACHER, Karlsruhe Institute of Technology

Classic Normal Mapping
24 seconds

Microfacet-Based Normal Mapping (ours)
27 seconds

Fig. 1. An image rendered at 256spp with classic normal mapping and our microfacet-based normal mapping (default settings). With classic normal mapping, back-facing normals (red) and energy leaks (green) produce artifacts such as black fringes and it can be challenging to obtain the desired appearance without hacking the data or the light transport algorithm. Our new normal-mapping model makes it possible to obtain a detailed reflective sphere not covered by these black regions without breaking light transport.

Normal mapping enhances the amount of visual detail of surfaces by using shading normals that deviate from the geometric normal. However, the resulting surface model is geometrically impossible and normal mapping is thus often considered a fundamentally flawed approach with unavoidable problems for Monte Carlo path tracing, such as asymmetry, back-facing normals, and energy loss arising from this incoherence. These problems are usually sidestepped in real-time renderers, but they cannot be fixed robustly in a path tracer: normal mapping breaks either the appearance (black fringes, energy loss) or the integrator (different forward and backward light transport); in practice, workarounds and tweaked normal maps are often required to hide artifacts.

We present microfacet-based normal mapping, an alternative way of faking geometric details without corrupting the robustness of Monte Carlo path tracing. It takes the same input data as classic normal mapping and works with any input BRDF. Our idea is to construct a geometrically valid microfacet surface made of two facets per shading point: the one given by the normal map at the shading point and an additional facet that compensates for it such that the average normal of the microsurface equals the geometric normal. We derive the resulting microfacet BRDF and show that it mimics geometric detail in a plausible way, although it does not replicate the appearance of classic normal mapping. However, our microfacet-based normal mapping model is well-defined, symmetric, and energy conserving, and thus yields identical results with any path tracing algorithm (forward, backward, or bidirectional).

CCS Concepts:
• Computing methodologies → Reflectance modeling

Additional Key Words and Phrases: Normal mapping, path tracing, microfacet theory

ACM Reference format:

1 INTRODUCTION

In this paper, we use the term normal mapping to describe the set of techniques that compute the shading of a surface with a perturbed normal instead of the actual geometric normal. The normal is usually provided by a high-resolution texture and allows for increasing the visual complexity of a surface without increasing the complexity of its geometry. A whole body of techniques and representations...
originated from the original work on bump mapping [Blinn 1978] and they are ubiquitous today in both video games and offline rendering engines. In this paper, we focus on their application for path tracing.

Use of Normal Mapping in Path Tracers. Some offline renderers use tessellation and displacement mapping for geometric details, but normal mapping remains a cheaper alternative that is still widely used in the offline rendering industry, especially for movies. This can be seen in the documentation of path tracers used for major productions, for example Pixar’s Renderman [Pixar 2015], Solid Angle’s Arnold Renderer [SolidAngle 2016], or Guerilla Render [Mercenaries 2016].

Problems of Normal Mapping. Despite its wide adoption in practice, normal mapping remains a “prehistoric” tool in modern frameworks of rendering, inherited from a period where mathematical coherence was less important for rendering. Physically based rendering is obtained by deriving a model of light transport within an appropriate mathematical framework [Veach 1997] which has to form a mathematically consistent whole. The consistency is particularly important for Monte Carlo path tracing where different techniques are often combined together to improve variance reduction. Symmetry and energy conservation are important constraints for light transport to be well-defined and consistent.

The problem of normal mapping is that it breaks this consistency by definition: it fakes the light transport of a surface by replacing its geometric normal with another perturbed normal. Faking light transport in this way affects both its symmetry and energy conservation. We review the classic problems of normal mapping and their consequences in Section 3.

An Underexplored Problem. These problems are well-known by practitioners who cannot use normal maps in path tracers without making compromises. In order to reduce artifacts, either normal maps are tweaked by artists or workarounds are made in the integrators that prevent the use of advanced rendering algorithms such as bidirectional path tracing. Veach [1996] proposed a way to fix the broken symmetry due to shading normals for bidirectional path tracing. However, the other problems, illustrated for a detailed sphere in Figure 1, remain unsolved, and they are the ones that result in visible artifacts. It is just impossible to obtain a correct or plausible appearance with classic normal mapping and the only way is to replace the normal map by a height map and use tessellation. However, this is not a viable solution: the goal of normal mapping is to have a cheap alternative without geometric or ray tracing overhead.

Position of Our Work. To our knowledge, there is no more recent work dedicated to these fundamental problems of normal mapping, although they are well-known and acknowledged to be painful by practitioners. One reason might be the habitual assumption that faking geometric details cannot come without breaking light transport. The objective of this paper is to demonstrate that it is actually possible to model fake geometric details over a flat surface with a correct light-transport formulation. We propose a simple microfacet model to be used as a replacement for a shading normal, and we derive its masking-shadowing function (Section 4) and its multiple-scattering BRDF with a random-walk approach (Section 5). With this alternative definition of normal mapping, the light transport of the surface is at the same time symmetric (works with bidirectional techniques) and energy conserving (no black fringes nor energy loss). In Section 6 we show that multiple modeling choices are possible and exemplarily explore two of them. In Section 7 we validate our model and propose one configuration with good properties to be used as default.

