Improving the Dwivedi Sampling Scheme

Johannes Meng\(^1\)\(^†\) Johannes Hanika\(^1,2\)\(^‡\) Carsten Dachsbacher\(^1\)\(^§\)

\(^1\) Karlsruhe Institute of Technology \(^2\) Weta Digital Ltd.

Abstract
Despite recent advances in Monte Carlo rendering techniques, dense, high-albedo participating media such as wax or skin still remain a difficult problem. In such media, random walks tend to become very long, but may still lead to a large contribution to the image. The Dwivedi sampling scheme, which is based on zero variance random walks, biases the sampling probability distributions to exit the medium as quickly as possible. This can reduce variance considerably under the assumption of a locally homogeneous medium with constant phase function. Prior work uses the normal at the Point of Entry as the bias direction. We demonstrate that this technique can fail in common scenarios such as thin geometry with a strong backlight. We propose two new biasing strategies, Closest Point and Incident Illumination biasing, and show that these techniques can speed up convergence by up to an order of magnitude. Additionally, we propose a heuristic approach for combining biased and classical sampling techniques using Multiple Importance Sampling.

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

1. Motivation
Rendering photorealistic characters is important in the entertainment industry, for instance to create digital doubles for visual effects in movies. Nowadays this is often done using Monte Carlo (MC) integration, i.e. by running random walks to form particle trajectories connecting the sensor and the light sources. This paper will focus on simulating sub-surface scattering in materials such as skin or wax.

Many fast approximations to rendering sub-surface scattering are known, but they tend to show artifacts near concavities and thin features in the geometry. The reason for this is that most analytic methods build on the assumption of simplified geometry, such as semi-infinite slabs. Results will be sub-optimal, and sometimes visually displeasing, whenever this assumption is violated grossly.
We observe that in such cases other means of biasing the random walk can perform much better (see Fig. 1).

Classical Monte Carlo simulation does not suffer from this problem, but tends to be slow in participating media: uninformed random walks may scatter many times before exiting the medium. Yet, one cannot simply terminate random walks just because they are long, because they may contribute significantly to the image. This is especially true in high-albedo participating media.

To solve this problem, Krivánek and d’Eon [Kd14] applied the zero-variance random walk theory by Dwivedi [Dwi82a,Dwi82b] to rendering. In this method, geometry is also approximated by semi-infinite slabs to compute an analytic approximation of the light field. However, this light field is not visualized directly. Instead, it is merely used to bias the sampling probability distributions employed for the Monte Carlo random walk. In essence, Dwivedi sampling will prefer directions towards the slab surface, and will move in larger steps if the random walk progresses in the direction of the slab normal. Note that the resulting estimator is unbiased despite the double use of the word bias in literature.

Krivánek and d’Eon [Kd14] place the semi-infinite slab so that its normal is aligned with the surface normal at the Point of Entry, which is the point where the current random walk first entered the medium (cf. Fig. 2 left). This generally leads to much shorter random walks, and thus improves convergence considerably.

However, this strategy is not optimal in certain scenarios. The Point of Entry really is an approximation to the closest surface point, which can fail in the presence of thin geometric features (center) or high surface curvature. Additionally, if there is a strong back light, PoE sampling will drag the random walk away from the light source (right). The slabs used by our proposed techniques are shown in red: Closest Point (center) and Incident Illumination (right).

As our main contributions,
- we propose a heuristic approach for combining Classical, Closest Point, and Incident Illumination sampling using Multiple Importance Sampling [VG95].

2. Previous Work

2.1. Radiative Transfer

Radiation in participating media can be described using the Radiative Transfer Equation (RTE) [Cha60]

$$\frac{d}{ds} + \mu_t(x) L(x, \omega) = L^e(x, \omega) + \mu_s(x) \int_{4\pi} \phi(\omega', \omega) L(x, \omega') d\omega'. \quad (1)$$

In Eqn. (1), $L(x, \omega)$ is the radiance at position $x$ that flows into direction $\omega$, while $L^e(x, \omega)$ is the emitted radiance. $\mu_t(x)$ and $\mu_s(x)$ are the extinction and scattering coefficients, and $\phi(\omega', \omega)$ is the phase function. The adjoint RTE can be obtained from Eqn. (1) by replacing radiance $L$ with importance $\Psi$, and emission $L^e$ with the detector response function $\Psi'$. In general, these equations also depend on the wavelength and time interval, but these are omitted here for brevity.

