

Manifold Next Event Estimation

Johannes Hanika, Marc Droske and Luca Fascione

Weta Digital, New Zealand



Figure 1: A Stanford Buddha in a glass sphere showcasing the robustness of our method in the presence of refractive caustics. Our technique can efficiently sample this transport: the two images are an equal time comparison of our method to bidirectional path tracing. While the raw speed impact is about $3\times$, an equal quality comparison would render our method many thousand times more efficient here.

Abstract

We present manifold next event estimation (MNEE), a specialised technique for Monte Carlo light transport simulation to render refractive caustics by connecting surfaces to light sources (next event estimation) across transmissive interfaces. We employ correlated sampling by means of a perturbation strategy to explore all half vectors in the case of rough transmission while remaining outside of the context of Markov chain Monte Carlo, improving temporal stability. MNEE builds on differential geometry and manifold walks. It is very lightweight in its memory requirements, as it does not use light caching methods such as photon maps or importance sampling records. The method integrates seamlessly with existing Monte Carlo estimators via multiple importance sampling.

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Raytracing

1. Introduction

Simulating light transport has many applications across different fields of science and the computer graphics industry. In recent years the movie industry has shown a growing interest in Monte Carlo path tracing, which lead to its deployment in a number of production renderers both commercially available (such as Solid Angle's *Arnold*, and Pixar's *RenderMan 19*) and internally developed (such as Disney's *Hyperion* or Weta Digital's *Manuka*).

In contrast to rendering for the architectural or manufacturing industries, where the dominant use case is still images, production renderers in the film industry are mainly

used to produce animations. This imposes stringent constraints on the amount of time that can be spent rendering each frame [Ste05]. For most productions this means renderers generally have to complete overnight and deliver sharp, low-noise images, with temporal stability being of primary importance [EW11]. The richness of the visual language commonly employed in movies (especially in VFX work) quickly translates into large numbers of highly complex assets, which often make ray tracing very expensive. For this reason, careful sample placement can yield large benefits.

One path configuration that is particularly hard to sample in the context of unbiased path tracing are so-called

specular-diffuse-specular (SDS) paths. These are light transport paths containing a diffuse or broadly glossy vertex between two specular (or tightly glossy) scattering events (see Fig. 2). This configuration is problematic because the specular events impose constraints on the half vectors [KHD14a] which are hard to satisfy. When sampling, this means simulating the diffuse scattering at the middle vertex (\mathbf{x}_b in Fig. 2) is unlikely to produce a connection to a light source subtending a small solid angle (as seen from \mathbf{x}_b). Indeed, in this case the common approach would be to use a different sampling technique to deterministically connect to a position on the light source (which is called *next event estimation*, NEE). It is easily seen that this cannot be employed for the configuration of Fig. 2 as all paths from \mathbf{x}_b to the light need to pass through a refractive interface.

While there are methods in literature which deal with SDS paths, such as Metropolis light transport (MLT) or photon mapping, these have drawbacks in terms of either temporal stability or large memory footprints and as such tend to be of limited practical use to render animations in a production context (see Sec. 2 for further details). This is especially frustrating for seemingly simple cases, such as water or sweat droplets on skin, or objects embedded in refractive media such as glass or water (see Fig. 1). Often admissible light paths are not deviating by much from the non-refracted connection which just ignores the interface (dashed line in Fig. 2). This is the path proposed by next event estimation and it fails because it violates Fermat’s law.

Our insight is to try and make NEE more “stubborn”, meaning that it will not give up a connection proposal (a *seed path*) quite as quickly, but try hard to get a contribution out of it transforming the seed path into an *admissible path*, which satisfies the half vector constraints. We achieve this by employing correlated sampling, i.e. randomly perturbing the seed path while remaining outside the context of Markov chain Monte Carlo. While on the one side this simplifies certain aspects of implementation and integration into existing systems, it also improves noise characteristics in between frames. The method works for purely specular materials as well as for rough scattering and through multiple layers of transmissive interfaces. It extends the next event estimation technique and integrates well into Monte Carlo path tracers with multiple importance sampling [VG95] without introducing bias. The technique is most effective in a range of simple cases which are relevant and very common in practice, and successfully combines low memory requirements and scalability, since its independence from caches lifts any potential thread synchronisation needs.

2. Background and Previous Work

The purpose of light transport simulations is to solve instances of the problem of radiative transfer [Cha60]. In the context of computer graphics, this was formulated as the rendering equation [Kaj86] and is often solved using path

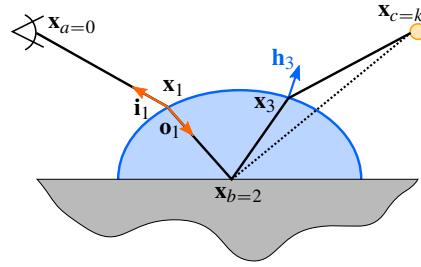


Figure 2: Notation used in this paper for vertices \mathbf{x} , incoming and outgoing directions \mathbf{i}, \mathbf{o} and half vectors \mathbf{h} . Because our new technique is most useful in the context of path tracing from the eye, we start numbering the vertices from there.

tracing, which is a Monte Carlo integration technique based on random walks [Erm75, Sob94]. The fundamental building block of this algorithm consists in constructing paths of light in a scene that connect light sources and the eye.

Path space The space of all the light transport paths is called *path space* and here denoted Ω [Vea98]. We express a path in Ω as a list of vertices \mathbf{x}_i and use the shorthand $\mathbf{X}_{a,b} = (\mathbf{x}_a, \mathbf{x}_{a+1}, \dots, \mathbf{x}_b)$ to denote *sub-paths* in it. Likewise, we denote half vectors [KHD14a] at vertex i as \mathbf{h}_i and use the shorthand $\mathbf{H}_{a,b} = (\mathbf{h}_a, \mathbf{h}_{a+1}, \dots, \mathbf{h}_b)$. We use \mathbf{i}_i and \mathbf{o}_i for the incoming (from the eye) and outgoing (to the light) directions at vertex \mathbf{x}_i , respectively (see Fig. 2). Measures are denoted $d\mathbf{x}$ for the vertex area measure, $d\mathbf{o}$ for solid angle and $d\mathbf{o}^\perp = \langle \mathbf{o}, \mathbf{n} \rangle d\mathbf{o}$ for projected solid angle. We will write $d\mathbf{h}$ for the domain of half vectors, which is usually projected solid angle [Jak13, KHD14a].

A path has $k + 1$ vertices, where $\mathbf{x}_{a=0}$ is on the eye, \mathbf{x}_b is where the next event estimation sampler is invoked (i.e. the last vertex created by extending the path via outgoing direction sampling) and $\mathbf{x}_{c=k}$ is the sampled point on the light source.

Intuitively, the purpose of light transport simulations is to detect the flux of photons arriving at a pixel j :

$$I_j = \int_{\Omega} f(\mathbf{X}) d\mathbf{X}, \quad (1)$$

where $d\mathbf{X} = \prod_{i=0}^k d\mathbf{x}_i$ is the product vertex area measure over all vertices \mathbf{x}_i in the path of length k , and Ω is the path space containing all paths of lengths $k \in [1, \infty)$.