We call the realization of this idea Microfacet-Based Normal Mapping which has the following properties:

- It fixes all the issues of classic normal mapping: asymmetry, black fringes, and violated energy conservation. To our knowledge, it is the first normal mapping model that does not cause problems with Monte Carlo path tracing. As a consequence, it does not replicate the appearance of classic normal mapping, mainly because of lower contrast due to its improved energy conservation. However, our model produces plausible results even in cases where classic normal mapping fails.
- Like classic normal mapping, it is implemented as a simple wrapper that takes an arbitrary normal map and an arbitrary BRDF (potentially with textured parameters) as input.
- Our default microfacet-based normal mapping model comes with optional parameters that provide control over the saturation and the contrasts of the normal-map appearance. In the worst case, it is about 70% costlier than classic normal mapping; in realistic scenes, its cost is largely amortized and the overhead is about 10%.
- The general model is evaluated stochastically and thus leads to additional variance in the renders. For some integrators it is also problematic that the model has no analytic sampling probability density function (PDF). Most notably, our model cannot be used with Metropolis Light Transport out-of-the-box. We discuss this issue and possible solutions further in Section 7.3.

2 RELATED WORK

Microfacet Theory. In Section 4 we propose the tangent facet microsurface model which is a variant of the V-cavity model introduced by Torrance and Sparrow [1967]. Our derivations are based on the work of Heitz [2014] who reviewed the equations that should be satisfied by microfacet models and their masking-shadowing functions to be geometrically correct. We also build upon the random-walk algorithm by Heitz et al. [2016] that they use for evaluating the multiple-scattering BRDF of surfaces based on the Smith model.

Mathematically Correct Fake Geometric Details. Our idea is conceptually similar to the model of Raymond et al. [2016] for textured scratches: they precompute several scratch-BRDFs and apply them...
over the surface. The resulting light transport is well-defined because the average normal of the scratches is exactly the geometric normal, i.e. the geometry of the fake details is consistent with the surface geometry. In contrast, our model uses a normal map to model scratches shaped as simple V-cavities from which we compute a valid BRDF on the fly. Our V-cavity model does not require any precomputation and we can thus use it for the general case where arbitrary normal maps and BRDFs are given as input.

Normal Map LODs. Like other textures, normal maps are usually prefiltered in order to avoid aliasing and reduce the memory footprint when the details are seen from far away. LOD representations for classic normal mapping are based on the same idea: the normal map details are merged into BRDF parameters, typically the “roughness” for glossy materials [Han et al. 2007; Toksvig 2005]. Some techniques use the same idea in other spaces, for instance LEAN Mapping [Dupuy et al. 2013; Olano and Baker 2010] represents the data as a Gaussian in slope space. In this case, the average slope yields a normal, and the covariance of the Gaussian yields the anisotropic roughness parameters of the BRDF. Since our model works with any normal map and any BRDF (and its parameters), it is compatible with all the aforementioned LOD representations.

Glint Rendering. Microfacet-based normal mapping is not relevant for the very specific case of glint rendering, which usually does not suffer from the black fringes and energy conservation problems of classic normal mapping: Glint rendering integrators are required for highly-specular subpixel-normal-mapped details [Yan et al. 2014, 2016] and in this case, large normal map footprints are considered for the glints where possibly occuring back-facing normals are virtually outweighed. Note that glint integrators only work for forward path tracing as the calculation of the footprints violates symmetry by definition. As we focus on fixing all problems (including symmetry), our microfacet-based approach is meant as a replacement for cases where classic normal mapping is applied.

3 PROBLEMS CAUSED BY NORMAL MAPPING

Classic normal mapping replaces the geometric normal by a perturbed normal for shading. However, this results in inconsistencies between the actual geometry and the one defined by the shading normals. In particular, this leads to non-symmetric scattering and violates the conservation of energy as described by Veach [1996].

3.1 Non-Symmetry Due to Shading Normals

The use of shading normals corresponds to a non-symmetric modification of the BRDF. This becomes easily visible when using bidirectional methods, e.g. bidirectional path tracing (see Figure 14).

To show this non-symmetry, let us first consider an intersection of an eye subpath with a normal mapped surface shown in Figure 2. For both eye and light subpaths, we refer to the eye direction as $\omega_i$ and to the light direction as $\omega_l$. Let $f_{\omega_l,o}$ be the BRDF evaluated with respect to the shading normal $\omega_o$ instead of the geometric normal $\omega_g$.

To achieve the desired effect of a shading normal, the foreshortening factor for $\omega_g$ must use $\omega_i$ instead of $\omega_l$, leading to the modified BRDF $f$:

$$\tilde{f}(\omega_i, \omega_l) = f_{\omega_l, o}(\omega_i, \omega_l) \frac{\langle \omega_l, \omega_o \rangle}{\langle \omega_o, \omega_g \rangle} \frac{\langle \omega_o, \omega_l \rangle}{\langle \omega_l, \omega_g \rangle}.$$  (1)

This modified BRDF is in general not symmetric, even if the original BRDF $f$ was symmetric.

For bidirectional rendering it would be necessary to efficiently sample the adjacent BRDF, but we consider this to be impractical. When constructing a light subpath, we can usually only importance sample $f_{\omega_l, o}(\omega_i, \omega_l)\omega_i\omega_o\omega_g$ instead of the modified BRDF to determine $\omega_l$. Consequently, we need to reweight the contribution:

$$\tilde{f}(\omega_i, \omega_l) = f_{\omega_l, o}(\omega_i, \omega_l) \frac{\langle \omega_l, \omega_o \rangle}{\langle \omega_o, \omega_g \rangle} \frac{\langle \omega_o, \omega_l \rangle}{\langle \omega_l, \omega_g \rangle}.$$  (2)

When $\omega_o$ is nearly perpendicular to $\omega_l$, or $\omega_i$ is nearly perpendicular to $\omega_o$, the weighting factor can get arbitrarily large. This causes high variance in renderings and does not solve the problem in practice1.