2.2. Sub-surface Scattering for Skin

Rendering skin efficiently and accurately has received a fair amount of attention in literature. The dipole method [JMLH01] in combination with irradiance caches [JB02] was very popular for a long time, and scattering parameters have been derived from measurements [Dj05, Dj06]. Fast approximations in form of sums of Gaussians have been researched [dLE07]. Diffusion profiles on the surface have been refined from the theoretical view [dI11] as well as from the practical side [CB15]. Those have been refined by using directional dipole models [FHK14]. Production work flows have been simplified considerably by removing the irradiance caches from the pipeline and by moving to importance sampling of diffusion profiles in the ray tracing context [KJKCF13]. Such a ray tracing scheme, however, does not simulate the full transport inside the tissue but...
only samples a new exit point on the surface, under the assumption of locally flat geometry, leading to problems with high curvature and near concavities. For example, light bleeding may occur or diffusion profiles may be distorted. To avoid such problems, it is possible to use these closed form solutions to guide Monte Carlo sampling, which we will detail in the next section.

2.3. Dwivedi Sampling

In this section, we review the main concepts of Dwivedi sampling [Dwi82a, Dwi82b] and its foundation, zero variance random walk theory.

The concept of zero variance random walks has been explored by various authors, e.g. [Dwi82a, Dwi82b, Boo87], but a good overview is given in a work by Hoogenboom [Hoo08]. The idea is appealing: compute a solution to the RTE (1) with Monte Carlo random walks, but in such a way that the estimator has no variance.

This can be done by modifying, or biasing, the sampling distributions for position and direction in the random walk. Note that the word bias is overloaded in this context: even though sampling distributions are biased, the resulting estimator is still unbiased in the sense that it has the correct expected value. This is similar to importance sampling the free path for the time dimension in the context of transient rendering [JMMn14].

Dwivedi [Dwi82b] showed that one can construct zero variance random walks by sampling a new position $x' \parallel x$ and direction $\omega'$ proportionally to the product of phase function $\phi$, transmittance $\tau$, and importance $\Psi$: 

$$p(x', \omega' | x, \omega) \propto \phi(\omega, \omega') \tau(x, x') \Psi(x', \omega'),$$

where 

$$\tau(x, x') = \exp \left(- \int_x^{x'} \mu(x') \, dx' \right).$$

Unfortunately, this requires knowledge of the importance function for every position and direction. This is a far more difficult problem that determining the sensor response for a given sensor, and so true zero variance random walks are mainly of theoretical interest.

However, as Dwivedi points out [Dwi82b], they can still be used to reduce variance. The idea is that an approximate importance function can be used to reduce variance. Such an importance function can be computed analytically for slab geometry using Singular Eigenfunction Expansions.

2.3.1. Singular Eigenfunction Expansions

Singular Eigenfunction Expansions can be seen as representing the general solution to the RTE or the adjoint RTE in a convenient basis. This can be done if some assumptions are introduced.

Similar to diffusion techniques, the geometric setup is assumed to be a semi-infinite slab. As shown in Fig. 3, only the depth in the medium $z$ and the cosine of the angle between the ray direction and the surface normal, $\omega_z$, are considered in this type of geometry. This is usually justified by the assumption of uniform external emission. In the case of the adjoint formulation, one may interpret the slab surface at $z = 0$ as a uniform sensor.

