

Importance Sampling Techniques for Path Tracing in Participating Media

Christopher Kulla¹ and Marcos Fajardo²

¹Sony Pictures Imageworks, Culver City, USA

²Solid Angle, Madrid, Spain

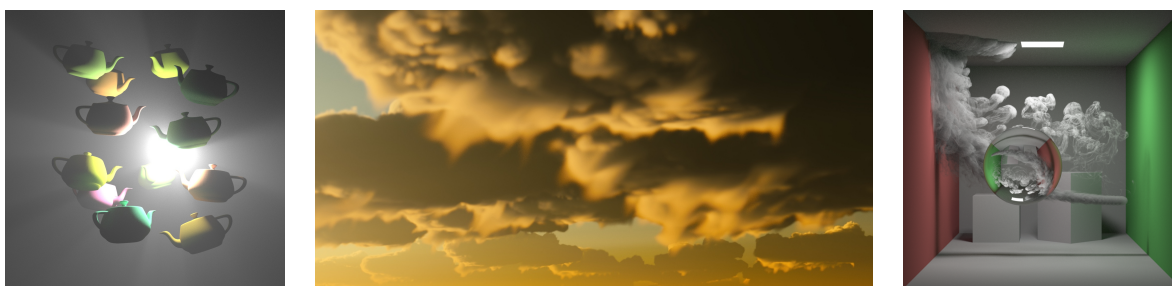


Figure 1: From left to right: soft shadows from a spherical light source (4 min), multiple scattering in clouds at sunset (2 hr 20 min), and a smoke simulation visible through refraction with one bounce of global illumination (10 min).

Abstract

We introduce a set of robust importance sampling techniques which allow efficient calculation of direct and indirect lighting from arbitrary light sources in both homogeneous and heterogeneous media. We show how to distribute samples along a ray proportionally to the incoming radiance for point and area lights. In heterogeneous media, we decouple ray marching from light calculations by computing a representation of the transmittance function that can be quickly evaluated during sampling, at the cost of a small amount of bias. This representation also allows the calculation of another probability density function which can direct samples to regions most likely to scatter light. These techniques are orthogonal and can be combined via multiple importance sampling to further reduce variance. Our method has very modest per-ray memory requirements and does not require any preprocessing, making it simple to integrate into production ray tracing based renderers.

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

1. Introduction

Rendering of participating media is critical to the realistic depiction of many natural effects including fog, steam, smoke and fire. These effects are particularly important in visual effects and animation where volumetric lighting is used to enhance the mood of the scene, and where smoke and explosions are common occurrences.

While the formal equations behind volumetric light transport have been known for quite some time [Cha50], the nested

integral structure of the equation remains computationally demanding. The recent survey by Raab et al. [RSK06] gives an overview of the best-known methods for importance sampling within homogeneous and heterogeneous media, allowing the incorporation into well-known light transport algorithms such as unidirectional and bidirectional path tracing, or Metropolis light transport. Our paper will illustrate alternative importance sampling strategies that can yield more robust estimators.

We will first highlight some cases where classical sampling methods can fail, then show how to distribute samples proportionally to incoming radiance along the ray in a way that avoids the weak singularity from the geometric term. We will then show how to apply the same sampling techniques to heterogeneous media by decoupling ray marching from light sample selection. The decoupling introduces a small amount of bias related to the step size. Finally, we will show how this decoupled representation of transmittance can be used to further reduce variance by combining a secondary probability density function constructed during ray marching via multiple importance sampling [VG95]. Higher-order scattering (from both the volume and other surfaces in the scene) will also be considered by the straightforward application of these techniques to path and light tracing.

2. Previous Work

Volume Rendering Equation: The volume rendering equation (first derived by Chandrasekhar [Cha50]) gives a formula for the radiance L seen at a position \mathbf{x} when looking along a direction $\vec{\mathbf{w}}$ as the sum of two terms:

$$L(\mathbf{x}, \vec{\mathbf{w}}) = T(s) L_s(\mathbf{x}_s, \vec{\mathbf{w}}) + L_v(\mathbf{x}, \vec{\mathbf{w}}) \quad (1)$$

The surface radiance L_s defined by the familiar rendering equation [Kaj86] is attenuated by transmittance T along the ray up to the distance s to the nearest surface intersection. The transmittance function is defined by the integral of the extinction coefficient σ_t along the ray:

$$T(t) = \exp\left(-\int_0^t \sigma_t(\mathbf{x}_w) dw\right) \quad (2)$$

To simplify the notation we write T as a univariate function, but it implicitly depends on the parameterized position $\mathbf{x}_t = \mathbf{x} + t\vec{\mathbf{w}}$ for any distance $t > 0$. Finally, light scattered inside the volume is accounted for by the volumetric radiance L_v :

$$L_v(\mathbf{x}, \vec{\mathbf{w}}) = \int_0^s \sigma_s(\mathbf{x}_t) T(t) \int_{\Omega_{4\pi}} \rho(\vec{\mathbf{w}}, \vec{\mathbf{v}}) L(\mathbf{x}_t, \vec{\mathbf{v}}) d\vec{\mathbf{v}} dt \quad (3)$$

where σ_s is the medium's scattering coefficient and ρ is the normalized phase function. The key difficulty in solving Equation 3 efficiently lies in the fact that the integration along the line contains two nested integrals: one for evaluating transmittance, the other for evaluating incoming light.

Brute force method: One of the first strategies introduced to deal with this difficulty was the classical ray marching algorithm [PH89]. By advancing along the ray in fixed sized steps, both the transmission and lighting terms can be evaluated in lockstep. Unfortunately, since the lighting calculation must recursively perform ray marching to compute shadowing, the brute force algorithm takes $O(N^2)$ steps on average, where N is the average number of steps along a ray. This fact was already recognized in the original publication which recommended caching the lighting as a preprocess so that it could be queried in constant time during rendering. Deep

shadow maps [LV00] later became a popular data structure for this task. The simplicity of this algorithm means it is still to this day employed in many production volume rendering systems [WBZC*10, WBZH*11]. However, caching lighting in this manner increases memory usage, blurs lighting details and complicates the support of area lights. Our aim is to emulate the success of this simple algorithm, and reduce its cost to linear time without relying on any preprocessing steps or sacrificing visual quality.