3.2 Black Fringes on Surfaces

Shading normals tilt the positive hemisphere of outgoing directions, leading to inconsistencies with the geometric hemisphere, as shown in Figure 3. The contributions of all directions that lie on any of the two negative hemispheres must be discarded. This causes black fringes to appear when the viewing direction is below the hemisphere around the shading normal. It also leads to considerable darkening of the surface, because of the reduced hemisphere of outgoing directions. This is an additional cause of black fringes for highly specular materials when the reflected viewing direction is below the surface.

Veach suggests to extend the BRDF to the entire sphere to eliminate undefined regions of the geometric hemisphere. While this may be reasonable for simple cases like a diffuse BRDF, it is unclear how this extension should be defined for others.

![Fig. 2. Intersection of an eye subpath. A ray from the incident direction $\omega_i$ leaves the surface in the outgoing direction $\omega_o$, which is determined by importance sampling $f_{\omega_l, o}(\omega_i, \omega_l)\omega_i\omega_o\omega_g$.](image)

![Fig. 3. Problems caused by normal mapping. With a naive approach, light can leak through the surface when evaluating the BRDF with respect to the shading normal (left). Furthermore, the BRDF is undefined for directions below the tilted hemisphere defined by the shading normal (right).](image)

1 In the Mitsuba renderer, Veach’s correction factor for shading normals was disabled because it comes with unacceptable variance. See https://www.mitsuba-renderer.org/repos/mitsuba/changeset/b766685c429.
3.3 Violation of Energy Conservation

Energy conservation imposes the following constraint on the modified BRDF:

\[
1 \geq \int f(\omega_i, \omega_o)(\omega_i \cdot \omega_o) \, d\omega_i = \frac{\dot{\omega}_p \cdot \dot{\omega}_s}{\omega_o \cdot \omega_o} \cdot \int f_{\omega_o}(\omega_i, \omega_o)(\omega_i \cdot \omega_o) \, d\omega_i
\]

We notice that the factor \(\frac{\dot{\omega}_p \cdot \dot{\omega}_s}{\omega_o \cdot \omega_o}\) can be moved outside the integral. Note that this factor can become arbitrarily large for outgoing directions nearly perpendicular to the geometric surface. Consequently the constraint is not satisfied and the use of shading normals violates the conservation of energy.

3.4 Standard Practices for Preventing Artifacts

We compare our model to two techniques that are commonly used in practice to hide the black fringe artifacts of classic normal mapping. Both methods define the BRDF for incident directions below the positive shading hemisphere:

- The **flipping technique** (Figure 4 left) flips the direction of the shading normal. In this way, incident directions always lie in the positive hemisphere of the shading normal and no incident directions are undefined.
- The **switching technique** (Figure 4 right) switches to an alternative, arbitrarily chosen BRDF. We use a diffuse BRDF with albedo \(\rho = 0.5\).

To complement the described techniques, we implement a technique suggested by Keller et al. [2017] that mitigates the effect of sampled directions lying below the geometric hemisphere. It works by changing the shading normal such that the reflection vector of the incident direction is always in the positive geometric hemisphere. While this does not prevent invalid directions from being sampled, it ensures that specular reflections do not go below the surface.

All of these techniques modify the BRDF in a way that further amplifies issues with energy conservation and symmetry of classic normal mapping. Further, sampling the adjoints of these techniques is problematic, because the hemisphere (flipping, [Keller et al. 2017]) or the BRDF (switching) may change depending on \(\omega_i\). When tracing from the light, we need to sample a \(\omega_i\) for a given \(\omega_o\). Since the hemisphere and BRDF are not fixed before sampling \(\omega_i\), choosing a good sampling strategy for \(\omega_i\) is difficult.

4 MICROSURFACE MODEL

In order to solve the problems caused by normal mapping, we propose a microfacet model as an alternative to shading normals. Instead of just replacing the geometric normal for lighting calculations, the perturbed normal determines the orientation of microfacets. In this way, the apparent orientation of the surface is changed locally, while maintaining validity of the BRDF through the application of microfacet theory.

4.1 Choice of the Microsurface Profile

We choose a model in which the microsurface is composed of tiny asymmetric cavities consisting of two opposing facets: While the first facet is oriented by the perturbed normal \(\omega_p\), the other one is perpendicular to the geometric surface and its corresponding micro-normal \(\omega_t := \omega_p(\omega_p)\) is tangent to the geometric surface. Figure 5 shows the resulting microsurface profile, which is akin to the V-cavity model by Torrance and Sparrow [1967] yet with asymmetric cavities.

![Fig. 5. Microsurface profile. We add a tangent facet \(\omega_t\) that compensates for the perturbed normal such that the average normal of the microsurface (\(\omega_p\) and \(\omega_t\) weighted by the areas of the respective facets) remains the geometric normal.](image)

We have chosen this configuration because it is the configuration with minimal area that is geometrically correct and produces a look similar to the desired normal map, i.e. where the desired shading normal \(\omega_p\) has the dominant influence on the appearance:

- It satisfies the geometric constraint that the average normal of the microsurface is the normal of the geometric surface.
- The distribution of microfacet normals is not smooth: note that the final appearance of the surface is a result of the microsurface and the input BRDF. Any smooth distribution would act as a low-pass filter on the input BRDF and thus specular appearance would not be possible. If the microsurface remains discrete (Dirac delta distributions) then high frequencies are preserved and the resulting BRDF is similar to the input BRDF.
- This configuration maximizes the surface area with the desired shading normal \(\omega_p\) and minimizes the remaining area (in this case the surface area with \(\omega_t\)).
- Lastly, in this configuration the tangent facet \(\omega_t\) is invisible to half of the incident directions onto the macro-surface, i.e. it does not interact with these directions directly.