The *measurement contribution function* $f(\mathbf{X})$ (used in product vertex area measure here) intuitively measures how many photons per second are transported via the differential areas of the vertices the path traverses. It takes its simplest form in product projected solid angle measure $d\mathbf{O}^\perp = d\mathbf{x}_0 \prod_{i=0}^{k-1} d\mathbf{o}_i^\perp$:

$$f_{d\mathbf{O}^\perp}(\mathbf{X}) = W(\mathbf{x}_0) L_e(\mathbf{x}_k) \prod_{i=1}^{k-1} f_r(\mathbf{i}_i, \mathbf{x}_i, \mathbf{o}_i). \quad (2)$$

The bidirectional scattering distribution function (BSDF) [NRH*77] is denoted f_r , the eye responsivity is $W(\mathbf{x}_0)$ and the emitted radiance at the light sources is $L_e(\mathbf{x}_k)$. *Geometry terms* are used as Jacobian determinants to convert from $d\mathbf{O}^\perp$ to $d\mathbf{X}$. We do not model volume scattering as we will not be explicitly dealing with it in this paper.

Deterministic connection Next event estimation (NEE) or computing direct lighting is the simplest and most important way of performing deterministic connections to construct a transport path, and much thought has been spent optimising it [SW91, UFK13]. Bidirectional path tracing [LW93, VG94] connects not only to the last point of the path on the light source, but any two vertices on two independent random walks, one started at the camera and one started at the lights. This results in many estimators, making it important to weight the techniques optimally using multiple importance sampling (MIS) [VG95] to avoid excessive noise. However, even bidirectional methods fail to connect through refractive interfaces in the SDS scenario.

Photon mapping One possible way to address this problem is to employ regularisation: forcefully smoothing the integrand (introducing bias). Regularisation can be applied in the angular domain by softening the BSDF [KD13b] or by using kernel estimation to search through cached photons [Jen96, HJ09, KD13a, HPJ12, GKDS12]. Algorithms employing photon mapping must shoot and store large numbers of photons, which can be prohibitively costly when rendering complex, slow to ray trace assets, as realistic counts range in the tens to hundreds of millions in order to attain the necessary image quality. Further, the efficiency of these techniques can be very low in practical scenarios, as typically only caustics need this kind of treatment, and these often cover only small (although very important) regions of the final image. A great many photons must be traced in order to place a small portion of them near the caustics, resulting in a large proportion of wasted computation. Some have proposed the use of virtual devices such *portals* or *photon attractors*: these can indeed help in simple configurations, but tend to have trouble in situations involving occlusion or indirect sources of illumination, greatly limiting the practicality of the technique.

Metropolis sampling Adaptive sampling in path space has most successfully been done in the context of *Metropolis light transport* (MLT) [VG97, KSKAC02]. The idea is to run a Markov chain of paths \mathbf{X}^i where the stationary distribution will lead to a *probability density function* (PDF) $p(\mathbf{X})$ following a chosen target, most often the measurement contribution normalised by image brightness: $p(\mathbf{X}) = f(\mathbf{X})/b$.

This is implemented by choosing a tentative proposal path \mathbf{X}^i using perturbations of the current path \mathbf{X}^{i-1} . Next, the state of the Markov chain \mathbf{X}^{i+1} is randomly set to either \mathbf{X}^i or \mathbf{X}^i ,

depending on the acceptance probability [MRR*53, Has70]

$$a = \min \left\{ 1, \frac{f(\mathbf{X}^i)/T(\mathbf{X}^i \rightarrow \mathbf{X}^i)}{f(\mathbf{X}^{i-1})/T(\mathbf{X}^{i-1} \rightarrow \mathbf{X}^i)} \right\}. \quad (3)$$

This procedure elegantly explores nearby paths in interesting regions of path space requiring only the measurement contribution and the pairwise transition probabilities $T(\cdot \rightarrow \cdot)$. However, the characteristic low-frequency noise of MLT is very objectionable in the context of animations, especially when only low sample counts are practical. This is partly because the mixing rates of the Markov chain are limited by the bidirectional mutation [VG97], which typically has acceptance rates below 1%.

Energy redistribution path tracing (ERPT) [CTE05] addresses this issue, but the parameters for the algorithm require extensive tuning and render times are directly proportional to the product of started chains multiplied by the desired mutation steps. Another avenue for improvement is to combine photon mapping and Markov chain techniques [HJ11], to try to deposit more photons where they will be most useful. While the exploration of SDS paths with ERPT or the rendering of a caustic using photons can work very well, both techniques depend on an existing seed path: ERPT in the context of a Markov chain (the current sample) and photon mapping by means of a cache that has to be filled by shooting many light paths. In other words, both methods fundamentally rely on an event of low probability (finding a path belonging to a small caustic). We directly sample the connections of all initiated random walks through transmissive interfaces towards the light sources without relying on rare events to encounter interesting configurations.

Differential geometry Single scattering through refractive interfaces [WZHB09] searches for the contribution of direct light through a single layer of a smooth dielectric interface by constructing a search tree over the geometry and using a pruning algorithm based on a half vector formulation. Such a search requires a dedicated acceleration structure for these refractive interfaces, is comparatively slow and is limited to a single interface. Our technique on the other hand works through multiple layers of potentially rough interfaces.

Manifold exploration [Jak13] and half vector space light transport [KHD14a] enable efficient exploration of indirect caustics in the context of path space MLT. This is achieved by re-parametrising a path in a small environment around a base path by its start- and endpoint and half vectors for all inner vertices. To mutate a path according to a perturbed set of half vectors, Newton's method is used and the required constraint derivatives at every vertex \mathbf{x}_i are found analytically by means of differential geometry [dC76]. More precisely, for every vertex \mathbf{x}_i it is computed how the half vector \mathbf{h}_i changes with respect to the adjacent vertices:

$$A_i = d\mathbf{h}_i/d\mathbf{x}_{i-1} \quad B_i = d\mathbf{h}_i/d\mathbf{x}_i \quad C_i = d\mathbf{h}_i/d\mathbf{x}_{i+1} \quad (4)$$

resulting in a block-tridiagonal matrix for the whole

path [Jak13], which can be inverted efficiently [Sal06]. We build on the same underlying theoretical foundation and develop a correlated Monte Carlo NEE technique which does not depend on Markov chains and, while somewhat less generally applicable, leads to more uniform convergence in cases of high visual importance.

Gradient Domain Light Transport The mechanism of manifold exploration has been applied to gradient domain light transport [LKL*13, KMA*15]. This method computes offset paths from seed paths via shift mappings to compute image space gradients. These gradients are then used, together with the noisy image, to reconstruct a smoother image via a screened Poisson solver. While the underlying theory is similar to our approach, in their work they depend on a valid seed path which is usually harder to find than ours.

Correlated sampling There are studies in *perturbation theory* and *correlated sampling* in the field of neutron transport [Mat63, Rie84, DGN85], but we note that in that context the focus is on perturbing the input parameters (scattering cross sections) and observing the difference in sensor response while leaving the transport path constant. In some sense, the meaning assigned by the neutron transport community to these phrases is dual to ours, as we perturb the transport path and the correlation is due to similarity between the base and perturbed paths.

3. Manifold Next Event Estimation

Our new technique creates sub-paths through transmissive interfaces. Conceptually, we sample an admissible path \mathbf{X} by perturbing a given seed path \mathbf{Y} , and do so outside the context of Markov chains. As we are in a standard Monte Carlo context, the probability density of \mathbf{X} becomes

$$p(\mathbf{X}) = \int_{\mathbf{Y} \in \Omega} p(\mathbf{X}|\mathbf{Y}) \cdot p(\mathbf{Y}) d\mathbf{Y}, \quad (5)$$

which requires costly integration to obtain the marginalised distribution $p(\mathbf{X})$. MLT elegantly avoids computing this integral by ensuring the equilibrium distribution is correct. One way to achieve this is to require the acceptance probability to only be dependent on the current and the tentative sample (the *detailed balance* condition).