Stochastic methods: A stochastic alternative to brute force ray marching was first described in the context of bidirectional path tracing [LW96] and later Metropolis light transport [PKK00]. This technique is commonly referred to as *distance sampling* because it randomly selects a distance along the ray at which to evaluate light scattering. Distance sampling is trivial in homogeneous media as Equation 2 is analytically evaluated and inverted in closed-form. The heterogeneous case, however, cannot be solved in closed-form and requires some form of ray marching. Assuming it takes $O(N)$ steps to choose a point along the ray, and then $O(N)$ steps again to estimate shadowing to the light source, the algorithm is now linear in the number of steps, but has traded this efficiency for extra noise in the result. Indeed, the calculation of transmittance and lighting are still *coupled* as it is not possible to increase the quality of one without the other.

More recent work [MBJ*06, RSK06] has proposed the use of the unbiased *Woodcock tracking* algorithm from particle physics [Col68]. The algorithm relies on knowing an upper bound on the extinction coefficient $\sigma_t(\mathbf{x})$ along the ray. This is problematic for shader driven rendering systems where this value is not known in advance and may be difficult to estimate. Moreover, the performance is heavily dependent on the accuracy of this upper bound.

Acceleration structures such as kd-trees [YIC*10] or uniform grids [SKTM11] have been proposed to determine a locally accurate upper bound per ray, allowing optically thin regions to take larger (and thus fewer) steps. Unfortunately, these data-structures have not yet been extended to support motion blurred media, and are more difficult to build robustly for procedurally defined volumes. One possibility is to use interval arithmetic which may be overly conservative and complicate the implementation of the shading system.

Caching methods: Other light transport strategies have also been extended to support participating media. The classic photon mapping algorithm [JC98] for example, which was improved by the use of photon beams by Jarosz et al. [JNSJ11] takes a density estimation approach which is able to handle complex light transport paths, such as specular-diffuse-specular caustics. The beam formulation dramatically reduces the memory requirements of the original algorithm as beams implicitly represent infinite collections of photons, but complicates the handling of heterogeneous media because a representation of the transmittance must be stored along

each beam (increasing memory usage). Progressive photon beams [JNT*11] introduced a variation of Woodcock tracking for this purpose, but inherits the associated difficulties mentioned above. Several details important to production rendering also remain unexplored such as the combination with motion blur and the performance on scenes in which the camera only sees a small subset of the illuminated volume (such as the sunset in Figure 1). Nonetheless, these methods are still unmatched in their ability to handle complex light transport paths efficiently. Our work instead focuses on improving the quality of lower order lighting effects (single-scattering and first-bounce multiple scattering) without the assistance of complex data-structures.

Radiance caching [JDZJ08] is another method which uses a cached representation of the illumination to accelerate rendering of participating media. Similarly to photon mapping, the cached representation consumes extra memory and complicates the implementation and quality controls to be exposed to the user. While the method produces excellent results when tuned by experts, artifacts may be difficult to diagnose for more casual users and appear as flickering in animation which is usually more objectionable than noise. Motion blurred media and light sources have also not been explored within this framework. However, since radiance caching relies on Monte Carlo methods to seed the cache, it may be possible to apply our results within such a setting.

Real-time methods: Real-time approaches to volume scattering are usually restricted to single scattering from point light sources. Analytical methods [SRNN05, PSS11] can solve the rendering equation exactly, but must ignore visibility and thus cannot produce volume shadows. Real-time shadowing approaches [CBDJ11, Wym11] on the other hand are based on shadow mapping or similar discretizations of the visibility function. This is problematic for production rendering as these methods do not generalize to area lights, heterogeneous media or multiple scattering in any straightforward way. So while these approaches achieve very high performance, their applicability is limited to very specific scenarios. Our work focuses on general-purpose techniques that impose as few restrictions as possible on the supported types of lighting and media.

3. Sampling using incoming radiance

As the general form of Equation 3 contains non-analytical integrals, we first consider the simpler problem of single scattering from point lights in a homogeneous medium. We will then generalize our method to other light source types in Section 3.2, and to heterogeneous media in Section 4.

3.1. Point and Sphere Lights

Rewriting Equation 3 for single scattering from a point light in a homogeneous medium yields:

$$L(x, \vec{\omega}) = \int_a^b \sigma_s e^{-\sigma_t(t+\Delta+\sqrt{D^2+t^2})} \frac{\Phi}{D^2+t^2} dt \quad (4)$$

Figure 2 describes the involved parameters. Note that to simplify the notation, we omit the inclusion of visibility and phase function in the equation. We also reparameterize t so the origin is at the orthogonal projection of the light onto the ray. This change modifies the integration bounds a and b (which can be negative) and adds an extra term Δ which is the distance between the real ray origin and the new one. The use of integration bounds a and b is convenient when dealing with spotlights as we can restrict integration only to the part of the ray that intersects the light's cone.

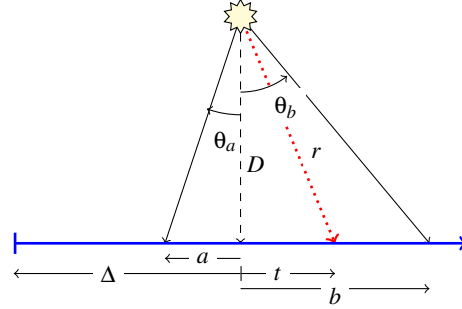


Figure 2: Single scattering from a point light source

While the transmission term is bounded by 1, the incoming radiance Φ/r^2 presents a weak singularity that starts to dominate as rays approach the light source. This reveals why distance sampling (distributing samples proportionally to $e^{-t\sigma_t}$) can be suboptimal. Distance sampling focuses most samples close to the viewer, missing the visually important light source (see Figure 3). In fact, the results for distance sampling can be made almost arbitrarily bad by increasing σ_t , pushing the sample points closer to the viewer and away from the light. As noted by Jarosz et al. [JNT*11], this problem persists in bidirectional methods when both light and viewer are far from a visually important point.

