Caustic Connection Strategies for Bidirectional Path Tracing

Sebastien Speierer, Christophe Hery, Ryusuke Villemin, Wenzel Jakob

Pixar Technical Memo #18-01



Figure 1: Caustic Connection strategies enable the bidirectional path tracer to render complex indirect caustics as we can observe in the right eye of this digital character. In this setup, the entire lighting is indirect, first bouncing on the wall before illuminating the character

Abstract

We propose a new type of sampling strategy for connection-based path tracing algorithms such as bidirectional path tracing. Classical bidirectional path tracing generally exhibits poor performance when sampling light paths involving specular transport (e.g. refraction through dielectrics). We therefore introduce specialized connection strategies that connect through chains of specular events. By applying mechanisms such as *manifold exploration*, we propose an efficient solution for connecting two points in the scene where a sequence of refractive specular surfaces lies between them. We also introduce a lightweight scheme for recursively computing multiple importance sampling weights during path creation. The resulting algorithm is easy to implement and leads to significant improvements in constructing complex caustic paths.

1 Introduction

While today modern rendering systems are very efficient at simulating complex light paths in complex environments, rendering refractive caustics still often takes too much time for production renders due to the low probability of sampling such paths.

Caustics are light subpaths composed of a chain of reflective or refractive specular events. Due to the sharp density distribution of these specular events, rendering algorithms mostly rely on directly sampling the BSDF at these surfaces to build these paths.

Bidirectional path tracing combines many different ways of constructing paths via a random walk from both endpoints that are then combined using a deterministic connection. In a state-of-the-art implementation of a bidirectional path tracer, each step is generally implemented using an ideal or close to ideal importance sampling scheme that reproduces the associated portion of the underlying integrand. Unfortunately, specular paths imply complex dependencies in path space that are impossible to capture by such an iterative manner. In practice, this means that the generated light and camera subpaths cannot be connected because little or no transport occurs over the connection segment. *Specular-diffuse-specular (SDS)* path contains sequences alternating diffuse and specular events. The construction of such path has to statisfy multiple half-vector contraints [Kaplanyan et al. 2014] simultaneously. As described on Figure 2, the current set of deterministic connection schemes will fail at constructing SDS paths.



Figure 2: Different connection strategies

Hanika et al. [2015] introduced an extension to the Next Event Estimation [Veach 1997] method to find refractive caustic connecting end of path to a light source. Their algorithm uses *manifold exploration* [Jakob and Marschner 2012] to construct a caustic subpath that satisfies the half-vector constraints. In practice, caustic subpaths can occur anywhere along the light path, while this method only construct caustic subpath at the end of the camera path.

In this paper, we extend the set of strategies of the bidirectional path tracing algorithm to correctly and efficiently handle caustic subpaths exclusively composed of refractive specular events, called *re-fractive caustics*. We present sampling strategies that use the *man-ifold exploration* during vertex connection in a bidirectional path tracer to sample caustic subpaths anywhere along the full path. This will allow the bidirectional path tracer to deal with any type of re-fractive caustic subpaths anywhere along the complete path, even if the path contains multiple refractive caustic subpaths which greatly improves the number of cases where the bidirectional path tracer is successful.

When combining a large variety of sampling strategies, efficient computation of multiple importance sampling weights is crucial. For instance, a naive computation that loops over all possible strategies results in a prohibitively expensive $O(N^4)$ implementation. Therefore, we also derive a lightweight scheme based on the work of Antwerpen [2011] to recursively compute the multiple importance sampling weights for the different strategies. The recursive mechanism of these computations can easily be implemented in modern rendering system.

The remainder of this paper is organized as follows: Section 2 looks at different methods addressing the rendering of refractive caustic paths, their strength and their faults. Section 3 discusses our new sampling strategies in detail and Section 3.3 describes a recursive scheme to efficiently compute the MIS weights. Finally, Section 4 presents some results and comparisons.

2 Previous work

Over the years, many groups have investigated techniques for better handling and sampling refractive caustic paths.

Bidirectional Path Tracing Bidirectional path tracing introduced by Veach [1997] uses a large set of sampling strategies to connect any two vertices on two subpath, one strating at the camera and one starting on a light source. Even with all the different sampling strategies used in the bidirectional path tracing, this method cannot properly handle paths containing an SDS sequence. This is mainly due to the specular and near-specular vertices, where it is highly unlikely that a connection satisfies their half-vector constraints.