4.2 Distribution of Normals

The distribution of normals (NDF) describes the statistical distribution of the microfacets’ normals [Heitz 2014]. This is equal to the area of the respective microfacets of a unit surface, which we derive directly from the microsurface profile. Since the model consists...
of only two different micro-normals, the distribution of normals becomes:

\[
D(\omega_m) = \frac{\delta_{\omega_p}(\omega_m)}{\omega_p \cdot \omega_g} + \frac{\delta_{\omega_t}(\omega_m) \sqrt{1 - (\omega_p \cdot \omega_g)^2}}{\omega_p \cdot \omega_g}.
\]  

(4)

This NDF satisfies the geometric constraint that its projected area onto the geometric normal should equal one [Heitz 2014]:

\[
\int_{\Omega} D(\omega_m) |\omega_m \cdot \omega_g| \, d\omega_m = 1.
\]  

(5)

### 4.3 Projected Areas and Intersection Probabilities

The projected areas of the facet oriented by \(\omega_p\) and the tangent facet onto an incident direction \(\omega_i\) are determined as

\[
a_p(\omega_i) = \frac{\langle \omega_i, \omega_p \rangle}{\langle \omega_p, \omega_g \rangle},
\]  

(6)

\[
a_t(\omega_i) = \frac{\langle \omega_i, \omega_t \rangle \sqrt{1 - \langle \omega_p, \omega_g \rangle^2}}{\langle \omega_p, \omega_g \rangle},
\]  

(7)

and the probabilities that a ray incident on direction \(\omega_i\) intersects \(\omega_p\) or \(\omega_t\), respectively, are proportional to the projected areas:

\[
\lambda_p(\omega_i) = \frac{a_p(\omega_i)}{a_p(\omega_i) + a_t(\omega_i)},
\]  

(8)

\[
\lambda_t(\omega_i) = \frac{a_t(\omega_i)}{a_p(\omega_i) + a_t(\omega_i)}.
\]  

(9)

Lastly, the distribution of visible normals (VNDF) [Heitz 2014] is:

\[
D_{\omega_i}(\omega_m) = \lambda_p(\omega_i) \delta_{\omega_p}(\omega_m) + \lambda_t(\omega_i) \delta_{\omega_t}(\omega_m).
\]  

(10)

### 4.4 The Masking Function

The masking function \(G_1(\omega_i, \omega_m)\) determines the fraction of micro-facets with normal \(\omega_m\) that are visible from direction \(\omega_i\). Similar to the Cook-Torrance V-cavity masking function [Heitz 2014] our microsurface profile has only two normal cases that can occur: (1) either both facets (with normal \(\omega_p\) and \(\omega_t\)) are fully visible and the masking function equals 1; or (2) one of them is backscattering (\(a_p(\omega_i)\) or \(a_t(\omega_i)\) equals 0), and the other one is partially masked (Figure 6).

This enables us to derive \(G_1(\omega_i, \omega_m)\): first of all, it has to fulfill the property of conservation of the projected area [Heitz 2014]:

\[
\omega_i \cdot \omega_g = \int_{\Omega} G_1(\omega_i, \omega_m)(\omega_i, \omega_m) D(\omega_m) \, d\omega_m,
\]  

(11)

which for our microsurface simplifies to

\[
\langle \omega_i, \omega_g \rangle = G_1(\omega_i, \omega_p) a_p(\omega_i) + G_1(\omega_i, \omega_t) a_t(\omega_i).
\]  

(12)

This condition is met by the following masking function which can conveniently be expressed in a single formula:

\[
G_1(\omega_i, \omega_m) = H(\langle \omega_i, \omega_m \rangle) \min \left[ 1, \frac{\langle \omega_i, \omega_g \rangle}{a_p(\omega_i) + a_t(\omega_i)} \right],
\]  

(13)

where \(H(\cdot)\) is the Heaviside function.

### 4.5 The Masking-Shadowing Function

For simplicity, we opted for the separable masking-shadowing function given by Equation 14 which assumes no correlation of masking and shadowing:

\[
G_2(\omega, \omega_0, \omega_m) = G_1(\omega, \omega_m)G_1(\omega_0, \omega_m).
\]  

(14)

This choice is motivated by the simple random-walk approach used to compute the multiple-scattering BRDF of the microsurface and is explained in the next section.

### 5 THE BRDF OF THE MICROSURFACE

In this section, we derive the BRDF of the microsurface given the BRDFs of the perturbed and tangent facets \(f_p\) and \(f_t\), respectively.