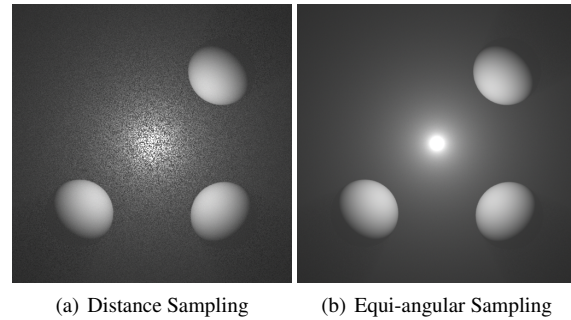


Figure 3: Point light source in homogeneous media (16 paths/pixel, each method rendered in 2 seconds)

It should be noted that photon beam density estimation methods [JNSJ11] do not suffer from this singularity as they connect light and eye paths by spatial proximity instead (at the cost of a small amount of bias in the result).

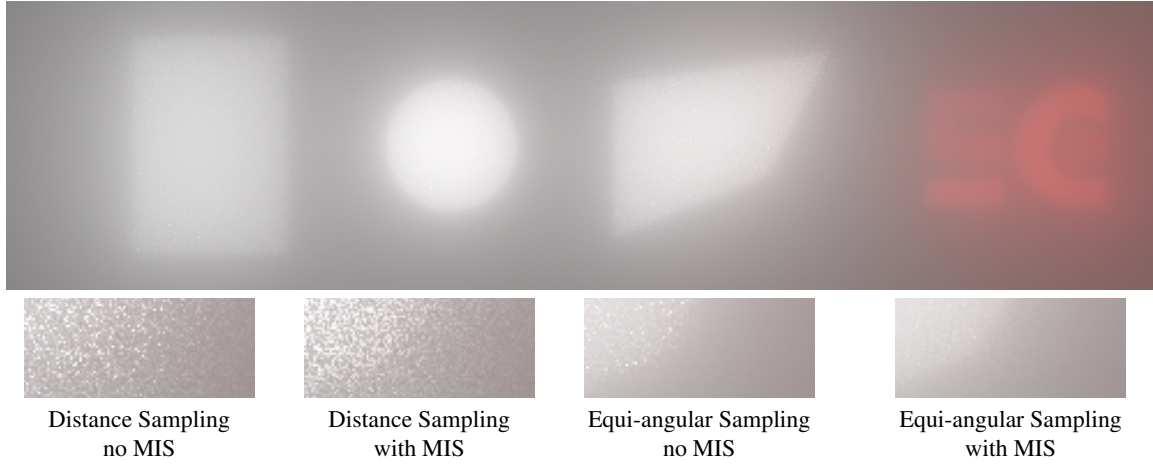


Figure 4: Single scattering from various area light shapes in homogeneous media (256 paths/pixel, ~2 min). Connection to the light can either be done by area sampling only (no MIS), or by sampling the phase function as well (with MIS). Equi-angular sampling makes MIS more effective as high variance is concentrated near the surface of the light.

We propose to design a pdf proportional to the $1/r^2$ term instead to focus more samples where the incoming light is strong and cancel out the weak singularity directly. This results in a Cauchy distribution which has the following normalized pdf and associated sampling function over the interval $[a, b]$ (ξ is a random number in $[0, 1]$):

$$\text{pdf}(t) = \frac{D}{(\theta_b - \theta_a)(D^2 + t^2)} \quad (5)$$

$$t(\xi) = D \tan\left((1 - \xi)\theta_a + \xi\theta_b\right) \quad (6)$$

$$\theta_x = \tan^{-1} x/D \quad (7)$$

Equation 6 reveals that this technique linearly interpolates the angle between the bounds of integration, so we refer to this technique as *equi-angular sampling*. The results in Figure 3(b) show the dramatic improvement compared to distance sampling. Due to the smoothness of the transmittance term, the noise present using our method quickly vanishes and yields nearly noise free images with as few as 16 paths per-pixel.

We can generalize this result to spherical light sources by noticing that the solid angle subtended by the sphere also varies as $1/r^2$. In fact, no change is required to Equations 5 and 6 to choose a sample along the ray. As long as the shadow ray traced to the light source is chosen proportionally to the solid angle of the sphere [SW91], the probabilities will cancel out and maintain the technique's efficiency. The image on the left of Figure 1 shows an example of soft shadows cast by a spherical light source.

3.2. Arbitrary Area Lights

For other types of area lights (quads, discs, triangles, etc.), Equation 4 gains a nested integral over the surface of the light. Most notably, the distance D which was constant before

is now a function of the position sampled from the light. Therefore, there is little hope of finding an analytical formula to importance sample all three dimensions of the problem at once in the general case.

We reduce the dimensionality of the problem by first choosing a sample point on the surface of the light (uniformly or according to an importance table for textured lights), and then connecting to the ray using equi-angular sampling. This can be viewed as a connection between a zero-length light path and the eye path. This approach eliminates the weak singularity in t as in the previous section, but there is now a second possible singularity due to the fact that D is no longer constant. This leaves some noise near the surface of the light which can be masked by applying multiple importance sampling [VG95] between the area light surface and the phase function. This is similar to the use of MIS for solving the final gather problem. Figure 4 shows the result for a variety of light shapes and shows the improvement over distance sampling, even when it is combined with MIS. Just as in the point light case, distance sampling can perform arbitrarily badly for these cases as optical depth increases.

Collections of multiple area light sources can either be handled by randomly choosing a point on the combined area of all light sources, or by handling them individually. We tend to prefer handling each light individually as this allows tuning the integration bounds $[a, b]$ along the ray for each light.