Additionally, the medium is assumed to be homogeneous, the phase function is assumed to be isotropic and no emission inside the medium is considered. Combined, these assumptions lead to the following adjoint RTE [Cha60]:

$$\left[ \omega_z \frac{\partial}{\partial z} + \mu \right] \Psi(z, \omega_z) = \frac{\mu_z}{2} \int_{-1}^{1} \Psi(z, \omega_z') \, d\omega_z'.$$

It can be shown that all solutions to this RTE can be written as the linear combination

$$\Psi(z, \omega_z) = A(v_0) \phi(v_0, \omega_z) e^{-\mu v_0}$$

$$+ \int_{-1}^{1} A(v) \phi(v, \omega_z) e^{-\mu v \, v_0} \, dv$$

$$+ A(-v_0) \phi(-v_0, \omega_z) e^{-\mu v_0}$$

where $\Psi(z, \omega_z)$ is a linear combination of coefficients, $\phi(v, \omega_z)$ is called an eigenfunction, and the $v$ are eigenvalues with $v_0 > 1$. All $v = \pm v_0$ are the discrete eigenvalues, while $v \in [-1, 1]$ are called the singular eigenvalues.

This is the Singular Eigenfunction Expansion for $\Psi(z, \omega_z)$. We show a partial derivation in supplemental material, but the full derivation can be found in texts such as [CZ67, MK73].

To apply this solution to zero variance random walks, previous work [Dwi82a, Kd14] now prunes the expansion aggressively.

First, the importance in a slab without detectors must decrease as the distance to the surface goes to infinity. Consequently, coefficients of non-decreasing modes in Eqn. (5), where $v < 0$, must be 0.

Second, the singular part of the expansion decreases faster than the discrete part, since $v_0 > 1$ and the integration domain ends at 1. Thus, assuming $z$ is large, the solution will be dominated by the discrete part.

This leads to the final approximation

$$\Psi(z, \omega_z) \propto \frac{1}{v_0 - \omega_z} e^{-\mu z / v_0}.$$
and positions from the distributions
\[
p(\omega^t | z, \omega_c) \sim \frac{1}{\nu_0 - \omega_c} \quad (7)
\]
\[
p(\omega^t | z, \omega_c) \sim \tau(z, \omega^t) e^{-\mu^t / \nu_0} \quad (8)
\]
The normalized directional distribution is
\[
p(\omega^t | z, \omega_c) = \frac{1}{\log \left( \frac{\nu_0 + 1}{\nu_0 - 1} \right) \nu_0 - \omega_c} \quad (9)
\]
which can be sampled using \( \xi \sim \text{unif}(0, 1) \) as
\[
\omega_c = \nu_0 - (\nu_0 + 1) \left( \frac{\nu_0 - 1}{\nu_0 + 1} \right) \xi \quad (10)
\]
The free path sampling pdf Eqn. (8) can be simplified using the assumption that the medium is homogeneous, which means that the transmittance term is a simple exponential in \( t = \| \omega' - z \| \). Using \( \omega' = z - \omega_c t \), we get
\[
p(t | z, \omega_c) \propto e^{-\mu t} e^{-\mu (z - \omega_c t) / \nu_0} = e^{-(1 - \omega_c / \nu_0) \mu t} \quad (11)
\]
The normalized distribution is a simple exponential distribution:
\[
p(t | z, \omega_c) = (1 - \omega_c / \nu_0) \mu e^{-(1 - \omega_c / \nu_0) \mu t} \quad (12)
\]
This distribution can be sampled as usual. It effectively has a modified extinction coefficient,
\[
\mu^t (\omega_c) = (1 - \omega_c / \nu_0) \mu_c \quad (13)
\]
In Fig. 4, we illustrate that the directional distribution prefers directions towards the slab normal \( \omega_c = 1 \), while the free path distribution has a longer tail (and thus prefers longer paths) if the direction is oriented along the slab normal.

### 2.4. Slab Orientation

One of the assumptions used in Dwivedi sampling is that the geometry can be approximated by a semi-infinite slab. In realistic scenes, this assumption is usually violated in general, but it may hold locally. It is thus important to consider how exactly a slab can be placed into a scene for good performance. Since the positional variable is not present in the final sampling distributions Eqns. (9) and (12), this amounts to choosing how the slab is oriented, or what its normal is.