#### 5.1 The Single-Scattering BRDF

The contribution of facets with a single micro-normal \(\omega_m\) and BRDF \(f_m\) to the BRDF of the entire surface can be obtained from the single-scattering microfacet BRDF [Heitz 2014]:

\[
f(\omega_i, \omega_0, \omega_m) = f_m(\omega_i, \omega_0, \omega_m)G_2(\omega, \omega_0, \omega_m)D(\omega_m).
\]  

(15)

The value of the single-scattering BRDF \(f_1\) is the integral of the contribution of each micro-normal over the hemisphere:

\[
f_1(\omega, \omega_0) = \int_{\Omega} f(\omega, \omega_0, \omega_m) \, d\omega_m.
\]  

(16)

In our case, the cosine-weighted BRDF expression simplifies to

\[
f_1(\omega, \omega_0) \langle \omega_i, \omega_g \rangle = \lambda_p(\omega_i) f_p(\omega_i, \omega_0) \langle \omega_p, \omega_0 \rangle G_1(\omega_0, \omega_p)
\]  

(17)

\[
+ \lambda_t(\omega_i) f_t(\omega_i, \omega_0) \langle \omega_t, \omega_0 \rangle G_1(\omega_0, \omega_t),
\]

which is the sum of the cosine-weighted microfacet-BRDFs multiplied by their respective intersection probabilities with the incident direction and their shadowing function in the outgoing direction.

The single-scattering BRDF \(f_1\) is symmetric and energy conserving if \(f_p\) and \(f_t\) fulfill these properties.

#### 5.2 The Multiple-Scattering BRDF

As shown for the Smith model [Heitz et al. 2016], the multiplescattering BRDF \(f_m(\omega, \omega_0)\) of a microsurface can be defined as the expectation of random-walks on the microsurface. Following this idea, we propose a similar random-walk method illustrated in Figure 7 and presented in Algorithm 1.
Random Walk Algorithm. In order to perform the random walk (and compute its result) we need to track the current ray direction \( \omega_r \), its energy throughput \( e \), the total radiance \( L_o \) collected from the direction \( \omega_o \) towards the light source, and the current facet (orientation) \( \omega_m \). The random walk starts either at the facet oriented by \( \omega_o \) by or \( \omega_t \), which are chosen according to the respective probabilities \( \lambda_p(\omega_t) \) and \( \lambda_t(\omega_t) \).

At each intersection, we increase \( L_o \) by the radiance reflected by the current facet \( \omega_m \) weighted by the current ray throughput \( e \) and the masking function \( G_t(\omega_t, \omega_m) \) in direction \( \omega_o \). Then we sample the next direction \( \omega_r \) from the BRDF of the current facet \( f_m \) and multiply the energy throughput by the weight of the sample. The ray leaves the microsurface with a probability given by the masking function \( G_t(\omega_r, \omega_m) \), or it intersects the respective other facet.

\[
\omega_m \leftarrow [U < \lambda_p(\omega_t) \ ? \ \omega_p \ : \ \omega_t] \quad \text{first intersection}
\]

\[
\text{while } \ U < G_t(\omega_r, \omega_m) \text{ do}
\]

\[
L_o \leftarrow L_o + e f_m(-\omega_r, \omega_o) \langle \omega_o, \omega_m \rangle G_t(\omega_o, \omega_m) \quad \text{eval}
\]

\[
(\omega_r', w) \leftarrow \text{sample } f_m(-\omega_r, \omega_o') \langle \omega_o', \omega_m \rangle \quad \text{sample}
\]

\[
\omega_r \leftarrow \omega_r' \quad \text{update direction}
\]

\[
e \leftarrow w e \quad \text{update throughput}
\]

\[
\text{if } U < G_t(\omega_r, \omega_m) \text{ then break}
\]

\[
\text{else}
\]

\[
\omega_m \leftarrow \text{other facet}(\omega_m) \quad \text{interact other facet}
\]

\[
\text{end if}
\]

\[
\text{end while}
\]

Our algorithm is conceptually similar to that of Heitz et al. [2016], but it is simpler for two reasons:

- We use a discrete NDF opposed to continuous NDFs, such as Beckmann or GGX, which need to be importance sampled at each intersection with the microsurface. As our microsurface is made of only two facets, we sample the first facet using the intersection probabilities of Equations 8, and the next intersection can only occur on the other facet.

- In order to simplify the computation, we have chosen a separable masking-shadowing function (Equation 14).

A random walk by explicitly tracing rays (computing geometric ray-facet intersections) and tracking the heights of the intersection points would be consistent with non-separable masking-shadowing functions that model the correlations due to the heights within the microsurface.

Intuitively, assuming separable masking-shadowing means that whenever a ray intersects a facet, its position is uniformly redistributed on the facet surface independently of the height of the intersection point [Heitz 2014]. Due to this assumption we can use the average intersection probability given by the shadowing function \( G_t(\omega_o, \omega_m) \) instead of computing explicit ray-facet intersections.

This assumption yields a simple and mathematically valid algorithm as it does not break the symmetry and the conservation of energy of light transport on the microsurface. Furthermore, in the case where there is either no masking or no shadowing (the first case in Figure 6) then the separable and non-separable masking-shadowing functions are equivalent and our random-walk algorithm is equivalent to an explicit ray tracing simulation with geometric ray-facet intersections.

**Properties of the Algorithm.** In practice, we use Algorithm 1 to either evaluate or sample the cosine-weighted multiple-scattering BRDF \( f_m(\omega_i, \omega_o)(\omega_o, \omega_i) \). Indeed, at the end of the algorithm, the value of \( L_o \) yields an unbiased estimate of the cosine-weighted multiple-scattering BRDF:

\[
f_m(\omega_i, \omega_o)(\omega_o, \omega_i) = E[L_o], \tag{18}
\]

and the direction \( \omega_r \) is a sample of it with weight \( e \).