3.3. Arbitrary phase functions

The equi-angular technique ignores the phase function in distributing samples and therefore works best when the phase function is isotropic. Anisotropic phase functions reduce the effectiveness of the sampling by flattening the shape of the weak singularity. While this small loss in effectiveness could

be ignored at moderate anisotropies (for example ($|g| < 0.5$), we have found it is beneficial to combine Equation 5 with classic distance sampling by multiple importance sampling. For consistency with Equations 5 and 6, we give the equations of distance sampling for $t \in [a, b]$:

$$\text{pdf}(t) = \frac{\sigma_t}{e^{(t-a)\sigma_t} - e^{(t-b)\sigma_t}} \quad (8)$$

$$t(\xi) = a - \frac{1}{\sigma_t} \log \left(1 - \xi(1 - e^{-(b-a)\sigma_t}) \right) \quad (9)$$

This brings the total number of sampling techniques up to four: two for choosing a point along the ray and two methods for integrating the light contribution from the chosen point. Note that we always choose the point along the ray before choosing a direction towards the light, and that the probabilities simply multiply each other as the two choices are independent (even when we correlate the equi-angular center with the light sample as mentioned in Section 3.2). The MIS weights are simply scaling factors for the respective probabilities and multiply each other as well. In our implementation we found that the power heuristic works well, and generate one candidate sample for each of the four methods. In practice, several will end up with a numerically zero weight due to their low probability and therefore do not require a full visibility calculation (for instance the direction sampled from the phase function does not always hit the light source). As we will show in Section 4.2, this combination of line-sampling techniques is useful in heterogeneous media as well.

4. Sampling in Heterogeneous Media

4.1. Decoupled Ray Marching

At first glance, the techniques described in the previous section seem to only be applicable in a homogeneous medium. Indeed, we rely on being able to quickly estimate the transmittance at arbitrary points along the ray by an analytical formula. We observe that we can still use our method when the homogeneous volume is confined to a region of space. Only the distances inside the exponential and the bounds of integration must be changed to reflect the intersection distances with the volume. Figure 5 shows an example of restricting an infinite homogeneous volume to a sphere.

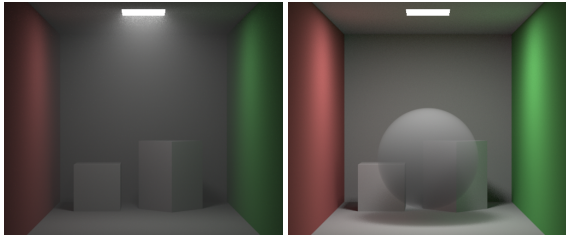


Figure 5: Locally confined homogeneous media are also renderable by the methods of Section 3.

This logic can be generalized to any number of homogeneous segments. We represent the transmission function $T(t)$

by an array of short segments which we assume to be homogeneous. Classical ray marching [PH89] can be interpreted as producing such an array, although previous formulations would not have stored it explicitly. For a given segment i we store volume properties $\sigma_{s,i}$ and $\sigma_{t,i}$, as well as the distance t_i to the start of the segment and its length Δ_i . The volume properties are obtained by executing a shader at a jittered position within the interval $[t_i, t_i + \Delta_i]$. We also compute the transmittance up to the front of each segment as we proceed from front-to-back:

$$T_0 = 1$$

$$T_i = T(t_i) = T_{i-1} \exp(-\Delta_{i-1} \sigma_{t,i-1})$$

This discretized representation of the volume along the ray makes it simple to evaluate $T(t)$ at *any* distance by using a binary search over the distances t_i to locate the segment containing the requested point. From the located segment k , the final expression of transmittance is simply:

$$T(t) = T_k \exp(-\min(t - t_k, \Delta_k) \sigma_{t,k}) \quad (10)$$

Figure 6 summarizes this process.

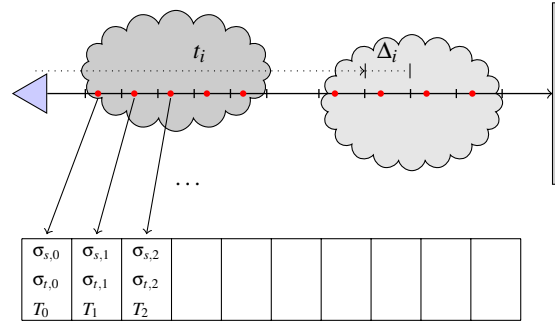


Figure 6: Volume properties are stored in an array prior to lighting calculations so $T(t)$ may be evaluated quickly. The segments do not need to be of equal length or contiguous.

Once we have built this small per ray data-structure, we can trace any number of shadow rays without having to restart the ray marching process. Each lighting call will need to perform a $O(\log(N))$ binary search to evaluate $T(t)$ but this cost is far lower than the $O(N)$ steps taken to evaluate transmittance along the shadow ray and is not a noticeable bottleneck, nor does it affect the time complexity of our algorithm. In most cases, N ranges from 10-100, while we only need to take 1-4 light samples.

In difference to the classic algorithm [PH89], we do not perform lighting calculations during ray marching, avoiding the $O(N^2)$ time complexity. In difference to stochastic distance sampling methods [LW96, YIC*10], we do not stop ray marching at a randomly chosen location but instead keep tracing all the way across the ray before tracing any shadow rays. We maintain a linear time complexity in the number of steps, but are free to take lighting samples at any location along the ray (in particular using Equation 6). The quality

of the estimate of the transmittance is also decoupled from the lighting as it is possible to increase the number of steps (thus resolving smaller details) without tracing any additional shadow rays.

We refer to this technique as *decoupled* ray marching because it separates the calculation of the transmission from the lighting. The combination of decoupled ray marching and the light sampling routines from Section 3 let us render single scattering in heterogeneous volumes as shown in Figure 7.

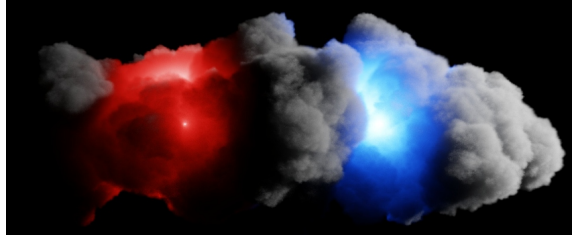


Figure 7: Heterogeneous medium rendered using decoupled ray marching (16 paths/pixel, ~5 min). Previous stochastic methods would have a hard time capturing single scattering from the light sources embedded in the medium as seen in Figure 3.