Křivánek and d’Eon [Kd14] use the surface normal at the Point of Entry (PoE) as the slab normal, i.e. the point where the random walk first enters the medium (cf. Fig. 2, left). The underlying assumption is that the normal at the PoE will usually point away from the medium. If a random walks is biased into this direction, it will probably exit the medium quickly. Křivánek and d’Eon show that this technique can effectively reduce variance in the resulting image. They employ MIS [VG95] to combine it with Classical (C) sampling, where the unbiased phase function and free path distribution are used for sampling.

PoE sampling can be problematic in scenes where the medium surface is not approximated well by a slab (cf. Fig. 2). If the surface is not really planar, the PoE may not be a good estimate of the closest surface point, and the normal at the PoE may not be a good estimate of the overall surface orientation. To alleviate this problem, we propose biasing the random walk towards the location of the true closest surface point.

Also, the Point of Entry technique has difficulties in scenarios where a light source opposite the PoE is strong enough to have a non-negligible contribution to the image. The biasing draws most samples away from the light source in this case, which leads to an increase in variance. We propose to use an additional biasing technique that shifts the random walk towards a point placed on a light source.

### 3. Improving Biased Sampling

In this section, we detail our proposed improvements to the Dwivedi sampling technique as introduced by Křivánek and d’Eon [Kd14]. We call their technique C+PoE, or Classical + Point of Entry sampling.

#### 3.1. Closest Surface Point

Instead of using the normal at the point of entry as the slab normal, we propose using the direction towards the Closest Surface Point (CP) \( \mathbf{x}_{CP} \).

We employ a Bounding Volume Hierarchy to efficiently compute the Closest Point Transform for the current vertex position \( \mathbf{x} \). We then use
\[
\mathbf{n}_{CP} = \frac{\mathbf{x}_{CP} - \mathbf{x}}{||\mathbf{x}_{CP} - \mathbf{x}||} \quad (14)
\]
as the slab normal. Note that in contrast to PoE sampling, we do not use the surface normal at the CP directly, because that can be problematic in scenes with highly displaced surface geometry.

We search the CP only once per random walk, specifically at the first vertex inside the medium, and always bias towards the same surface point. However, the direction \( \mathbf{n}_{CP} \) is updated at every vertex. It is possible to also recompute a new \( \mathbf{x}_{CP} \) at every vertex, but our tests did not show a benefit that would justify the increase in cost.
3.2. Incident Illumination

To handle non-uniform illumination better, one could use next event estimation from the volume to connect to light sources. However, this may not be practical due to refractive surfaces such as rough dielectrics [WMLT07] found on skin [HDF15, Hol15], or because the medium is so dense that transmittance to the surface is practically zero.

Instead, we propose sampling one light vertex per random walk, and biasing towards it at each step. Using importance sampling, we can choose light vertices proportional to their emission. We ignore visibility in this step. In scenes with strong backlights, this technique biases most paths towards the brightest light sources in the scene, leading to much better convergence in translucent regions.

The slab direction can be computed in a fashion similar to Eqn. (14) for area lights or point lights. For directional light sources such as environment maps, we can use the light direction as the slab normal: \( \mathbf{n}_{DL} = (\mathbf{s}_{DL} - \mathbf{x})/||s_{DL} - \mathbf{x}|| \)

3.3. Multiple Importance Sampling

Similar to Krivánek and d’Eon [Kd14], we combine the Classical, Closest Point and Incident Illumination sampling techniques using Multiple Importance Sampling with the one sample model and the balance heuristic [VG95].

This is imperative with biased sampling techniques: since the transmittance and phase functions do not cancel out of the contribution function with non-perfect importance sampling, low-probability random walks with a high contribution lead to excessively bright pixels, or fireflies. Multiple Importance Sampling can detect such paths, and reduce their weight automatically.