**Properties of the Multiple-Scattering BRDF.** The BRDF \( f_m \) defined as the expectation of random walks on the microsurface in (Equation 18) is symmetric and energy conserving:

\[
f_m(\omega_i, \omega_o) = f_m(\omega_o, \omega_i), \tag{19}
\]

\[
\int_{\Omega} f_m(\omega_i, \omega_o) \cos \theta_o d\omega_o \leq 1. \tag{20}
\]

Furthermore, it achieves 100% energy conservation:

\[
\int_{\Omega} f_m(\omega_i, \omega_o) \cos \theta_o d\omega_o = 1, \tag{21}
\]

if so do \( f_p \) and \( f_t \).

6 CHOOSING THE BRDF OF THE TANGENT FACET

In Section 5 we derived the single- and multiple-scattering BRDFs of the microsurface which depend on the BRDFs \( f_p \) and \( f_t \) of the facets. Since our goal is to obtain the appearance of a normal map defined by an input perturbed normal \( \omega_p \) and its BRDF \( f_p \), it is meaningful to keep \( f_p \) as provided by the input. However, the choice of the BRDF \( f_t \) of the tangent facet is more arbitrary. Ideally, \( f_t \) would interfere as little as possible with the average material appearance.
Along this idea, we explore two meaningful options for \( f_t \): using either the same BRDF \( f_p \), or using a purely specular material.

### 6.1 Choosing a Tangent Facet of the Same Material

The first option is to use the same BRDF for the tangent facet and the perturbed facet, i.e. \( f_p = f_t \) (Figure 8). The reasoning is that the tangent facet then contributes to the transport similarly as the surrounding perturbed facets would (assuming that material properties vary slowly across the surface).

**Evaluation and Sampling.** In this case, the evaluation and sampling of the multiple-scattering exactly follow Algorithm 1 with \( f_m = f_p \) at each intersection.

**Analytic Diffuse Multiple Scattering.** If the BRDF is diffuse, i.e. \( f_m(\omega_i, \omega_o) \approx \bar{f} \), we can compute the multiple-scattering BRDF analytically using a radiosity method. Recall that classic radiosity [Goral et al. 1984] solves the light transport between diffuse patches with uniform energy distribution which is precisely the assumption of the separable masking-shadowing function (discussed in Section 5.2). By inversion of the resulting \( 2 \times 2 \) transport matrix (for the perturbed and the tangent facet) we obtain the radiosity of each microfacet after an infinite number of reflections as a function of their initial radiosities, i.e. the direct illumination. By this the multiple-scattering BRDF can be evaluated analytically and we use Algorithm 1 only for importance sampling. We provide more details in our supplemental material.

### 6.2 Choosing a Specular Tangent Facet

The second option is to use a purely specular BRDF for the tangent facets with

\[
f_t(\omega_i, \omega_o | \omega_t, \omega_o) = \delta_{\omega_t}(\omega_o),
\]

(22)

where \( \omega_t = \text{reflect}(\omega_i, \omega_o) \), i.e. \( \omega_t \) is the reflection of \( \omega_i \) at the tangent facet with normal \( \omega_t \). Using this option, the tangent facet is a reflector of the perturbed facet and the appearance of the resulting microfacet surface is driven by the BRDF \( f_p \) and the orientation of \( \omega_p \). Note that this choice for \( f_t \) also affects the random walk in Algorithm 1 as there is no need to sample the specular reflection.

**Evaluation and Sampling.** Directly applying Algorithm 1 does not work because of the evaluation of the specular BRDF \( f_t \). However, we can easily modify the algorithm for this case. This is illustrated in Figure 10: we evaluate \( f_p \) with the direction \( \omega_o' \), \( \omega_o \) reflected at the tangent facet \( \omega_t \). We weight this contribution by the intersection probabilities for this path:

- the ray leaving \( \omega_p \) in direction \( \omega_o' \) intersects the tangent facet \( \omega_t \) with a probability given by the complement of the masking function: \( 1 - G_1(\omega_o', \omega_p) \), and
- the ray leaves \( \omega_t \) in direction \( \omega_o \) without further intersection with a probability given by the masking function \( G_1(\omega_o, \omega_t) \).

These modifications of Algorithm 1 are highlighted in blue in Algorithm 2.

**Algorithm 2 Random Walk on the Microsurface with Specular \( \omega_t \)**

\[
\begin{align*}
L_0 &= 0 & \triangleright \text{radiance collected from } \omega_0 \\
\omega_r &= -\omega_i & \triangleright \text{ray direction} \\
e &= 1 & \triangleright \text{ray throughput} \\\no_m &\left[ \mathcal{U} < \lambda_p(\omega_i) \ ? \ \omega_p : \omega_t \right] & \triangleright \text{first intersection} \\
\text{while true do} & \\
\text{if } \omega_m = \omega_p \text{ then} & \\
L_0 &= L_0 + e f_p(-\omega_r, \omega_o) (\omega_o, \omega_p) G_1(\omega_o, \omega_p) & \triangleright \text{eval} \\
L_0 &= L_0 + e f_p(-\omega_r, \omega_o') (\omega_o', \omega_p) (1 - G_1(\omega_o', \omega_p)) G_1(\omega_o, \omega_t) & \triangleright \text{eval} \\
(\omega_o', \omega'_r) &\left( \omega_o', \omega'_r \right) & \triangleright \text{sample} \\
\omega_r &\left( \omega_r \right) & \triangleright \text{update direction} \\
e' &= \omega_r e & \triangleright \text{update throughput} \\
\text{else} & \\
\omega_r &\left( \omega_r \right) & \triangleright \text{reflect} \\
\text{end if} & \\
\text{if } \mathcal{U} < G_1(\omega_r, \omega_m) \text{ then} & \triangleright \text{leave microsurface} \\
\text{break} & \\
\text{else} & \\
\omega_m &\left( \omega_m \right) & \triangleright \text{intersect other facet} \\
\text{end if} & \\
\text{end while} & \\
\end{align*}
\]
Analytic Single and Double Scattering. Thanks to the simplifications due to the specular material \( f_t \) explained above, the single- and double-scattering BRDF has the analytic expression