Computing the Marginalised Distribution In the Monte Carlo context, we achieve our goal by introducing a few constraints. First, we only do a single-step perturbation, where the seed path \mathbf{Y} is always deterministically defined by the current shading point \mathbf{x}_b and the point on the light source \mathbf{x}_c . As the seed path is solely used to initiate the construction of the admissible transport path, there doesn't need to be any energy balance. Further, we assume that given a set of half vectors $H_{b+1,c-1}$ there exists a bijection between the seed path \mathbf{Y} and the admissible path \mathbf{X} . This means the integrand in Eq. 5 is non-zero only for a single path

\mathbf{Y} and $p(\mathbf{X}|\mathbf{Y}) = \delta(\mathbf{Y})$. This renders the evaluation of the marginalised distribution trivial.

Naturally, to implement our technique in an unbiased path space renderer, we need to provide routines for sampling and PDF evaluation to integrate into the MIS framework. The rest of this section is a high level overview of sampling and PDF computation. Details specific to our case are expanded in the following sections.

Sampling Direct Illumination Sampling a path suffix through transmissive interfaces is summarised in Alg. 1: We use standard techniques [SW91] to sample from \mathbf{x}_b a position on the light source, potentially doing normal culling and including distance to the shading point. Next, we construct the seed path building a straight path from \mathbf{x}_b towards \mathbf{x}_c , and collecting all transmissive events along this line (potentially on rough transmissive surfaces). Then we sample half vectors for all rough vertices $b+1, \dots, c-1$, if any. In the end, a Newton solver with the constraint derivative matrix (also called *manifold walk* [Jak13]) is used to find the admissible path matching such half vectors (Sec. 3.4). The contribution of the admissible path is then evaluated using the optimised measurement contribution in half vector space [KHD14a].

Algorithm 1 Next event estimation sampling.

Require: path prefix $\mathbf{X}_{a,b}$ **return** full path $\mathbf{X}_{a,c}$
 $\mathbf{x}_c, p_{d\mathbf{x}}(\mathbf{x}_c) \leftarrow$ sample point on light source
 $\mathbf{Y}_{b+1,c-1} \leftarrow$ vertices on straight line $(\mathbf{x}_b, \mathbf{x}_c)$ // Sec. 3.1
 $\mathbf{H}_{b+1,c-1}, p_{d\mathbf{H}}(\mathbf{H}) \leftarrow$ sample half vectors // Sec. 3.2
 $\mathbf{X}_{b+1,c-1} \leftarrow$ h_to_positions($\mathbf{H}_{b+1,c-1}, \mathbf{Y}$) // Sec. 3.4
 // compute the measurement (Sec. 3.3)
return $f_r(\mathbf{x}_b) \cdot f_{d\mathbf{H}}(\mathbf{X}_{b,c}) \cdot L_e(\mathbf{x}_c) / (p_{d\mathbf{H}}(\mathbf{H}) \cdot p_{d\mathbf{x}}(\mathbf{x}_c))$

Probability Density The PDF evaluation routine receives an admissible path \mathbf{X} and needs to find the seed path \mathbf{Y} . This proceeds as follows: first, the seed path is constructed connecting \mathbf{x}_b to \mathbf{x}_c . Then the computation proceeds as in the case for sampling to ensure the Newton solver will actually converge back to \mathbf{X} (see Alg. 2). We do not converge the walk to machine precision but to some small finite epsilon for performance reasons. Thus, the PDF of the half vectors $p_{d\mathbf{H}}(\mathbf{H})$ should be evaluated on the half vectors of the input path, not on those the manifold walk converged to, this ensures MIS will use consistent weights.

3.1. Creating the Seed Path

Creating the sub-path $(\mathbf{x}_{b+1}, \dots, \mathbf{x}_c)$ is straightforward. The vertex \mathbf{x}_c is sampled in the same way as computing direct lighting. Among other things, this ensures that MNEE degenerates to regular NEE if there are no refractive vertices that would block visibility. There are two important details illustrated in Fig. 3: First, if direct lighting is using any culling with respect to the surface normal at \mathbf{x}_b , this needs

Algorithm 2 PDF evaluation.

Require: path suffix $\mathbf{X}_{b,c}$ **return** PDF $p_{d\mathbf{x}}(\mathbf{X}_{b,c})$
 $p_{d\mathbf{x}}(\mathbf{x}_c) \leftarrow$ PDF of point \mathbf{x}_c on light source
 $\mathbf{Y}_{b+1,c-1} \leftarrow$ vertices on straight line $(\mathbf{x}_b, \mathbf{x}_c)$ // Sec. 3.1
 $p_{d\mathbf{H}}(\mathbf{H}) \leftarrow$ PDF of half vectors on input path
 $\mathbf{X}'_{b+1,c-1} \leftarrow$ h_to_positions($\mathbf{H}_{b+1,c-1}, \mathbf{Y}$) // Sec. 3.4

if $|\mathbf{X}_{b+1,c-1} - \mathbf{X}'_{b+1,c-1}| > \varepsilon$ **then**
 return 0 // fail: found a different path
end if

$p_{d\mathbf{X}}(\mathbf{X}_{b+1,c-1}) \leftarrow p_{d\mathbf{H}}(\mathbf{H}) \cdot \left| \frac{d(\mathbf{h}_{b+1} \cdot \mathbf{h}_{c-1})}{d(\mathbf{x}_{b+1} \cdot \mathbf{x}_{c-1})} \right|$ // Sec. 3.3
return $p_{d\mathbf{X}}(\mathbf{X}_{b+1,c-1}) \cdot p_{d\mathbf{x}}(\mathbf{x}_c)$

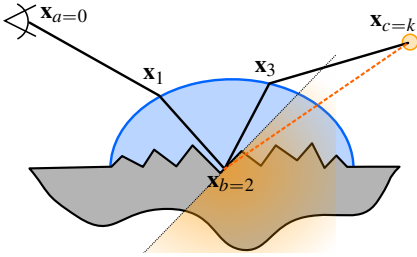


Figure 3: Handling occlusion: normal culling used with MNEE to sample \mathbf{x}_c needs special care not to cull light sources which are only visible for the admissible path (culling region for \mathbf{x}_b indicated in orange). Also, the seed path (dashed in orange) may be occluded, while the admissible path is not. For this reason, self-occlusion is checked only after determining the admissible positions of $\mathbf{X}_{b+1,c-1}$.

an additional tolerance margin to handle the configuration in this figure. Second, the seed path might be blocked by a heavily displaced surface. We detect such cases and ignore self-occlusion until the solver has converged and the admissible positions of $\mathbf{X}_{b+1,c-1}$ are known. General occluders such as additional objects under the water droplet in our example are not a common use case for us, but ignoring them has a big impact on performance as it causes the Newton solver to run on paths that will in many cases still be occluded after convergence, wasting computation.

3.2. Sampling Half Vectors

Sampling is performed in the half vector domain, i.e. we know $p_{d\mathbf{H}}(\mathbf{H})$. In the MLT context, the path space is explored well given an approximate step size proposal, e.g. chosen by assuming the half vector distribution of a surface point is Gaussian [KHD14a]. In contrast to that, in a pure Monte Carlo context it is imperative to sample the half vector distribution of the underlying BSDF very carefully. A good option is to use the sampling function of the BSDF, assuming that it draws from a global distribution and not from the visible normals [HD14]. This is because after the walk has

converged, the incoming direction will be different from the one the half vector was sampled with, which will typically change the set of visible normals.