A new sampling technique is selected randomly at every vertex of the random walk, but biasing techniques have zero probability before the first volume vertex. This means that the first volume vertex is always sampled using Classical sampling.

3.4. Selection Probability for Classical Sampling

In Krivánek and d’Eon [Kd14], a fixed percentage of samples is allocated to Classical sampling. We follow this approach for constant phase functions, where we use a probability of 0.1 for selecting Classical sampling.

In media with anisotropic phase functions, such as the Henyey-Greenstein phase function [HG41] with mean cosine \( g \neq 0 \), directional biasing creates random walks with a low probability density if the slab normal and the ray direction are not aligned. This is demonstrated in Fig. 5. We use an adaptive criterion that reduces the probability for selecting a biased sampling technique for anisotropic scattering. It basically lets the sampling degrade gracefully to pure Classical sampling. We found that
\[
p^c = \max \left\{ 0.1, |g|^{1/3} \right\}
\]
yields results with low noise over the full range of possible mean cosines. Note that this may differ if other phase functions are used, especially complex, highly anisotropic phase functions. However, it should be noted that biased sampling may generally not be the best choice for media with such phase functions.

3.5. Selection Probabilities for Biased Sampling

In a medium with extinction coefficient \( \mu_e \), the expected number of volume events on distance \( t \) is \( \mu_e \cdot t \). If a direct random walk of length \( t \) is created using perfect Classical sampling, the contribution of that random walk will be weighted by a factor of \( \alpha^{w_{DL} \cdot t} \).

We use this value to define the probability for selecting one of the three biasing techniques:
\[
p_i^b = (1 - w^c) \frac{\alpha^{N_i}}{\sum_j \alpha^{N_j}},
\]
where \( N_i = \mu_e \cdot t_i \) is the expected number of events to the slab surface of technique \( i \). In essence, if the medium has a high single-scattering albedo \( \alpha \), travelling a large distance and across multiple scattering events may be worthwhile.

When Incident Illumination biasing is used, random walks tend to get rather long, since they may traverse the whole medium before exiting. This is generally more expensive than PoE or CP biasing, which exit at a nearby surface. However, if the contribution is large enough, it may still be worth performing such long walks.

**Committing to Incident Illumination Sampling.** It would be problematic if the two different types of biasing techniques competed for the random walk, sending it back and forth. For this reason, we
set the probability of selecting either PoE or CP to zero once II is selected for the first time. This means that a random walk commits to biasing towards the light source. We always keep a minimum probability of sampling using the Classical technique, however.

4. Results

All results were rendered on a machine with eight Intel Xeon E7-8867 v3 processors running at 2.5 GHz. We used a total of 256 concurrent threads. Render times are given in thread hours.

We implemented the reference and proposed techniques in our research renderer, which uses single-wavelength spectral transport. The ray tracing backend is Embree [WWB*14]. The renderer does not impose a maximum length for random walks. Terminating paths after a given number of bounces will increase performance in practice, but tends to miss important long, but high-contribution light paths that occur in the dense scattering scenes we show in our paper.

Our scenes feature a large dynamic range, and so we had to apply some amount of tone mapping to ensure acceptable printing. Still, we recommend viewing results on a computer screen. Both tone mapped and out-of-renderer images are provided in supplemental material as OpenEXR files.

We will now describe our test scenes, and analyze them afterwards. Material parameters for our scenes are summarized in Tab. 1.

### 4.1. EAR Scene

The EAR scene contains three area light sources: two fill lights illuminate the scene from the front, but are not directly visible. A strong backlight is visible in frame, and causes visible sub-surface scattering in the ear lobe. The ear lobe is approximately 5 cm tall.

We show the full rendering and enlarged crops from the cheek and the ear in Fig. 6. The crops were taken from roughly equal-time renders, at 12 thread hours each. Convergence is shown in the RMSE plots in Fig. 7.