\[
\begin{align*}
    f_s(\omega_i, \omega_o) (\omega_o, \omega_g) &= \\
    &\lambda p(\omega_i) f_p(\omega_i, \omega_o, \omega_p) (\omega_o, \omega_p) G_1(\omega_o, \omega_p) \\
    &+ \lambda p(\omega_i) f_p(\omega_i, \omega_o, \omega_p) (\omega_o', \omega_p') (1 - G_1(\omega_o', \omega_p')) G_1(\omega_o, \omega_p) \\
    &+ \lambda f(\omega_i) f_p(\omega_i', \omega_o, \omega_p) (\omega_o, \omega_p) G_1(\omega_o, \omega_p),
\end{align*}
\]

(23)

where

- the first line accounts for the single scattering that contains only the path \( i \to p \to o \) because the path \( i \to t \to o \) evaluates to 0 since \( \text{reflect}(\omega_o, \omega_t) \) is in the lower hemisphere,
- the second line accounts for the paths \( i \to p \to t \to o \),
- and the third line accounts for the paths \( i \to t \to p \to o \).

Since the reflected incident direction \( \omega_o' = \text{reflect}(\omega_i, \omega_t) \) is in the lower hemisphere we have an implicit multiplication by \( 1 - G_1(\omega_o', \omega_t) = 1 \).

7 RESULTS

7.1 Validation

We rendered all the results using the Mitsuba renderer [Jakob 2010] with the classic normal mapping-plugin `normalmap` provided in Mitsuba as well as with our microfacet-based normal mapping-plugin `normalmap_microfacet` which is provided in our supplemental material. The supplemental document also contains more results.

**Fig. 11.** Comparison of classic normal mapping and single-bounce path tracing to multiple scattering in our model and path tracing.

Removing Artifacts. In Figure 11, we rendered a detailed surface with a specular material with both height and normal mapping.

Classic normal mapping produces black regions where specular reflections are oriented towards the surface due to the problem illustrated in Figure 1 center and Figure 3 (green regions). These black regions are also present on the height-mapped surface if path tracing is cut after the first bounce and they disappear if all the bounces are computed. Thanks to its multiple-scattering component, microfacet-based normal mapping is able to produce a plausible appearance without black regions similar to the full path-traced simulation on the height-mapped surface. This is impossible to achieve with classic normal mapping.

However, note that the microscopic multiple-scattering component of our microfacet-based normal mapping model does not match the macroscopic multiple scattering that occurs on the displaced surface. These are fundamentally different effects and scales, and produce different results. We believe, however, that a multiple-scattering component is mandatory for normal mapping to achieve a plausible appearance without artifacts. Figure 13, 16 and 17 show that including up to the 2nd-order scattering already removes almost all artifacts.

Energy Conservation. Figure 13 shows a white furnace test in order to verify that our model is energy conserving. We used a specular BRDF that is 100% energy conserving, and lit the surface with a constant white environment. As explained in Section 3.3, classic normal mapping does not preserve energy resulting in visible darkening. Our BRDF loses less energy thanks to its multiple-scattering term, and accounts for 100% of this energy and passes the white furnace test if all the scattering orders are computed.

Symmetry. In practice Veach’s correction factor [Veach 1996] is often not used in rendering engines because it results in unacceptable variance, however, the consequence of this is non-symmetric light transport. It is also disabled in Mitsuba (see footnote of Section 3.1) and the differences between forward (plugin path) and backward path tracing (plugin `prtracer`) easily become apparent (see Figure 14). That is, classic normal mapping does not converge to the same result and thus cannot be used with bidirectional path tracing techniques as the renderer computes a different result whenever the sampling parameters change. In contrast, our microfacet-based normal mapping is symmetric and both techniques converge to the same result. This problem can also be seen in Figure 15 where we compare with advanced light transport algorithms.

Performance. Our microfacet-based normal mapping can be up to 70% more costly than classic normal mapping with our implementation (Figure 16). However, this is the worst case due to the minimal geometry in this scene with only two triangles. The analytic 2nd-order scattering for the specular tangent facet-model, for example, is 35% slower than classic normal mapping in this case. Figure 17 shows a scene with more complex geometry consisting of 20000 triangles. In this case, the overhead of our normal mapping is already largely amortized and the rendering time is about 10% more than with classic normal mapping.

7.2 Choosing a Default Model

Exploring Appearance. In Figure 16 we compare the appearance obtained with our model for different input BRDFs; please also see
We provide it in our supplemental material as the Mitsuba plugin walk. By construction of the random walk, the PDF of the generated (Equation 23) and is thus faster than the random-walk approach and 7.3 Implementation in a Path Tracer property of random walk sampling algorithms. have to evaluate a PDF in practice to compute these weights. The out in the expression of the importance sampling weight, we do not each bounce on the microsurface. Because the BRDF and PDF cancel is the energy throughput $e$ of the non-absorptive BRDF. The stochastic evaluation does not have a closed-form PDF for the di-
al mapping (Algorithm 1, plugin normalmap_microfacet_default) is shown in Figure 15. normalmap_microfacet_default (Section 6.2) with 2nd-order scattering. Note that this model can be evaluated analytically (Equation 23) and is thus faster than the random-walk approach and comes without additional variance. We believe that these properties qualify it as the default substitution for classic normal mapping. We provide it in our supplemental material as the Mitsuba plugin normalmap_microfacet_default.