This decoupling does introduce some bias, however it is explicitly controlled by the step size parameter and can be made arbitrarily small. In fact, as the step size approaches the Nyquist rate of the volume, our algorithm converges on the ground truth result. In our implementation each volume primitive controls how many ray marching samples are required within the bounds it occupies along the ray to allow for volumes of different frequency contents to be mixed in the same scene. We also point out that the homogeneous limit case is handled without bias as a single ray marching sample captures the volume properties exactly.

4.2. Discrete Density PDF

Distributing samples according to the light intensity is not always optimal when the heterogeneous medium is thick. In Figure 8 we see that there is some noise at the “surface” of the volume because the sampling is not aware of the sharp change in $T(t)$ as we enter the volume. As a consequence, samples may be distributed deep inside the volume where their contributions are low. Another example would be the space between the volumes in Figure 6 which could receive light samples even though it contains no medium.

Decoupled ray marching lets us improve this situation by computing a second pdf at little additional cost. During ray marching, we simply assign each segment a discrete probability $p_i = \Delta_i \sigma_{s,i} T_i$. This approximates the term in front of the integral in Equation 3, automatically skips over regions that do not scatter light and assigns fewer samples deep inside the volume. After ray marching, this piecewise constant pdf can be turned into a cdf in linear time. Samples can then be taken

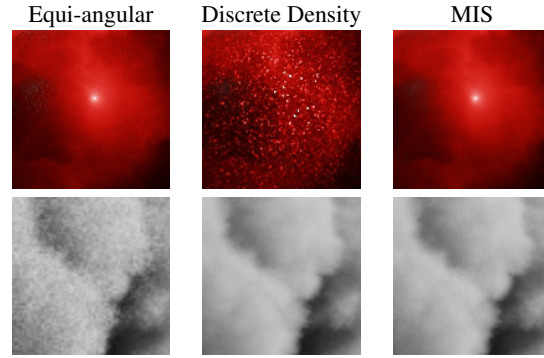


Figure 8: Closeups of Figure 7. MIS combines the strengths of each of the line-sampling techniques.

according to this pdf by using a binary search for cdf inversion. The choice of the position within the segment $[t_i, t_i + \Delta_i]$ can be made by appropriately normalized distance sampling (Equations 8 and 9). Without having changed the time complexity of our algorithm, we can now generate samples that focus calculations on visually important regions of the ray. Figure 8 shows the resulting improvement. Unfortunately the noise near the origin of the embedded lights is worse compared to equi-angular sampling (as would be the case with previous methods).

As we now have two line-sampling techniques with differing strengths, it is natural to combine them by multiple importance sampling. As shown in Figure 8, this gives a clean result in all cases without having to increase the number of samples. As discussed in Section 3.3 this combination of line-sampling methods also helps with highly anisotropic phase functions that reduce the effectiveness of equi-angular sampling. In homogeneous media, where we only need a single segment for our discrete pdf, it becomes equivalent to classical distance sampling.

Our density based pdf has a number of advantages over previous techniques. We only generate samples in regions where $\sigma_s > 0$, focusing computation on regions that actually scatter light. Earlier methods could only generate samples where $\sigma_t > 0$ because they implicitly sample a pdf proportional to T alone (which is stochastically estimated at the same time). Our pdf can also be queried for arbitrary distances t , enabling the use of MIS. Previous methods were based on rejection sampling and therefore only allowed the *generation* of samples according to a pdf that is never explicitly calculated. This precludes previous work from being combined with equi-angular sampling.

5. Multiple Scattering

The two previous sections focused on single scattering, however decoupled ray marching easily handles higher order bounces by plain path tracing. The discrete density pdf computed for direct lighting can be reused to generate sampling

locations for indirectly scattered light. Figure 9 shows an example of single bounce global illumination rendered this way. As our technique is easily integrated into a ray tracer, we automatically capture the indirect illumination from nearby surfaces as well as internal volume scattering. Emissive volumes such as fire are easily accounted for during ray marching, allowing them to cast illumination onto nearby geometry.



Figure 9: Indirect lighting on a procedural volume and a fire simulation rendered by path tracing (36 and 256 paths/pixel respectively, ~15 min and ~2 hrs)

Equi-angular sampling is also helpful in light tracing (the adjoint of path tracing) as it allows scattered rays to be efficiently connected back to the camera. The geometric connection term there also varies as $1/r^2$, so equi-angular sampling can noticeably reduce variance (see Figure 10). The bounds a and b in Equations 5 and 6 can be set by clipping the ray against the camera’s frustum to focus calculations on rays that map to screen pixels. Reflections and refractions are missing from the image because we used pure light tracing and a pinhole camera. Attempts at equal quality comparisons revealed that even with 8 times more samples, distance sampling still contains a noticeable amount of spiked noise, which is avoided by equi-angular sampling.

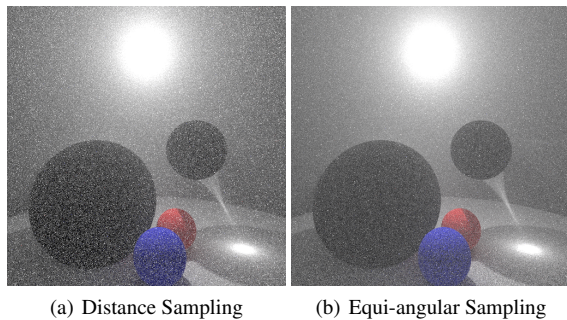


Figure 10: Variance of light tracing paths can be reduced by improved sampling of explicit camera connections (64 paths/pixel, ~12 seconds)

In scenes which require a high number of volumetric bounces such as clouds, simple path tracing may still be too computationally demanding. The method of Bouthors

et al. [BNM*08, WBZH*11] can be easily adapted to our framework as it approximates the effect of light bleeding through the volume by simply changing the volume properties along shadow rays. We can combine this approach with one real bounce to better approximate light diffusion through the medium (see Figure 11). The image of clouds at sunset in Figure 1 was also rendered using this method. The orange color shift is due to absorption through a procedural representation of the atmosphere. A single light source simulates the sun and all indirect lighting is due to approximate multiple scattering. This particular case would be challenging for bidirectional methods [LW96, JNSJ11] because of the large spatial extent of the scene. The atmosphere is modeled as an earth-sized sphere so the distant light representing the sun would cast many rays that do not contribute to the image.