Photon mapping Density estimation methods like [Georgiev et al. 2012; Hachisuka and Jensen 2009] create a data structure called *photon map* and store a large number of photons shot from the light sources in a pre-rendering phase. Later, at render-time, the renderer can query this map to estimate the density of incoming energy at a specific point on a surface in the scene.

Density estimation techniques unfortunately suffer from intensive memory usage and are biased. Furthermore, if the refractive caustics only lie on a small portion of the scene, it will be a waste of computations to trace millions of photons onto the entire scene, while only a small fraction of them will end up near these caustic regions.

Markov Chain Monte Carlo techniques Based on Markov chains, MCMC techniques like [Hachisuka et al. 2014; Bitterli et al. 2017; Šik et al. 2016], provide an elegant procedure to explore the path space based on previous iterations. At every iteration, they use mutations to change the structure of the path and perturbations to move vertices by small distances. The drawback of using Markov Chains in this context is their lack of temporal coherence, and the fact that the integrator relies on exploration, therefore making judging a non fully converged image impossible. The goal of this paper is to pursue methods that are suitable for industrial applications where MCMC is not traditionally used, hence the need for a scheme that works in a pure Monte Carlo context.

The *manifold exploration* framework, [Jakob and Marschner 2012], introduces mutation strategies in the MLT context, efficient at rendering indirect caustics. Using an iterative process similar to Newton's method that is informed by the differential geometry of the light path, it computes a new path that satifies all the physical constraints sequence of specular events.

Manifold Next Event Estimation - MNEE Hanika et al. [2015] introduced an extension of the Next Event Estimation technique, using *manifold exploration* to connect surfaces to light sources across a sequence of refractive events. It doesn't suffer from intense memory usage, neither from temporal stability issues, which makes it

suitable in a production context.

The downside of this method is that it only handles refractive caustic subpaths in direct illumination with the light sources. Often, a large amount of the energy coming into the scene results from several bounces, illuminating the scene indirectly. Unfortunately, MNEE is of no help in this case.

3 Caustic Connection strategies

Our method introduces new sampling strategies in the context of bidirectional path tracing, exploiting similar mechanisms as in the MNEE method to determistically connect vertices across refractive surfaces. Given the very low probability of a connection to satisfy the half-vector contraints at specular surfaces, these new strategies help finding refractive caustic subpaths where other strategies often fail.

As described in Algorithm 1, our strategies can easily be added to the already existing set of strategies in any implementation of a bidirectional path tracer. Whenever the connection fails due to obstruction, our method is more persistent and attempts a *manifold exploration* to resolve any refractive caustic subpath between the two vertices if the obstruction was due to refractive objects. We then compute the appropriate MIS weight and add the contribution of the computed path.

Algorithm 1 Connection procedure	
1:	procedure CONNECTION
2:	for $i = 1$ to C.length do
3:	for $j = 1$ to L.length do
4:	if InvalidConnection (i, j) then
5:	continue
6:	if VisibilityCheck (i, j) then
7:	$res \leftarrow RegularConnection(C, L, i, j)$
8:	else
9:	$res \leftarrow ManifoldWalkConnection(C, L, i, j)$
10:	$w \leftarrow \text{ComputeMISWeight}(C, L, i, j, \text{res})$
11:	AddContribution(res, w)

3.1 Throughput computation

It is important to define how to compute the contribution of a path sampled by these strategies. Given the camera subpath

$$\bar{p}_t = p_0, p_1, \dots, p_{t-1},$$

and the light subpath

$$\bar{q}_s = q_0, q_1, \dots, q_{s-1},$$

Veach [1997] describes the contribution of a path sampled using a regular connection as

$$P(\bar{q}_s \bar{c}_n \bar{p}_t) = L_e T(\bar{q}_s) T(\bar{p}_t) W_e$$

$$\times f(q_{s-2} \to q_{s-1} \to p_{t-1}) f(q_{s-1} \to p_{t-1} \to p_{t-2})$$

$$\times G(q_{s-1} \leftrightarrow p_{t-1}),$$

where $T(\bar{p}_t)$ and $T(\bar{q}_s)$ are the throughput of the camera subpath and light subpath respectively.