3.3. Computing the Measurement and PDF

To evaluate the pixel contribution of a sampled path we need to evaluate the value of the Monte Carlo estimator, which is the ratio of the measurement contribution to the PDF (both in product vertex area measure) $\hat{I} = f(\mathbf{X})/p_{d\mathbf{X}}(\mathbf{X})$. To improve numerical stability in the evaluation of the measurement, we compute $\hat{I} = f_{d\mathbf{H}}(\mathbf{X})/p_{d\mathbf{H}}(\mathbf{H})$ in half vector space [KHD14a, Eq. (4) and (5)] as this formulation contains no geometry terms for the sub-path $\mathbf{X}_{b+1,c}$:

$$f_{d\mathbf{H}}(\mathbf{X}_{b,c}) = \left| \frac{d\mathbf{o}_b}{d\mathbf{x}_c} \right| \prod_{i=b+1}^{c-1} f_r(\mathbf{x}_i) \left| \frac{d\mathbf{o}_i}{d\mathbf{h}_i} \right| \left| \frac{\langle \mathbf{o}_i, \mathbf{n}_i \rangle}{\langle \mathbf{h}_i, \mathbf{n}_i \rangle} \right|, \quad (6)$$

where $|d\mathbf{o}_i/d\mathbf{h}_i|$ is well known from microfacet theory [WMLT07], i.e. $4|\langle \mathbf{o}_i, \mathbf{h}_i \rangle|$ for reflections. Computing $|d\mathbf{o}_b/d\mathbf{x}_c| = G(\mathbf{x}_b, \mathbf{x}_{b+1})|T_{b+1}|$ via the determinant of the transfer matrix T_{b+1} pushes the accuracy limits of single-precision floating point arithmetic: double precision or a careful eye to implementation and compiler optimization parameters are recommended here.

Computing the PDF for MIS requires us to convert the half vector PDF to product vertex area measure

$$p_{d\mathbf{X}}(\mathbf{X}) = p_{d\mathbf{H}}(\mathbf{H}) \left| \frac{d\mathbf{H}}{d\mathbf{X}} \right|. \quad (7)$$

The routine `solve_matrix_h_to_x`, as outlined in [KHD14b, Fig. 2], inverts the block-tridiagonal constraint derivative matrix with the blocks $\mathbf{A}, \mathbf{B}, \mathbf{C}$ for every vertex. We use the same LU decomposition, and the full Jacobian determinant $|d\mathbf{H}/d\mathbf{X}|$ can be computed very cheaply inside this routine. More precisely, we evaluate $|d\mathbf{H}/d\mathbf{X}|$ during the first LU decomposition step as a product of determinants [Mol07]:

$$\left| \frac{d(\mathbf{h}_{b+1} \cdot \mathbf{h}_{c-1})}{d(\mathbf{x}_{b+1} \cdot \mathbf{x}_{c-1})} \right| = \prod_{i=b}^{c-1} |\Lambda_i|, \quad (8)$$

where the difference to the original code listing [KHD14b] is that our index i runs over the sub-path $i \in [b, c-1]$ instead of the whole path $[0, k-1]$. In order to reduce numerical issues, it is worth computing this product in double precision.

Specular Case In the specular case, we sample the half vectors with Dirac deltas $p_{d\mathbf{h}}(\mathbf{h}_i) = \delta_{d\mathbf{h}}(\mathbf{h}_i)$. When evaluating the pixel contribution \hat{I} , these deltas cancel with the measurement contribution. In our implementation, the BSDF evaluation returns the BSDF in delta-half vector space (instead of the regular space of outgoing directions $d\mathbf{o}$). That is, a specular dielectric will return κ (the Fresnel term R for reflection). This avoids computing the per-vertex Jacobian in

the half vector space measurement Eq. (6) since

$$f_{r,\text{dh}} = f_r(\mathbf{x}) \left| \frac{d\mathbf{o}}{d\mathbf{h}} \right| |\langle \mathbf{o}, \mathbf{n} \rangle| = \kappa \cdot \delta_{\text{dh}}(h). \quad (9)$$

Now that the delta function lives in half vector domain (instead of the commonly used domain for BSDF, \mathbf{do}), it cancels with the PDF and never needs to be evaluated. For MIS this means that we do not compute the PDF in vertex area domain, but lack a factor of $\delta_{\text{dh}}(\mathbf{h}) \left| \frac{d\mathbf{h}}{d\mathbf{o}^\perp} \right|$ for every specular vertex.

Layered BSDFs with Mixed Specular and Glossy Lobes

This change of measure when moving from glossy to specular poses a problem for BSDFs which contain both scattering modes at the same surface point (such as car paint with a clear coat layer) since we would be comparing different measure spaces in MIS.

To analyse this, we compare the values of a mollified specular PDF [KD13b] to a regular glossy PDF and note how the mollified PDF will outweigh any other values as the mollification parameter tends to zero. This means that in the limit the MIS weights will be exclusively determined by the specular PDF, and the glossy contributions can be ignored. On the other hand, the glossy lobe will only interfere with the specular one on a measure zero set and thus the specular lobe can be ignored in the reverse case.

We extend path space for such cases, so that the class of scatter mode (glossy or specular) at every vertex is treated as an additional dimension. That is, we randomly pre-determine whether to use the specular or glossy components, and ignore all BSDF layers of the other class when doing MNEE.

3.4. Convergence of the Predictor/Corrector

The predictor/corrector method employed to find an admissible path \mathbf{X} from the seed path \mathbf{Y} and a set of perturbed half vectors is very similar to the one employed in manifold exploration [Jak13] and half vector space light transport [KHD14a]. First the half vector distance $\Delta\mathbf{H}^i$ to the target half vectors is transformed to a position offset $\Delta\mathbf{X}$ using the first order approximation provided by the constraint derivate matrix (using the routine `solve_matrix_h_to_x`). The projection $\mathbf{X}^{i+1} \leftarrow P(\mathbf{X}^i + \Delta\mathbf{X})$ used to move the displaced vertex $\mathbf{X}^i + \Delta\mathbf{X}$ back to the surface could be performed by ray tracing. Our experiments have revealed that using a closest point search instead (i.e. finding the closest point on the surface of the same shape which the seed path vertex originally belonged to) makes the projection robust, as it avoids the problem of missing geometry when intersection testing a ray and a surface which is nearly parallel to it. Also, it can be faster than casting a ray for complex geometry since the closest point search can be limited to a small radius $|\Delta\mathbf{x}_i|$ and implemented as a traversal of the same bounding volume hierarchy used for ray tracing.

iterations:	5	10	20	35	50
time (A)	352	410	527	646	680
time (B)	352	407	501	583	637
avg it. (A)	3.55	4.21	5.41	5.87	5.94
avg it. (B)	3.55	4.39	6.51	8.45	9.50
success (A)	33%	39%	45%	46%	46%
success (B)	33%	41%	52%	58%	59%

Table 1: Performance numbers for Fig. 4, depending on the maximum allowed number of iterations. We show time in seconds, average iterations until successful convergence and success rate of the predictor/corrector. All tests are run both without (A) and with (B) allowing the error to increase.