The fire example in Figure 9 is another tricky case for bidirectional methods as it is difficult to emit rays from the fire itself. These examples took longer to render than our other examples as higher sampling rates were required to fully smooth out the indirect lighting by brute force path tracing. More sophisticated Markov Chain Monte Carlo methods such as Metropolis Light Transport [PKK00] could be combined with our importance sampling scheme to more effectively search the path space in these hard to sample cases. While we have not explored this combination in our renderer, we do not believe it would pose any fundamental difficulties.

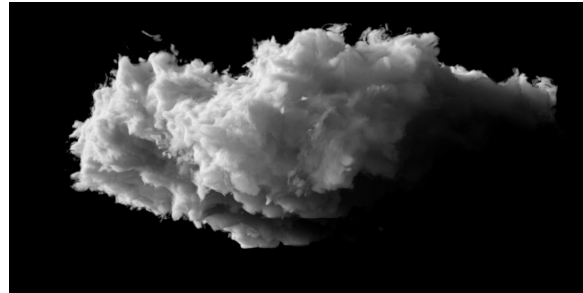


Figure 11: High orders of multiple scattering can be approximated by manipulating medium properties along shadow rays after the first bounce (64 paths/pixel, 4 min)

6. Implementation Details and Results

Our work is implemented in the Arnold renderer, a Monte Carlo path tracer for production rendering [KFC*10]. Our implementation augments the traditional “first hit” record common to any ray tracer with a linked list of segments describing intersections with volume primitives. Volume primitives are part of the scene hierarchy along with surfaces, but report *interval* hits instead of *surface* hits. After all hits are gathered, we sort and decompose the intervals into non-overlapping segments. Each segment is further divided according to a per-primitive step size. We compute the number of potential steps along the ray ahead of time so that we may pre-allocate the memory for the array required by decoupled ray marching (Section 4.1).

The creation of these temporary linked lists and arrays for every ray is managed by a per-thread memory pool with a stack interface. When we enter the volume integration routine, we save the state of the memory pool by pushing it on a stack so that we can quickly reclaim the memory as soon as the calculation is done. Note that only rays performing light integration need to perform any allocations. Shadow rays simply evaluate Equation 2 by ray marching. This means our memory pools rarely grow to more than a few kilobytes in size.

Many other opportunities for optimization exist within our framework, following commonly used tricks of other production systems [WBZH*11]. For example, ray marching may be terminated as soon as we reach a low enough transmission. This can be done relative to an upper bound on light intensity, or stochastically, to avoid introducing extra bias. As noted above, homogeneous regions only need to take a single step as long as their extents can be bounded. A common volume primitive is a grid of density values [WBZC*10], which is trivially preprocessed to identify empty blocks at loading time. By using a 3D-DDA traversal on these blocks, we reduce the length of volume intervals and thus the number of ray marching steps that need to be taken. For sparse data such as the smoke simulation on the right of Figure 1 the speedup can be significant.

Since our method is purely sampling based, it is trivial to support effects such as motion blur without any special handling (see Figure 12). Previous methods based on caching such as deep shadow maps [LV00] would incorrectly blur lighting detail in such cases. To efficiently bound density grids advected by velocity fields (as written by most fluid simulators) we expand the bounding box of non-zero blocks by the length of the longest velocity vector. As most smoke simulations do not fill their entire domain, this still allows large amounts of empty space to be skipped.

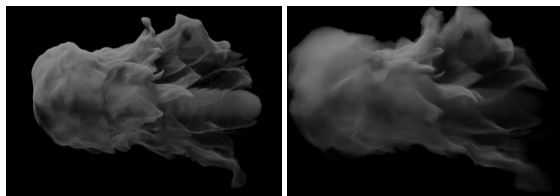


Figure 12: A smoke simulation rendered both without and with Eulerian motion blur (16 paths/pixel, 1 min and 3 min respectively)

Rendering times quoted in the figures in this paper are from a dual quad-core Intel Xeon E5420 2.5Ghz workstation. Image resolutions were kept around 1 million pixels for all examples. Final quality frames often render within a few minutes, while noisy previews can be obtained in a few seconds. When combined with a progressive sampling of the image plane, artists can see the render improve over time and interactively adjust camera, light and volume shader parameters.

This is a key advantage for film production rendering which requires both fast artist feedback and very high visual fidelity.

While there are still cases involving indirect lighting and frame-filling volume effects that remain challenging and can take multiple hours, noisy previews can still be obtained rapidly enough to make lighting decisions before committing to final quality renders. The simplicity of the quality controls (step size and number of samples) means that there is less room for error when increased quality is needed.

Finally, we include some rendered frames from the film *Men in Black 3* which demonstrate our method scales up to very large data sets (see Figure 13). The geometry of the rocket and environment total over 75 million triangles, while the fluid simulations are stored in voxel grids between 256^3 and 512^3 in resolution, totalling around 3GB per frame. Rendering times for steam elements ranged from 2 to 4 hours on a single Intel CPU core.

7. Limitations and Future Work

While the presented importance sampling techniques are very effective at rendering a wide variety of common cases, they do not form a complete light transport solution. For example, while we can achieve noise free results for single scattering in the scenes of Figures 3(b) and 7, calculating an extra bounce would cause a large amount of noise because we would not be able to detect the hotspot of scattered lighting surrounding the light source. This is not a weakness in our methods but a general limitation of unidirectional path tracing. We believe that our estimators will be helpful in the context of bidirectional path tracing as equi-angular sampling will allow for connections from path *segments* to path vertices instead of only between path vertices. Since connecting paths always presents a $1/r^2$ geometric term, we believe that convergence could be improved. As the number of possible connection techniques increases, care must be taken to ensure MIS weights between all methods can still be computed efficiently.