In the case of caustic connection strategies, we also need to account for the extra refractive interfaces along the connection. With a connection subpath

$$\bar{c}_n = c_0, \dots, c_{n-1}$$
 where $c_{-1} = p_{t-1}$ and $c_n = q_{s_1}$,

and based on Jakob [2013], we can extend this equation using the generalized geometry term:

$$P(\bar{q}_s \bar{c}_n \bar{p}_t) = L_e T(\bar{q}_s) T(\bar{p}_t) W_e$$

$$\times f(q_{s-2} \to q_{s-1} \to c_{n-1}) f(c_0 \to p_{t-1} \to p_{t-2})$$

$$\times \underbrace{\bar{G}(q_{s-1} \leftrightarrow \dots \leftrightarrow p_{t-1})}_{\text{generalized geometry term}}$$

$$\times \underbrace{\prod_{i=0}^{n-1} f(c_{i+1} \to c_i \to c_{c-1})}_{\text{independent of the observed of the set of the observed of the set of the$$

throughput along the specular chain

3.2 Direct Geometry Term

If a connection involves a vertex with a Dirac delta BSDF such as a specular event, it will never succeed since there is zero probability of sampling the direction where this delta distribution is not zero. In practice, a bidirectional path tracer will not attempt any connections with specular vertices and will not account for these delta dirac terms during the MIS weight computation for numerical stability reasons since they will always occur both in the numerator and the denominator.

We can see a specular event as light path router which doesn't bring any additional "sampling entropy" but simply re-orient the light rays. Therefore, the sampled direction at a diffuse vertex, following by a sequence of specular events, deterministically defines the position where the ray hits on the next diffuse surface along this path. Since we are dealing with probability in area unit, we need a way to relate the solid angle at the first diffuse vertex to an area at the diffuse vertex at the end of the specular chain. The *geometry term* describes this change of variables, computing the derivative of projected solid angle at the first diffuse vertex with respect to area at the following diffuse vertex. However, it is not taking the sampling densities of the specular vertices in the MIS weights computations into account.

Jakob [2013] introduces a so called *generalized geometry term* which directly links two diffuse surfaces across a sequence of specular events. This *generalized geometry term* represents the "derivative of solid angle at one end of the specular chain with respect to area at the other end of the chain, considering the path as a function of the positions of the endpoints.". However, its computation is not trivial and can result in a significant drop of performance since it involves inversion of large matrices and will need to be computed frequently.



Figure 3: Representation of a direct geometry term across a single refractive interface

Therefore, we propose to approximate this term with a new geometry term called the **direct geometry term**, which is easier to compute and still accounts for most of the density ratio along the specular chain.

For a path \bar{p}_t composed of t vertices, including the specular vertices

$$\bar{p}_t = p_0, p_1, \dots, p_{t-1}$$

we define two functions that link a vertex index to the index of the closest diffuse vertices along the path:

$$D_i^- =$$
 index of the last diffuse vertex before p_i along \bar{p}_i

 $D_i^+ =$ index of the next diffuse vertex after p_i along \bar{p}_t

The *direct geometry term* between two consecutive diffuse vertices along a path is given by:

$$G(p_i \leftrightarrow p_{D_i^+}) = \frac{|\cos(\theta_i)||\cos(\theta_{D_i^+})|}{\|p_i - p_{D_i^+}\|^2}$$
$$G(p_{D_i^-} \leftrightarrow p_i) = \frac{|\cos(\theta_{D_i^-})||\cos(\theta_i)|}{\|p_{D_i^-} - p_i\|^2}$$

Notice that in the case where the two consecutive vertices are not seperated by any specular vertices, the direct geometry term becomes the regular geometry term. Figure 3 illustrates the different angles used to compute a direct geometry term across a single refractive interface.

When dealing with rough dielectric surfaces, since their density functions are not delta dirac functions, it makes perfect sense to use the standard geometry term and attempt regular connection involving these surfaces. However, it is still unlikely to correctly connect "near-specular" surfaces and therefore, in our implementation, we decided to have a clear separation between *connectable* vertices and *non-connectable* vertices by adding a attribute to the surface properties. Specular vertices are always *non-connectable* and the rough dielectric might be depending on their roughness.