Another difference to the use case of manifold walks in MLT is the starting configuration of the Newton solver: our seed path \mathbf{Y} is in general not admissible and potentially far away from the admissible solution \mathbf{X} whereas in MLT all accepted iterations start and end on admissible paths. Fig. 4 and Tab. 1 show how this affects the number of required iterations. In areas where \mathbf{Y} and \mathbf{X} are reasonably close together, we need very few iterations, typically 2–4. In the dimmer part of the caustic the seed path is a poor initial guess for the manifold walk and the solver needs to jump over local minima to reach the required half vectors. This means the seed path is in a different subspace, and the constraint derivative matrix at the seed \mathbf{Y} is meaningless for a local environment around the final path \mathbf{X} . We can still find \mathbf{X} in such cases by *successive over-relaxation*, i.e. allowing the step size to grow above one and allowing the error of \mathbf{X}^{i+1} to increase temporarily instead of reverting to \mathbf{X}^i . However, this can result in a very large number of iterations (500 or more for successful convergence), making some individual samples very expensive. Our experiments indicate that a maximum iteration count of 15 works well in practice: depending on the setup, going from 15 to 50 iterations adds less than 10% successful convergence, and from 50 to 500 less than 1%.

3.5. Outlier Removal

If noise free images are more important than unbiased results, it is possible to detect cases where MNEE should have been more efficient than the other techniques (i.e. paths with a transmissive suffix to the light source), but failed (because the walk did not converge or in case of ambiguous paths, cf. Fig. 7). In such cases, we can discard the contributions of all other techniques, too, since we know we will not be able to sample these efficiently. This is similar to the use of path space regular expressions to discard transmitted caustics from all techniques but MNEE. We implement this in the MIS weight calculation: a sample from the regular path tracer computes the probability $p(\mathbf{X})$ of MNEE (see Alg. 2). If this evaluates to zero even though the class of the path has a transmissive suffix, the MIS weight of the current technique is set to zero, too, introducing bias.

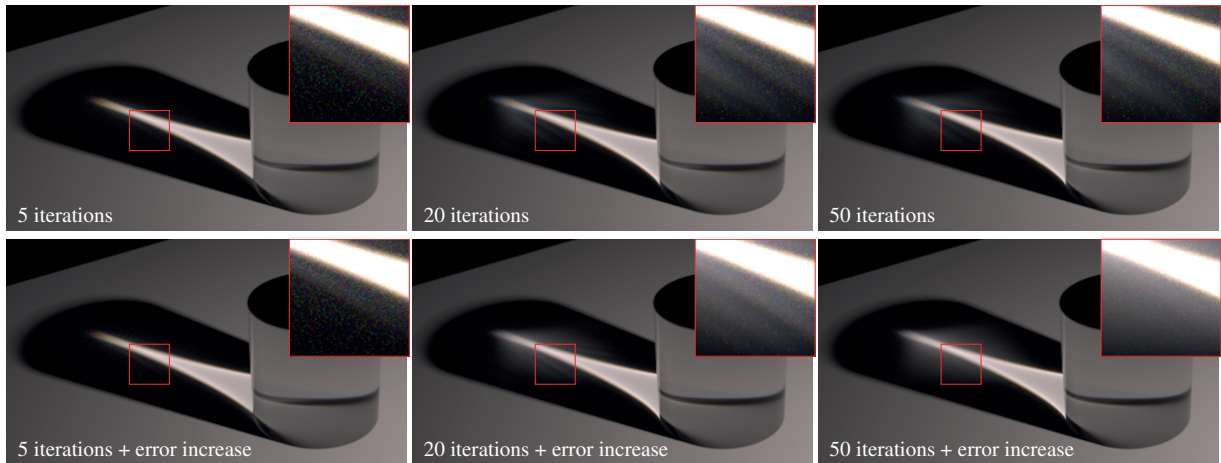


Figure 4: A challenging case for the predictor/corrector method: a caustic where the initial seed path is too far off and the iteration fails to jump over the bright caustic to find the correct path without increasing the error temporarily. All images are rendered at 1024 spp, see Tab. 1 for average iteration counts and render times. For better illustration, we only show caustic paths with 5 vertices found by MNEE and omit the other techniques and reflected caustics. The insets exposure is brighter by 2 stops.

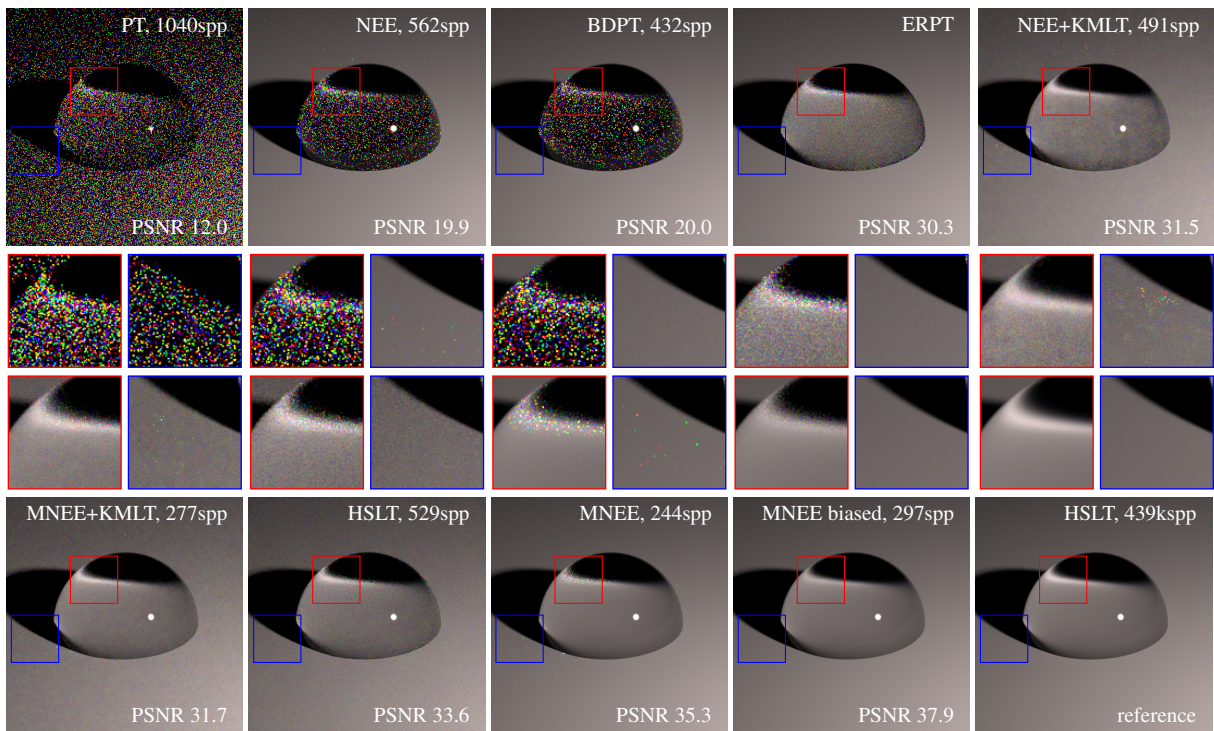


Figure 5: Equal time comparison renders (one minute) of a simple setup showing a water droplet on a flat plane, sorted by peak signal to noise ratio (PSNR). The closeups show the transmitted caustic (red) and the very subtle reflected caustic (blue). The scene consists of a diffuse ground plane and a sphere with IOR 1.33 (Abbe number 40) and Beckmann roughness $\alpha = 0.02$. See Fig. 7 for a schematic and Section 4 for details on the nine algorithms used



Figure 6: A sweaty character from a recent movie production demonstrating MNEE on complex geometry. MNEE was 45% slower than NEE here.