### 4.2. CANDLE Scene

The only source of light in the CANDLE scene is a brightly emitting object approximating a flame. It causes sub-surface scattering throughout the candle, which is approximately 12 cm tall. The wax is a heterogeneous volume with spatially varying density, which results in a pattern of brighter and darker spots.

We show the full rendering and crops from the bottom part of the candle in Fig. 8 (right). The crops are roughly equal time at about 4.25 thread hours each. Convergence is shown in the plots in Fig. 9.

### 4.3. BUDDHA Scene

The statuette in the BUDDHA scene is roughly 20 cm tall. Again there is only one source of illumination, a strong light illuminating the scene from behind.

Again, we show the full rendering and crops from the bottom part of the statuette in Fig. 8 (left). The crops are roughly equal time at about 4.25 thread hours each. RMSE is shown in the plots in Fig. 9.

### 4.4. Analysis

All scenes highlight a typical failure case of Point of Entry sampling: large contributions from random walks that cross the medium and do not return to the PoE. This is especially striking in the crops of the CANDLE and BUDDHA scenes, where PoE sampling performs far worse than Classical sampling. The EAR scene is the only one where PoE performs best in some parts: random walks entering at the cheek must exit there to have measurable contribution. Because of the low curvature, PoE and CP are almost identical, but the PoE can be computed more efficiently.

It is this problem that motivated our work in the first place. Our first proposed technique, Closest Point sampling, improves matters slightly. This technique has at least some chance of finding the light paths that PoE actively tries to avoid: If the first vertex inside the medium is closer to the back side than the front, CP will bias towards a back light. Consequently, a slight improvement can be seen in the EAR and CANDLE scenes. However, CP is no magic bullet. It also tends to favor turning around to exit the way it entered the medium.

Adding Incident Illumination sampling helps in scenarios where the illumination is not uniform, and where the light sources happen to not be in the direction of PoE or CP sampling. We find that biasing using II can speed up convergence by an order of magnitude in problematic regions when compared to Classical sampling. Compared to PoE biasing, the speed up can be even more dramatic, e.g. in the CANDLE scene where almost no random walks make it to the inside of the candle.

5. Discussion and Future Work

In this paper, we demonstrated several practical improvements to the Dwivedi sampling technique as introduced by Krivánek and d’Eon [Kd14].

We proposed replacing the Point of Entry biasing technique with a more accurate Closest Point sampling, and showed that non-uniform illumination can be handled more robustly with Incident Illumination sampling. We also proposed a heuristic approach for combining these biasing strategies with Classical sampling using Multiple Importance Sampling.
Some open questions remain. For example, our current selection criterion only implicitly considers the strength of incident illumination by importance sampling light sources. It is conceivable that using a simple analytic model, such as a diffusion approach, to estimate incident illumination from a given slab direction, could yield further improvements.

Non-constant phase functions are still difficult to handle with biased sampling. Our solution is to gracefully degrade to Classical sampling with increasing anisotropy. However, it would be more elegant, as suggested by Krivánek and d’Eon [Kd14], to use more exact analytical expressions that incorporate anisotropy directly.

Dwivedi sampling uses the assumption of a homogeneous medium. We showed in the Classical heterogeneous media. However, further applications such as clouds warrant future investigation.

6. Acknowledgements

The first author of this paper was funded by Weta Digital Ltd.

References


[DH05] Donner C., Jensen H. W.: Light diffusion in multi-layered translucent materials. ACM Trans. Graph. 24, 3 (July 2005), 1032–1039. 2


Figure 6: The Ear scene, which features a strong backlight behind a translucent ear. Equal-time comparisons at roughly 9.9 thread hours each are shown on the right.

Figure 7: Root Mean Squared Error for the full Ear scene as shown in Fig. 6 (top), the cheek crop (middle), and the ear crop (bottom).
Figure 8: The CANDLE and BUDDHA scenes. The enlarged regions show roughly equal-time comparisons at about 4.25 thread hours each.

Figure 9: Root Mean Squared Error for the CANDLE (left) and BUDDHA (right) scenes as shown in Fig. 8.