Our implementation of the *manifold exploration* used for the caustic connection strategies only explores chains of *connectable* vertices. Since connections never start or end at a *non-connectable* vertex, we ensure that for a given specular chain, there won't be any other estimators constructing the same path by connecting two vertices accross a sub-chain of this specular chain, which greatly simplifies the MIS weight computations. We can say that for a given path, every specular vertex belongs to a unique specular chain and there exists a unique caustic connection strategy that can find a connection accross this specular chain.

3.3 Multiple Importance Sampling

Multiple importance sampling is a variance reduction technique introduced by Veach [1997] and is highly used in rendering systems. It allows one to efficiently combine multiple sampling strategies in a Monte Carlo context, while keeping the results unbiased and getting the best from each strategy. Bidirectional path tracing extensively takes advantage of this method to combine its different estimators.

In the interest of reducing the computational cost of such computations, there has been a lot of effort put into computing these weights more efficiently, such as the work from [Antwerpen 2011] where the authors propose a framework for recursively computing the MIS weights during the path construction. In this section, we will elaborate on an extention of this framework which accounts for the new sampling strategies.

3.3.1 Balance heuristic formulation

In the same way that Jakob [2013] reduces the dimensionality of the path space by taking out specular events along the path, we are going to abstract the specular vertices along our path during the MIS weights computations and link the diffuse vertices with direct geometry terms instead for MIS weights computations.

Using the same notation as in Section 3.2, we can approximate the sampling area densities at the vertices using direct geometry terms:

$$p_{\overrightarrow{A}}(p_i) = p_w(p_{(D_i^-+1)} - p_{D_i^-}) \frac{|\cos(\overleftarrow{\theta_i})|}{||p_{D_i^-} - p_i||^2}$$
$$p_{\overleftarrow{A}}(p_i) = p_w(p_{(D_i^+-1)} - p_{D_i^+}) \frac{|\cos(\overrightarrow{\theta_i})|}{||p_{D_i^+} - p_i||^2}$$

Moreover, if p_i is not connectable, then

$$p_{\overrightarrow{A}}(p_i) = p_{\overleftarrow{A}}(p_i) = 1$$

since we use direct geometry terms and consider nonconnectable vertices as *light path routers*.

The approximation of the probability of sampling a path \bar{p}_t connecting p_s and $p_{D_s^+}$ using either a regular connection if $D_s^+ = s + 1$ or a caustic connection otherwise is given by the following equation:

$$p_s(X) = \prod_{\substack{i=-1\\ \text{camera subpath}}}^{s-1} p_{\overrightarrow{A}}(p_{i+1}) \times \underbrace{\prod_{\substack{i=D_s^++1\\ \text{light subpath}}}^{k+1} p_{\overleftarrow{A}}(p_{i-1})}_{\text{light subpath}}$$

As we know, multiple importance sampling can be applied using a multitude of different heuristics to compute its weights. Results will stay unbaised while the weights are computed in a consistent way for all the different estimators. Therefore, the algorithm will still converge if we use an approximation of the sampling probability density function of a path for the balance heuristic introduced by Veach [1997].

Antwerpen [2011] introduces a recursive scheme for computing the balance heuristic weights, which greatly improves the effiency these computations and only requires information at the last vertex on both camera and light subpath. We have demonstrated that we can use the same recursive scheme with our approximate area sampling densities to compute MIS weights that account for our new strategies. The recursive terms are computed as follow:

$$d_s^E = \begin{cases} \frac{1}{p_{\overrightarrow{A}}(p_1)} & s = 1\\ \frac{p_{\overrightarrow{A}}(p_{s-1})d_{s-1}^E}{p_{\overrightarrow{A}}(p_s)} & p_s \text{ is not connectable}\\ \frac{1+p_{\overrightarrow{A}}(p_{s-1})d_{s-1}^E}{p_{\overrightarrow{A}}(p_s)} & \text{otherwise} \end{cases}$$

and

$$d_s^L = \begin{cases} \frac{1}{p_{\overleftarrow{A}}(p_k)} & s = k-1\\ \frac{p_{\overrightarrow{A}}(p_{s+2})d_{s+1}^L}{p_{\overleftarrow{A}}(p_{s+1})} & p_{s+1} \text{ is not connectable}\\ \frac{1+p_{\overrightarrow{A}}(p_{s+2})d_{s+1}^L}{p_{\overleftarrow{A}}(p_{s+1})} & \text{otherwise} \end{cases}$$