4. Results

We implemented MNEE in two rendering systems: a small research renderer and Weta Digital's *Manuka* (see Fig. 6). Equal time comparisons were done using 12 threads of an Intel Core i7-3930K CPU @ 3.20GHz hexacore machine. We use spectral rendering with continuously sampled wavelength. We do not use hero wavelength spectral sampling [WND*14] for easier comparisons between specular and rough interfaces, showing only the differences caused by MNEE and not by improved transport of wavelengths for the rough case.

Fig. 5 shows an equal time comparison of nine Markov chain- and pure Monte Carlo path space sampling methods. We compare pure path tracing (PT), path tracing with next event estimation and MIS (NEE), bidirectional path tracing (BDPT) [VG94], energy redistribution path tracing (ERPT) [CTE05], primary sample space MLT [KSKAC02] both with NEE and MNEE as sampler (NEE+KMLT and MNEE+KMLT, respectively), half vector space light transport (HSLT) [KHD14a] and MNEE with full MIS and in the biased variant (see Sec. 3.5).

The Markov chain methods KMLT and HSLT are not suitable for animations as discussed previously. KMLT shows low frequency noise even in these still images. HSLT will show similar effects for more complex scenes where discovering the interesting sub-regions of path space is a challenge, even though the distribution for the individual regions is more uniform due to a step size explicitly governed by ray differentials. HSLT can use a larger step size than KMLT, because the half vector space formulation will keep the point \mathbf{x}_k on the light source, whereas KMLT depends on small steps to achieve this. Smaller steps result in obvious blotches when rendering an animation. Both KMLT and HSLT use simple path tracing as seed, KMLT with and HSLT without NEE.

ERPT is using 512 mutations per chain and chains are only started for paths that touch the glossy sphere. ERPT suffers from sparse seed paths (the highlight has not yet been

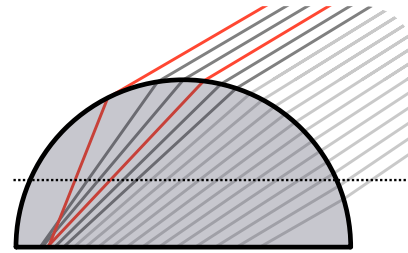


Figure 7: Vertical cross section of the water droplet shown in Fig. 5, with actual rays traced against the same interface (IOR 1.33). The red ray near grazing angle does not only compress the density on the plane, but connects to the same point as the other red ray. This kind of caustic folding onto itself results in ambiguous paths and MNEE as described in this paper can only find one of them.

discovered in this time frame because of the high mutation count) and the fact that the Markov chain is cut off after a fixed number of mutations. We started enough chains to bring down the average accumulation value to 0.2. Note that these parameters (number of mutations, number of chains) have a big impact on speed and our experiments have shown PSNR ranging from 14 to 30.

The perturbation used in MNEE to connect a point \mathbf{x}_b on a caustic to the light source efficiently reproduces the light under the water droplet. However, being a specialised technique, it cannot find all transport paths occurring in this scene. Reflected caustics (blue inset) are a problem as well as overlapping caustic paths (red inset, also see Fig. 7). Note that this means MNEE by itself is not an unbiased rendering algorithm, but needs to be embedded into a path tracer via MIS to find the paths for which MNEE has a zero sampling probability density. Although it is possible to use KMLT on top of a MNEE sampler to mutate these remaining paths, doing so will introduce MLT noise, resulting in a drop in PSNR. While it might be possible in this scene to run ERPT selectively or use some kind of adaptive sampling to resolve the remaining noise, using the biased version of MNEE and just discarding ambiguous paths (one of the red paths in Fig. 7) results in a clean, temporally stable image (this is why biased MNEE is a bit dimmer in the red inset). There are no such complications if the ground plane is moved up just a little (see dashed horizontal line in Fig. 7).

Fig. 8 shows the behaviour of MNEE with increasing roughness, in an equal time comparison to NEE. For very rough interfaces, regular next event estimation becomes more efficient because it can compute more samples in the same time, and increasingly rough surfaces make direct connections to the light sources more effective. We note that this is just an example, the actual difference in efficiency of either method depends strongly on the specifics of the scene, the ray tracing cost, and the relative size of objects on screen requiring MNEE to resolve caustics. MNEE is most effec-

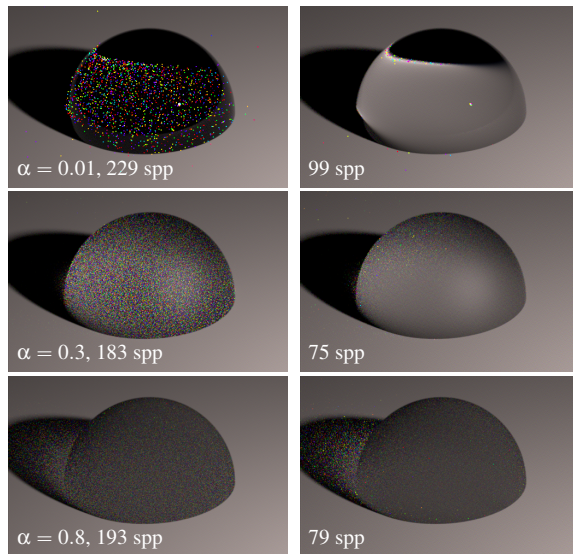


Figure 8: The droplet rendered at different roughnesses α , equal time comparison (20s) NEE (left) vs MNEE (right).

tive in contexts where ray tracing costs are high and there are not many small objects.

5. Limitations

We would like to underline how MNEE can work extremely well in difficult cases, while relying very strongly on the assumptions exposed earlier to hold. As is the case for other methods depending on differential geometry, and especially in the presence of glossy transport, a quality differential geometry framework is required, with a differentiable structure with good numerical stability (see Fig. 9 and 10). This has a tight relationship with the details of how geometric primitives are represented in the rendering system, and it can be easier to achieve for large analytical primitives than for finely tessellated surfaces, especially when run through programmable displacement. A number of approaches have been proposed in various domains that could be useful in this context: one is the introduction of partial curvature information as done for example in the case of PN triangles [VPBM01], another, more direct approach to the problem would be moving to advanced data structures based on discrete exterior calculus (DEC) theory either in the classical context following Whitney [Whi57] or in more modern formulations [Hir03, DHLM05]. In any event, for best results the displacement subsystem should be intimately connected with the geometric representation, and would need to be able to operate without corrupting its integrity. Whereas this is relatively easy to achieve for PN triangles, we are unaware of research in this domain in the context of DEC theory.

Next, note that the present formulation of MNEE's implementation does not consider reflected caustics or refracted

ones which lie outside the shadow of the transmissive object: this is because in these cases our method does not construct a seed path to start the process (\mathbf{x}_b and \mathbf{x}_c are not occluded from each other, see Fig. 11). We note that a procedure that would construct such seed paths is all that is needed to handle these configurations.

Lastly, the manifold walk will fail in two cases: first, if the base path is in a different subspace, the constraint derivatives will be meaningless. This can still converge but at high cost (see Fig. 4). It will depend on the complexity of the rest of the scene setup whether the net rendering efficiency will be higher or lower. Second, if there are multiple admissible paths for a given seed path (see Fig. 7) the Newton solver approach as described in this paper will only converge to one of them, and the use of a multimodal solver might become necessary. Despite all these limitations, a full implementation will still be unbiased (up to the notes in Sec. 3.5) because when MNEE fails to converge the PDF evaluates to zero, letting MIS hand over completely to the other techniques.