One weakness of our work over previous stochastic distance sampling routines is that we must traverse the entire ray instead of randomly stopping in the middle. While this is what allows the transmission calculation to be decoupled from the lighting, it also means that we cannot scale to high numbers of bounces and must resort to approximate techniques instead. It would be interesting to see if randomized termination heuristics could be applied without sacrificing the advantages of decoupled ray marching.

As with previous ray marching techniques (except Woodcock tracking), the bias induced by decoupled ray marching can be a weakness. If the user chooses the step size too large, the volume loses density. If the step size is too small, render times can increase substantially. While we do not expect to find a fully unbiased solution, we would like to explore adaptive ray marching strategies that vary the step size based on

volume contents to spend more time in sampling regions of higher variance.

The ideas behind decoupled ray marching can be extended beyond participating media. We have had some success in applying similar logic to discrete random media such as hair. By first tracing through all hairs and creating a pdf based on the relative transmitted visibility, we can stochastically decide which to shade, trading fewer lighting calculations for slightly increased noise.

Finally, we would like to explore implementing decoupled ray marching on the GPU. While the memory requirements of our method are very modest for a CPU based implementation, this memory may not be easy to keep in the local store of current GPUs since they run many more concurrent threads of execution. Choosing the optimal memory layout for this data, and dealing with the uneven workload caused by rays of different lengths will be the main challenge.

8. Conclusion

We have presented new importance sampling techniques for efficient rendering of participating media. Our approach retains most of the advantages of Monte Carlo based methods in that it does not require any form of pre-computation nor extra illumination storage. While our method cannot surpass state of the art special purpose volume renderers [WBZH*11] in speed, we come close enough that the improved visual quality of accurate ray traced lighting, combined with the benefit of having volume rendering fully integrated within a general purpose renderer has been a major productivity boost for our users and raised the bar of the imagery they are able to create.

Acknowledgements

We would like to thank Alejandro Conty, Alan King, Thiago Ize, Clifford Stein, and Magnus Wrenninge for helpful comments in reviewing early drafts of this work. We would also like to thank the anonymous reviewers for their suggestions.

References

- [BNM*08] BOUTHORS A., NEYRET F., MAX N., BRUNETON E., CRASSIN C.: Interactive multiple anisotropic scattering in clouds. *I3D* 2008, pp. 173–182. [7](#)
- [CBDJ11] CHEN J., BARAN I., DURAND F., JAROSZ W.: Real-time volumetric shadows using 1d min-max mipmaps. *I3D* 2011, pp. 39–46. [3](#)
- [Cha50] CHANDRASEKHAR S.: *Radiative Transfer*. Clarendon Press, Oxford, 1950. [1, 2](#)
- [Col68] COLEMAN W.: Mathematical verification of a certain Monte Carlo sampling technique and applications of the technique to radiation transport problems. *Nucl. Sci. Engin.* (1968), 76–81. [2](#)
- [JC98] JENSEN H. W., CHRISTENSEN P. H.: Efficient simulation of light transport in scences with participating media using photon maps. *ACM SIGGRAPH* 1998, pp. 311–320. [2](#)
- [JDZJ08] JAROSZ W., DONNER C., ZWICKER M., JENSEN H. W.: Radiance caching for participating media. *ACM Transactions on Graphics* 2008, pp. 7:1–7:11. [3](#)
- [JNSJ11] JAROSZ W., NOWROUZEZAHRAI D., SADEGHI I., JENSEN H. W.: A comprehensive theory of volumetric radiance estimation using photon points and beams. *ACM Transactions on Graphics* 2011, pp. 5:1–5:19. [2, 3, 7](#)
- [JNT*11] JAROSZ W., NOWROUZEZAHRAI D., THOMAS R., SLOAN P.-P., ZWICKER M.: Progressive photon beams. *ACM SIGGRAPH Asia* 2011, pp. 181:1–181:12. [2, 3](#)
- [Kaj86] KAJIYA J. T.: The rendering equation. *ACM SIGGRAPH* 1986, pp. 143–150. [2](#)
- [KFC*10] KRÍVÁNEK J., FAJARDO M., CHRISTENSEN P. H., TABELLION E., BUNNELL M., LARSSON D., KAPLAYAN A.: Global illumination across industries. *ACM SIGGRAPH* 2010 Courses. [7](#)
- [LV00] LOKOVIC T., VEACH E.: Deep shadow maps. *ACM SIGGRAPH* 2000, pp. 385–392. [2, 8](#)
- [LW96] LAFORTUNE E. P., WILLEMS Y. D.: Rendering participating media with bidirectional path tracing. *Eurographics Workshop on Rendering* 1996, pp. 91–100. [2, 5, 7](#)
- [MBJ*06] MORLEY R. K., BOULOS S., JOHNSON J., EDWARDS D., SHIRLEY P., ASHIKHMIN M., PREMÖZE S.: Image synthesis using adjoint photons. *Graphics Interface* 2006, pp. 179–186. [2](#)
- [PH89] PERLIN K., HOFFERT E. M.: Hypertexture. *ACM SIGGRAPH* 1989, pp. 253–262. [2, 5](#)
- [PKK00] PAULY M., KOLLIG T., KELLER A.: Metropolis light transport for participating media. *Rendering Techniques* 2000, pp. 11–22. [2, 7](#)
- [PSS11] PEGORARO V., SCHOTT M., SLUSALLEK P.: A Mathematical Framework for Efficient Closed-Form Single Scattering. In *Graphics Interface 2011* (2011), pp. 151–158. [3](#)
- [RSK06] RAAB M., SEIBERT D., KELLER A.: Unbiased global illumination with participating media. *Monte Carlo and Quasi-Monte Carlo Methods* 2006, pp. 591–606. [1, 2](#)
- [SKTM11] SZIRMAY-KALOS L., TÓTH B., MAGDICS M.: Free path sampling in high resolution inhomogeneous participating media. *Computer Graphics Forum* 30, 1 (2011), 85–97. [2](#)
- [SRNN05] SUN B., RAMAMOORTHY R., NARASIMHAN S. G., NAYAR S. K.: A practical analytic single scattering model for real time rendering. *ACM SIGGRAPH* 2005, pp. 1040–1049. [3](#)
- [SW91] SHIRLEY P., WANG C.: Direct lighting calculation by Monte Carlo integration. *Eurographics Workshop on Rendering* 1991, pp. 54–59. [4](#)
- [VG95] VEACH E., GUIBAS L. J.: Optimally combining sampling techniques for Monte Carlo rendering. *ACM SIGGRAPH* 1995, pp. 419–428. [2, 4](#)
- [WBZC*10] WRENNINGE M., BIN ZAFAR N., CLIFFORD J., GRAHAM G., PENNEY D., KONTKANEN J., TESSENDORF J., CLINTON A.: Volumetric methods in visual effects. *ACM SIGGRAPH* 2010 Courses. [2, 8](#)
- [WBZH*11] WRENNINGE M., BIN ZAFAR N., HARDING O., GRAHAM G., TESSENDORF J., GRANT V., CLINTON A., BOUTHORS A.: Production volume rendering 2: Systems. *ACM SIGGRAPH* 2011 Courses. [2, 7, 8, 9](#)
- [Wym11] WYMAN C.: Voxelized shadow volumes. *HPG* 2011, pp. 33–40. [3](#)
- [YIC*10] YUE Y., IWASAKI K., CHEN B.-Y., DOBASHI Y., NISHITA T.: Unbiased, adaptive stochastic sampling for rendering inhomogeneous participating media. *ACM SIGGRAPH Asia* 2010, pp. 177:1–177:8. [2, 5](#)