We can finally compute the inverse MIS weight for a strategy starting a connection at p_s :

$$\frac{1}{w_s(X)} = 1 + p_{\overleftarrow{A}}(p_s) \times d_s^E + p_{\overrightarrow{A}}(p_{D_s^+}) \times d_{(D_s^+ - 1)}^L$$

In these formulations, when the camera or light subpath contains a sequence of specular vertices, we assume that there exists a caustic connection strategy that could sample this path across that specular chain. However, it is not always true in practice due to the complexity of the manifold exploration when dealing with total internal reflection, complex geometries, or a long sequence of specular interfaces. For this method to stay unbaised, whenever we account for caustic connection strategy in the MIS recursive term, we should test if this connection could have happened in pratice. Therefore, we run a manifold exploration for only a few iterations and compare the found caustic subpath with the specular chain we have sampled during the random walk. If these subpaths are close enough, then we can account for this caustic connection strategy in the MIS weight.

3.4 Optimizations

Russian roulette

Computing a caustic connection using the *manifold exploration* algorithm is relatively expensive due to the many shader evaluations at each iteration before finding the correct path. Therefore, as an optimization, we propose to apply the concept of russian roulette to our strategies in order to reduce the amount of *manifold explorations* at render time while keeping the method unbiased.

We have tried different heuristics to determine the threshold to use for the russian roulette, such as the length of the camera and light subpaths, their throughput, and combination of these two. The simple heuristic based on the length of the two subpaths works relatively well in general and requires less user adjustment. In practise, it reduces the amount of time spent in *manifold exploration* by 30%.

The *efficiency-optimized russian roulette* from [Veach 1997] computes the roulette threshold value based on the efficiency metric of the estimator so as to maximize the efficienty of the resulting estimator. This should perform better than the standard russian roulette and remove the burden of having to choose the threshold value ourself. However, this method requires the integrator to keep track of the average variance and cost of a given estimator (per pixel). In practice, the overhead cost and complexity of using this method proved not to be beneficial compared to our heuristic.

Manifold exploration caching

Another way to reduce the cost of running many *manifold explorations* is to use some caching mechanism, exploiting the similarities of different caustic subpaths and reuse the solution of previous *manifold explorations*. The idea is to use those cached solutions as initial state for the current *manifold exploration* rather than using the *seed path*. This should lead to an initial caustic subpath closer to the true solution and therefore reduce the total number of iterations.

This raises the questions of how to best represent a connection configuration, its solution, and what distance metric should we use to evaluate the fitness of a cached solution given the current configuration. The caustic subpath is enterly determined by the position of its two endpoints and the orientations of the specular surfaces. Moreover, the curvatures of those surfaces define the condition number



Figure 4: The Caustic Connection strategies efficiently handle any types of refractive caustic subpaths, anywhere along the full light path and clearly outpass a regular bidirectional path tracer.

of this configuration, a crutial information defining the reliability of a cached solution.

After rendering scenes of various complexity, we made the observation that the *manifold exploration* was taking 4 iterations in average, where each iteration involves shadow tests and shader evaluations.

In the *manifold exploration*, each iteration assumes constant curvature at the specular surfaces. In a similar way, our caching mechanism uses a distance metric that only relies on local information, and therefore, also assumes constant properties on the surfaces. We obeserve that in the best case, our caching mechanism would play a similar role as the first iteration of the *manifold exploration*.

In conclusion, the *manifold exploration* is too ill-conditioned to be compatible with such caching mechanism. On top of the overhead costs of this caching mechanism, the reduction in the number of iterations for the *manifold explorations* was minimal.