6. Conclusion and Future Work

We presented *manifold next event estimation* (MNEE), a technique to compute direct lighting through multiple, potentially rough refractive interfaces by correlated sampling in a pure Monte Carlo context, without resorting to Markov chains. The technique is relatively light weight and works well on important special cases such as water droplets on skin, and integrates well into the MIS framework for unbiased path tracing. In the future, we would like to lift the limitations outlined in the previous section, extend the technique to work with reflected caustics, and to explore the rich field of multimodal optimization techniques to improve the robustness to cases where multiple admissible paths correspond to a single seed path.

Acknowledgements We are grateful to Warner Bros. and Wingnut Films for their kind permission to use Gandalf the Grey from The Hobbit trilogy in the rendering for Fig 6. We would like to thank Guillaume François to model and setup the shot for us, and Joe Letteri for making this work possible.

References

- [Cha60] CHANDRASEKAR S.: *Radiative Transfer*. Dover Publications Inc., 1960. ISBN 0-486-60590-6. 2
- [CTE05] CLINE D., TALBOT J., EGBERT P. K.: Energy redistribution path tracing. *ACM Trans. on Graphics (Proc. SIGGRAPH)* 24, 3 (2005), 1186–1195. 3, 8
- [dC76] DO CARMO M.: *Differential Geometry of Curves and Surfaces*. Prentice-Hall, 1976. 3
- [DGN85] DEJONGHE G., GONNORD J., NIMAL J.: Perturbation calculations by the correlated samples method. In *Monte Carlo Methods and Applications in Neutronics, Photonics and Statistical Physics*, vol. 240 of *Lecture Notes in Physics*. Springer Berlin Heidelberg, 1985, pp. 311–323. 4

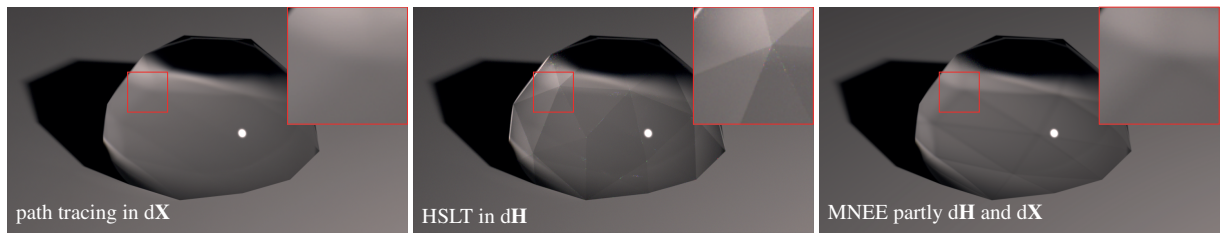


Figure 9: Bias introduced by non C_2 continuous geometry: since the normal derivatives are part of the measurement contribution in half vector space via the transfer matrix in Eq. (6), approximate derivatives will lead to approximate path contributions. This requirement to the renderer is shared between manifold exploration, HSLT, and MNEE. Note that the tessellation here is purposefully extremely poor. Continuous normal derivatives can be achieved e.g. by using PN triangles [VPBM01].

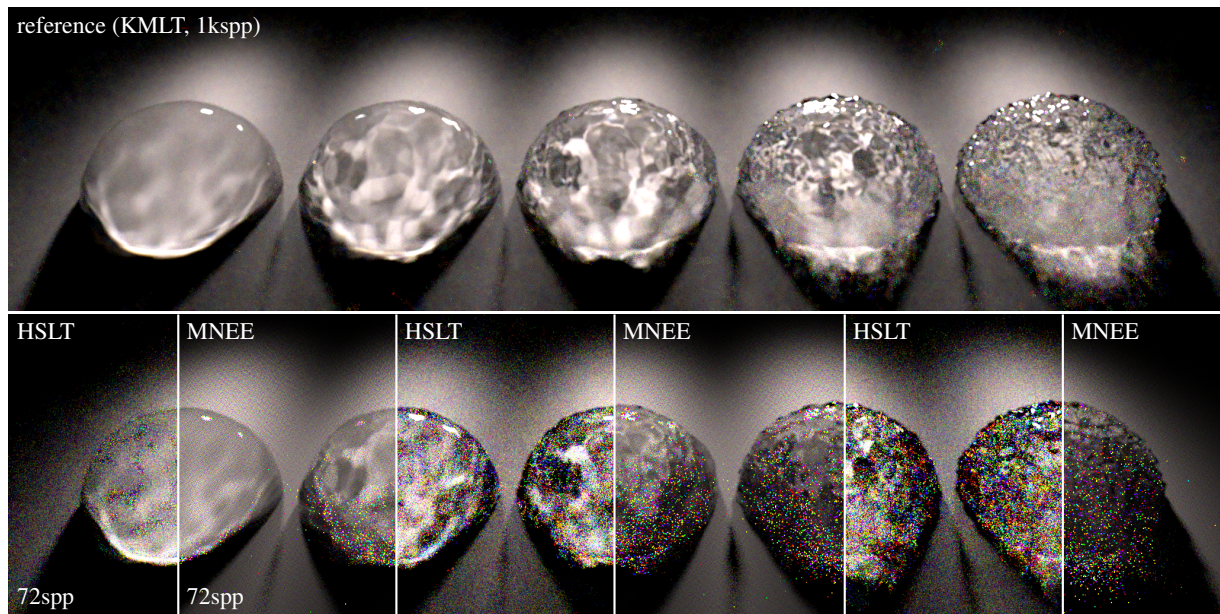


Figure 10: Fail case for MNEE due to displacement: equal sample count comparison of renders of water droplets with displacement texture with increasing frequency (each droplet has twice the frequency than the droplet to its left), illustrating the capability of different algorithms to find paths in this difficult setting. Fine displacement makes it hard to exploit geometric derivatives, so both MNEE and HSLT struggle with the rightmost droplet. Since the seed paths are closer to the next sample and constitute valid paths in the case of HSLT, it works a little better than MNEE. Note that even KMLT has a hard time exploiting coherence in path space in this case.

[DHLM05] DESBRUN M., HIRANI A. N., LEOK M., MARSDEN J. E.: Discrete exterior calculus, 2005. [arXiv:arXiv:math/0508341](https://arxiv.org/abs/math/0508341). 9

[Erm75] ERMAKOW S. M.: *Die Monte Carlo Methode und verwandte Fragen*. VEB Deutscher Verlag der Wissenschaften, 1975. 2

[EW11] ENDERTON E., WEXLER D.: The workflow scale: Why 5x faster might not be enough. In *Computer Graphics International Workshop on VFX, Computer Animation, and Stereo Movies* (2011). 1

[GKDS12] GEORGIEV I., KRIVANEK J., DAVIDOVIC T., SLUSALLEK P.: Light transport simulation with vertex connection and merging. *ACM Trans. on Graphics (Proc. SIGGRAPH Asia)* 31, 6 (2012), 192:1–192:10. 3

[Has70] HASTINGS W. K.: Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57, 1 (1970), 97–109. 3

[HD14] HEITZ E., D'EON E.: Importance sampling microfacet-based BSDFs using the distribution of visible normals. *Computer Graphics Forum (Proc. Eurographics Symposium on Rendering)* 33, 4 (July 2014), 103–112. 5

[Hir03] HIRANI A. N.: *Discrete Exterior Calculus*. PhD thesis, California Institute of Technology, May 2003. URL: <http://resolver.caltech.edu/CaltechETD:etd-05202003-095403>. 9

[HJ09] HACHISUKA T., JENSEN H. W.: Stochastic progressive photon mapping. *ACM Transactions on Graphics* 28, 5 (2009), 1–8. 3