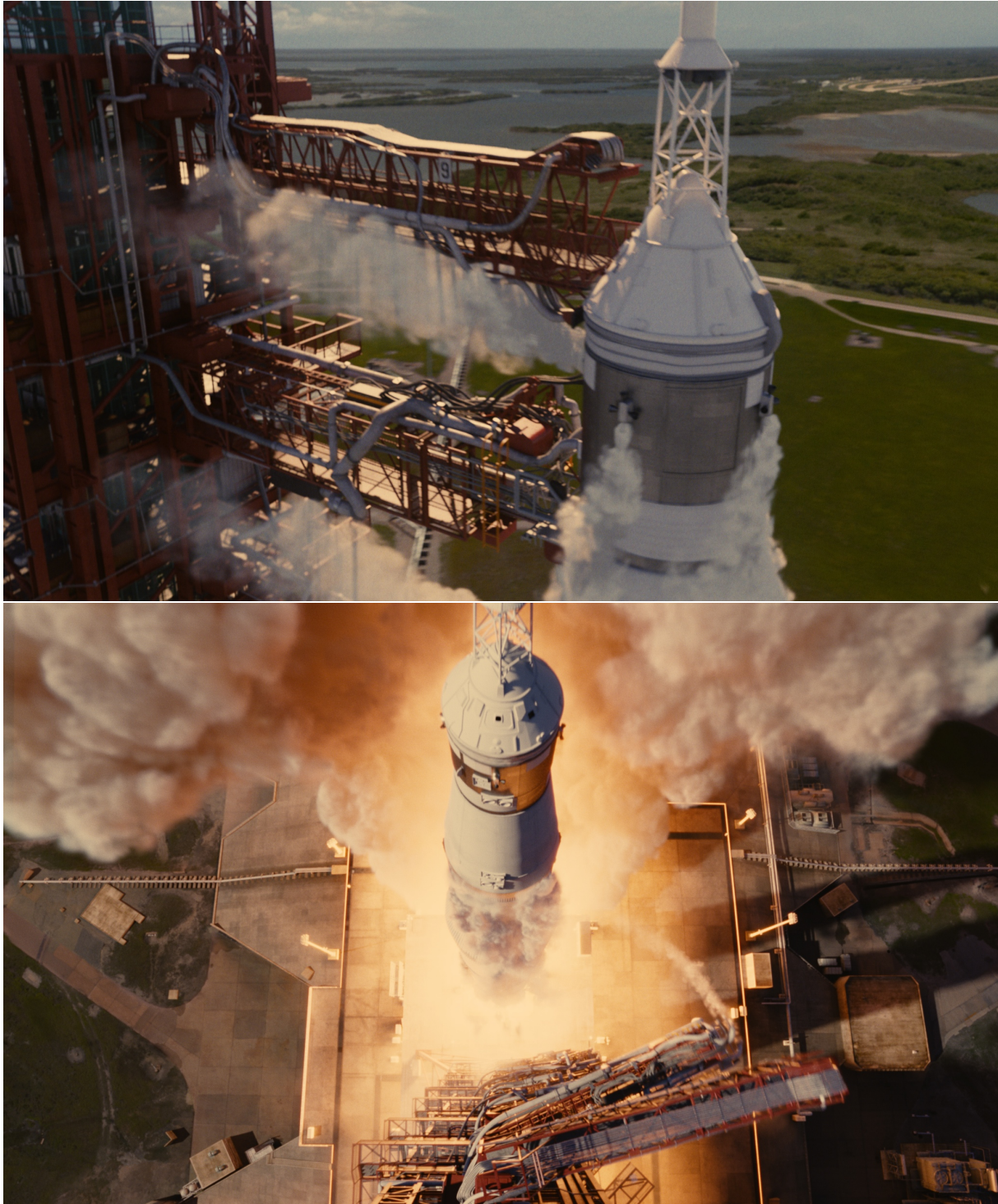


Figure 13: Computer generated stills of the Apollo 11 launch sequence from the film Men In Black 3. The sunlit geometry casts detailed shadows onto the smoke which would be difficult to capture with shadow maps due to the large spatial extent of the scene. Point lights embedded in the fluid simulation simulate the glow of the exhaust.
©2012 Columbia Pictures Industries, Inc. All Rights Reserved.