4 Results

Figure 4 shows equal sample per pixel comparisons of the regular BDPT and BDPT with our new strategies under different lighting conditions. In Figure 4a, all the light illuminating the scene first bounces on diffuse panel. BDPT properly handles this situation of indirect lighting for diffuse-diffuse interactions but struggles finding the refractive caustic path since it cannot do any connections on the specular hemisphere. Moreover, the MNEE method in this case would be of no help since it can only connect diffuse surfaces across specular chains directly to the light sources. On the other hand, the caustic connection strategies easily find the refractive caustics, by connecting the diffuse floor to the diffuse panel across the refractive hemisphere. As we can see, even with only a few samples, with our method, the caustic regions are almost as converged as the noncaustic regions. Figure 4b compares another case where all the light paths contain at least one specular event. This time, the BDPT can only rely on the strategies that sample most of the path, either from the camera or the light, since it can't connect anything in between. The caustic connection strategies add many ways to construct the path in this case, which greatly help the renderer. Finally, Figure 4c shows a case where all the light paths end with a SDS subpath. While this is a complex scenario for any rendering algorithm, our connection strategies can find this path in many ways: either by connecting the diffuse floor to the diffuse panel across some refractive interfaces, or by connecting the diffuse panel directly to the light source across the refractive hemisphere. Once more, even with a small number of samples, our method finds most of the refractive caustic paths and clearly outpass the regular BDPT.

Figure 1 demonstrates a use case of the new connection strategies introduced in this paper in a production setup. Thanks to the caustic connections strategies, the bidirectional path tracer properly renders indirect caustics on the irises of the digital character.

5 Conclusion

In this paper, we presented a novel set of sampling strategies for bidirectional path tracer. These strategies are specialized in connecting diffuse vertices across a sequence of refractive interfaces and help the BDPT to sample paths containing complex refractive caustic subpath. Our method solves the problem of rendering indirect refractive caustics, where state-of-the-art techniques such a Manifold Next Event Estimation fail. It also doesn't suffer from temporal uncoherency, and runs with a very low memory impact which makes it really suitable for rendering in the industry.

Our new strategies could easily be integrated in other connectionbased rendering algorithms such as VCM. We leave this for future study.

Acknowledgements

Thanks to Mike Seymour and Disney Research for the digital character dataset and to Junyi Ling for helping setting up the renders.

References

- ANTWERPEN, D. 2011. Recursive mis computation for streaming bdpt on the gpu. Tech. rep., Technical report/Delft University of Technology–2011.
- BITTERLI, B., JAKOB, W., NOVÁK, J., AND JAROSZ, W. 2017. Reversible jump metropolis light transport using inverse mappings. ACM Trans. Graph. 37, 1 (Oct.), 1:1–1:12.
- GEORGIEV, I., KRIVÁNEK, J., DAVIDOVIC, T., AND SLUSALLEK, P. 2012. Light transport simulation with

vertex connection and merging. ACM Trans. Graph. 31, 6, 192-1.

- HACHISUKA, T., AND JENSEN, H. W. 2009. Stochastic progressive photon mapping. *ACM Trans. Graph.* 28, 5 (Dec.), 141:1–141:8.
- HACHISUKA, T., KAPLANYAN, A. S., AND DACHSBACHER, C. 2014. Multiplexed metropolis light transport. *ACM Trans. Graph.* 33, 4 (July), 100:1–100:10.
- HANIKA, J., DROSKE, M., AND FASCIONE, L. 2015. Manifold next event estimation. In *Computer Graphics Forum*, vol. 34, Wiley Online Library, 87–97.
- JAKOB, W., AND MARSCHNER, S. 2012. Manifold exploration: a markov chain monte carlo technique for rendering scenes with difficult specular transport. ACM Transactions on Graphics (TOG) 31, 4, 58.
- JAKOB, W. A. 2013. *Light transport on path-space manifolds*. PhD thesis, Cornell University.
- KAPLANYAN, A. S., HANIKA, J., AND DACHSBACHER, C. 2014. The natural-constraint representation of the path space for efficient light transport simulation. ACM Trans. Graph. 33, 4 (July), 102:1–102:13.
- PHARR, M., JAKOB, W., AND HUMPHREYS, G. 2016. *Physically based rendering: From theory to implementation*. Morgan Kaufmann.
- VEACH, E. 1997. *Robust monte carlo methods for light transport simulation.* PhD thesis, Citeseer.
- ŠIK, M., OTSU, H., HACHISUKA, T., AND KŘIVÁNEK, J. 2016. Robust light transport simulation via metropolised bidirectional estimators. ACM Trans. Graph. 35, 6 (Nov.), 245:1–245:12.