- [HJ11] HACHISUKA T., JENSEN H. W.: Robust adaptive photon tracing using photon path visibility. *ACM Trans. on Graphics* 30, 5 (2011), 114. 3
- [HPJ12] HACHISUKA T., PANTALEONI J., JENSEN H. W.: A path space extension for robust light transport simulation. *ACM Trans. on Graphics (Proc. SIGGRAPH Asia)* 31, 6 (2012), 191:1–191:10. 3
- [Jak13] JAKOB W.: *Light transport on path-space manifolds*. PhD thesis, Cornell University, 2013. 2, 3, 4, 6
- [Jen96] JENSEN H. W.: Global illumination using photon maps. In *Proc. Eurographics Workshop on Rendering* (1996), pp. 21–30. 3
- [Kaj86] KAJIYA J. T.: The rendering equation. *Computer Graphics (Proc. SIGGRAPH)* (1986), 143–150. 2
- [KD13a] KAPLANYAN A., DACHSBACHER C.: Adaptive progressive photon mapping. *ACM Trans. on Graphics* 32, 2 (Apr. 2013), 16:1–16:13. 3
- [KD13b] KAPLANYAN A. S., DACHSBACHER C.: Path space regularization for holistic and robust light transport. *Computer Graphics Forum (Proc. of Eurographics)* 32, 2 (2013), 63–72. 3, 6
- [KHD14a] KAPLANYAN A., HANIKA J., DACHSBACHER C.: The natural-constraint representation of the path space for efficient light transport simulation. *ACM Trans. on Graphics (Proc. SIGGRAPH)* 33, 4 (2014), 1–13. 2, 3, 4, 5, 6, 8
- [KHD14b] KAPLANYAN A., HANIKA J., DACHSBACHER C.: Supplemental material to “The Natural-Constraint Representation of the Path Space for Efficient Light Transport Simulation”. *ACM Trans. on Graphics (Proc. SIGGRAPH)* 33, 4 (2014), 1–13. 5
- [KMA*15] KETTUNEN M., MANZI M., AITTALA M., LEHTINEN J., DURAND F., ZWICKER M.: Gradient-domain path tracing. *ACM Trans. on Graphics (Proc. SIGGRAPH)* (2015), to appear. 4
- [KSKAC02] KELEMEN C., SZIRMAY-KALOS L., ANTAL G., CSONKA F.: A simple and robust mutation strategy for the Metropolis light transport algorithm. *Computer Graphics Forum* 21, 3 (2002), 531–540. 3, 8
- [LKL*13] LEHTINEN J., KARRAS T., LAINE S., AITTALA M., DURAND F., AILA T.: Gradient-domain Metropolis light transport. *ACM Trans. on Graphics (Proc. SIGGRAPH)* 32, 4 (2013), 95:1–95:12. 4
- [LW93] LAFORTUNE E., WILLEMS Y.: Bi-directional path tracing. In *Proc. of COMPUGRAPHICS* (1993), pp. 145–153. 3
- [Mat63] MATTHES W.: Monte Carlo calculations of the nuclear temperature coefficient in fast reactors. In *Proc. 3rd Symposium “Reactor Theory” of EURATOM, EUR-309-e* (1963). 4
- [Mol07] MOLINARI L. G.: Determinants of block tridiagonal matrices. [arXiv:0712.0681\[math-ph\]](https://arxiv.org/abs/0712.0681). 5
- [MRR*53] METROPOLIS N., ROSENBLUTH A. W., ROSENBLUTH M. N., TELLER A. H., TELLER E.: Equation of state calculations by fast computing machines. *The Journal of Chemical Physics* 21, 6 (1953), 1087–1092. 3
- [NRH*77] NICODEMUS F., RICHMOND J., HSIA J., GINSBERG I., LIMPERS T.: Geometrical considerations and nomenclature for reflectance. *Final Report, National Bureau of Standards, Washington, DC. Inst. for Basic Standards* (1977). 3
- [Rie84] RIEF H.: Generalized Monte Carlo perturbation algorithms for correlated sampling and a second-order Taylor series approach. *Annals of Nuclear Energy* 11, 9 (1984), 455 – 476. 4

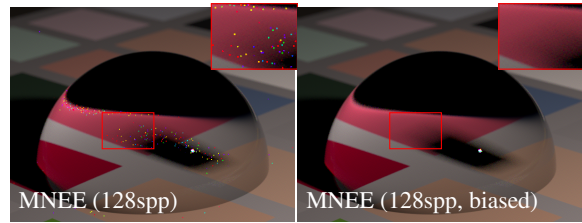


Figure 11: Fail case for MNEE due to occlusion: a little cube (upper right corner outside the viewing frustum) occludes the seed path, so we do not attempt to perturb it even though this would result in an admissible path. Note that our outlier removal transparently detects such cases, too (right image).

- [Sal06] SALKUYEH D. K.: Comments on "a note on a three-term recurrence for a tridiagonal matrix". *Applied Mathematics and Computation* 176, 2 (2006), 442–444. 4
- [Sob94] SOBOL' I.: *A Primer for the Monte Carlo Method*. CRC Press, 1994. 2
- [Ste05] STEPHENSON I.: *Production rendering: design and implementation*. Springer, London, Berlin, Heidelberg, 2005. 1
- [SW91] SHIRLEY P., WANG C.: Direct lighting calculation by monte carlo integration. In *Proc. Eurographics Workshop on Rendering* (1991), pp. 54–59. 3, 4
- [UFK13] UREÑA C., FAJARDO M., KING A.: An area-preserving parametrization for spherical rectangles. *Computer Graphics Forum (Proc. Eurographics Symposium on Rendering)* 32, 4 (2013), 59–66. 3
- [Vea98] VEACH E.: *Robust Monte Carlo methods for light transport simulation*. PhD thesis, Stanford University, 1998. AAI9837162. 2
- [VG94] VEACH E., GUIBAS L.: Bidirectional estimators for light transport. In *Proc. Eurographics Workshop on Rendering* (1994), pp. 147–162. 3, 8
- [VG95] VEACH E., GUIBAS L. J.: Optimally combining sampling techniques for Monte Carlo rendering. *Proc. SIGGRAPH* (1995), 419–428. 2, 3
- [VG97] VEACH E., GUIBAS L. J.: Metropolis light transport. *Proc. SIGGRAPH* (1997), 65–76. 3
- [VPBM01] VLACHOS A., PETERS J., BOYD C., MITCHELL J. L.: Curved PN triangles. In *Proceedings of the 2001 Symposium on Interactive 3D Graphics* (2001), I3D '01, pp. 159–166. 9, 10
- [Whi57] WHITNEY H.: *Geometric Integration Theory*. Princeton University Press, 1957. 9
- [WMLT07] WALTER B., MARSCHNER S., LI H., TORRANCE K.: Microfacet models for refraction through rough surfaces. In *Proc. Eurographics Symposium on Rendering* (2007), pp. 195–206. 5
- [WND*14] WILKIE A., NAWAZ S., DROSKE M., WEIDLICH A., HANIKA J.: Hero wavelength spectral sampling. *Computer Graphics Forum (Proc. Eurographics Symposium on Rendering)* 33, 4 (July 2014), 123–131. 8
- [WZHB09] WALTER B., ZHAO S., HOLZSCHUCH N., BALA K.: Single scattering in refractive media with triangle mesh boundaries. *ACM Trans. on Graphics (Proc. SIGGRAPH)* 28, 3 (July 2009), 92:1–92:8. 3