Default Model. Our analytic default model (Section 7.2, Equation 23, plugin normalmap_microfacet_default) works straightforwardly and consistently with any light transport algorithms. This is shown in Figure 15.

General Random-Walk Model. Our general microfacet-based normal mapping (Algorithm 1, plugin normalmap_microfacet) with stochastic evaluation does not have a closed-form PDF for the directions sampled by the random walks. However, not knowing the PDF is not a problem for common light transport algorithms such as forward path tracing or bidirectional path tracing.

For computing the importance sampling weights we do not need to evaluate a PDF. Instead, we use a random walk for importance sampling and the BRDF is precisely the expectation of this random walk. By construction of the random walk, the PDF of the generated samples is the cosine-weighted BRDF of a non-absorptive material

\[
PDF = \text{BRDF} \cdot \cos
\]

and what remains in the sample weight

\[
\text{weight} = \frac{e \cdot \text{BRDF} \cdot \cos}{PDF} = \frac{e \cdot \text{BRDF} \cdot \cos}{\text{BRDF} \cdot \cos} = e
\]

is the energy throughput $e$ that is multiplied by the reflectance after each bounce on the microsurface. Because the BRDF and PDF cancel out in the expression of the importance sampling weight, we do not have to evaluate a PDF in practice to compute these weights. The weight being given by the remaining energy throughput is a classic property of random walk sampling algorithms.

Since our evaluation yields an unbiased stochastic estimate and our importance sampling technique is also unbiased as explained above, we can use Multiple Importance Sampling (MIS) [Veach and Guibas 1995] to combine both techniques in a way that remains unbiased. Note that the PDF is usually used for computing the MIS weights of different sampling strategies, but using an approximate PDF does not introduce bias. Indeed, the necessary and sufficient conditions for MIS to be unbiased are:

- the weights of the different techniques should be deterministic,
- the weights should be non-negative, and
- the weights should add up to exactly 1.

Hence, there is no need for an MIS weight to use the exact PDF. For example, common implementations of MIS actually use PDF$^2$ instead of the PDF itself, which is known as the power heuristic.

The choice of weights impacts only the variance reduction, but does not introduce bias as long as the three aforementioned conditions are satisfied. Hence, we proceed similarly to Heitz et al. [2016] who approximate the PDF of their multiple-scattering BRDF by the closed-form single-scattering BRDF with a small diffuse contribution to account for multiple scattering. We found this to work well in terms of variance reduction and it makes the model work consistently with forward and bidirectional path tracing, as shown in Figure 15.

It is also possible to make such random-walk BRDFs work with Metropolis Light Transport by incorporating the space of the random walks in the path space explored by the algorithms. The contribution and PDF of the paths (random walks) inside the microsurface can be evaluated and mutated by changing the random numbers used to generate them. We explain this in more detail in a supplemental document and leave a further exploration thereof for future work.

Note that our stochastic random-walk evaluation produces results with higher variance than classic normal mapping. Figure 12 shows comparisons of the noise with the same number of samples per pixels. However, note that a non-negligible portion of the additional variance is due to the increased reflectance of the surface and not solely a consequence of the stochastic evaluation.

![Fig. 12. Comparison of the variance due to the stochastic evaluation with our general random-walk model.](image)
Fig. 13. Validation of the energy conservation of light transport with a white furnace test. The material is specular and lit by a constant white environment.

Fig. 14. Validation of the symmetry of light transport. We compare forward and backward path tracing; the bottom row shows 4× difference images. The material is a copper conductor with GGX roughness $\alpha = 0.3$. Unlike classic normal mapping, microfacet-based normal mapping produces symmetric BRDFs and we obtain black difference images (the small amount of noise is due to the remaining variance of the renders).

Fig. 15. Comparison of normal-mapping consistency with different light transport algorithms. Classic normal mapping does not produce the same image with forward path tracing, bidirectional path tracing, and Metropolis Light Transport. The difference increases considerably when standard practices for fixing artifacts are used. Our default microfacet-based normal mapping, which is an analytic model, works consistently with any light transport algorithms. Our general microfacet-based normal mapping with stochastic evaluation works straightforwardly with forward and bidirectional path tracing. Making it work with MLT is possible in theory, but requires modifications to the integrator. Note that the small amount of noise is due to the remaining variance of the renders.
Fig. 16. Appearances and performance of our model. We compare the appearances achieved by our model with a tangent facet made of the same material (4th and 5th column, Section 6.1) or specular material (6th and 7th column, Section 6.2). The timings are measured for the gold material shown in the first row, but they are roughly equal across materials shown. We measured the 2nd-order scattering of the specular facet-model using both the random-walk evaluation and its analytic expression.

Fig. 17. Our model applied on a more complex geometric model (20000 triangles) with textured GGX roughness. Classic normal mapping (left) shows distracting black fringing. Note the little overhead of our microfacet-based normal mapping in geometrically more complex scenes.
We thank Jonas Wagner (https://29a.ch) for providing the normal map of the coin.

ACKNOWLEDGEMENTS
We thank Jonas Wagner (https://29a.ch) for providing the normal map of the coin.

REFERENCES