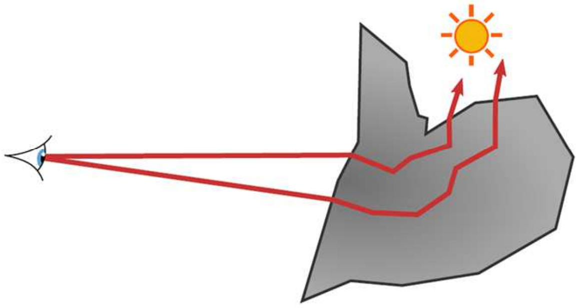
Zero variance-based sampling schemes (a.k.a. path guiding)

Jaroslav Křivánek
Charles University, Prague – Render Legion | Chaos Group

Computer Graphics Charles University

- This part of the course addresses the idea of “path guiding” – and how it can be used to reduce variance of Monte Carlo volumetric light transport simulation.
- We will also show that the theory of zero-variance random walks, originally developed in the neutron transport literature, can serve as a convenient theoretical framework for path guiding methods.

Path guiding – The idea



- The intuitive idea of path guiding is straightforward: sample paths in such a way that they can preferably reach `important' parts of the scene (e.g. reach the light sources, if we sample paths from the camera, as in path tracing).

Zero-variance path sampling in volumes

- A theoretical framework for path guiding
- Set of **local sampling rules** yielding **globally optimal path sampling**

- In order to achieve that, we need to design appropriate probability distributions to be used in path sampling.
- The theory of zero-variance random walks provides such sampling distributions.
- More precisely, the ZV theory provides a set of local sampling distributions that provably yield globally optimal sampling of the path space (in the sense that the resulting estimator will have zero variance)

- The idea that a random walk can be constructed in such a way that it always yields the correct answer with absolutely no variance has been around for almost as long as MC methods themselves.
- Despite the zero variance theory being old, Hoogenboom's recent 2008 NSE article (*Zero-variance Monte Carlo Schemes Revisited*) is very important: he corrects some misconceptions about the uniqueness of zero-variance walk construction that have lingered for several decades, and includes discussions about boundary crossing and track-length estimators as well.
- Booth’s 2012 article (*Common misconceptions in Monte Carlo particle transport*) further clarifies and generalizes some of the concepts of zero-variance schemes. He argues that Hoogenbooms’ conclusion concerning the uniqueness of the zero-variance construstions are not correct and that there are multiple ways in which a zero-variance walk can be constructed.

Zero-variance path sampling in volumes

- Theoretical construct – ZV cannot be achieved in practice
  - Requires knowing the radiance solution everywhere
- But it provides a guideline for variance reduction
  - Even approximate radiance solution yields low variance
- Obtaining the approximate solution
  - MC samples – Machine/statistical learning
    - [Vorba et al. 2014, 2016] – direction sampling, path termination and splitting (surfaces)
    - [Herholz et al., conditional accept] – all decisions **(volumes)**
  - Analytic solution
    - [Křivánek and d'Eon 2014] – subsurface scattering

- While the ZV theory provides a convenient framework, it is a mere theoretical construct that cannot be readily applied in practice, the reason being that it needs the radiance solution everywhere in the scene.
- But the radiance solution is unknown up front since that is what we are attempting to compute in the first place.

- While this vicious circle may sound hopeless, the ZV theory does provide a certain guideline for thinking about path guiding, an ideal to strive to.
- In practice, this is realized by replacing the radiance solution by some convenient approximation, which then yields an approximation of the ZV scheme.

- How can we obtain such an approximate solution?
- One option is to use statistical/Machine learning techniques to reconstruct the solution and the ZV-based sampling distributions directly from the Monte Carlo samples used in the rendering itself (or in a separate pre-pass). Our work [Vorba et al. 2014, 2016] applies this idea to surface light transport. Our recent work [Herholz et al.] generalizes the idea to volumetric transport.

- Another, very different approach to obtain the approximate solution, is to employ an analytical light transport solution in a canonical case that resembles the situation under consideration.
- In the specific case that I will be talking about, we use this idea in MC subsurface scattering and the appropriate canonical case here is a flat, semi-infinite medium (half-space).

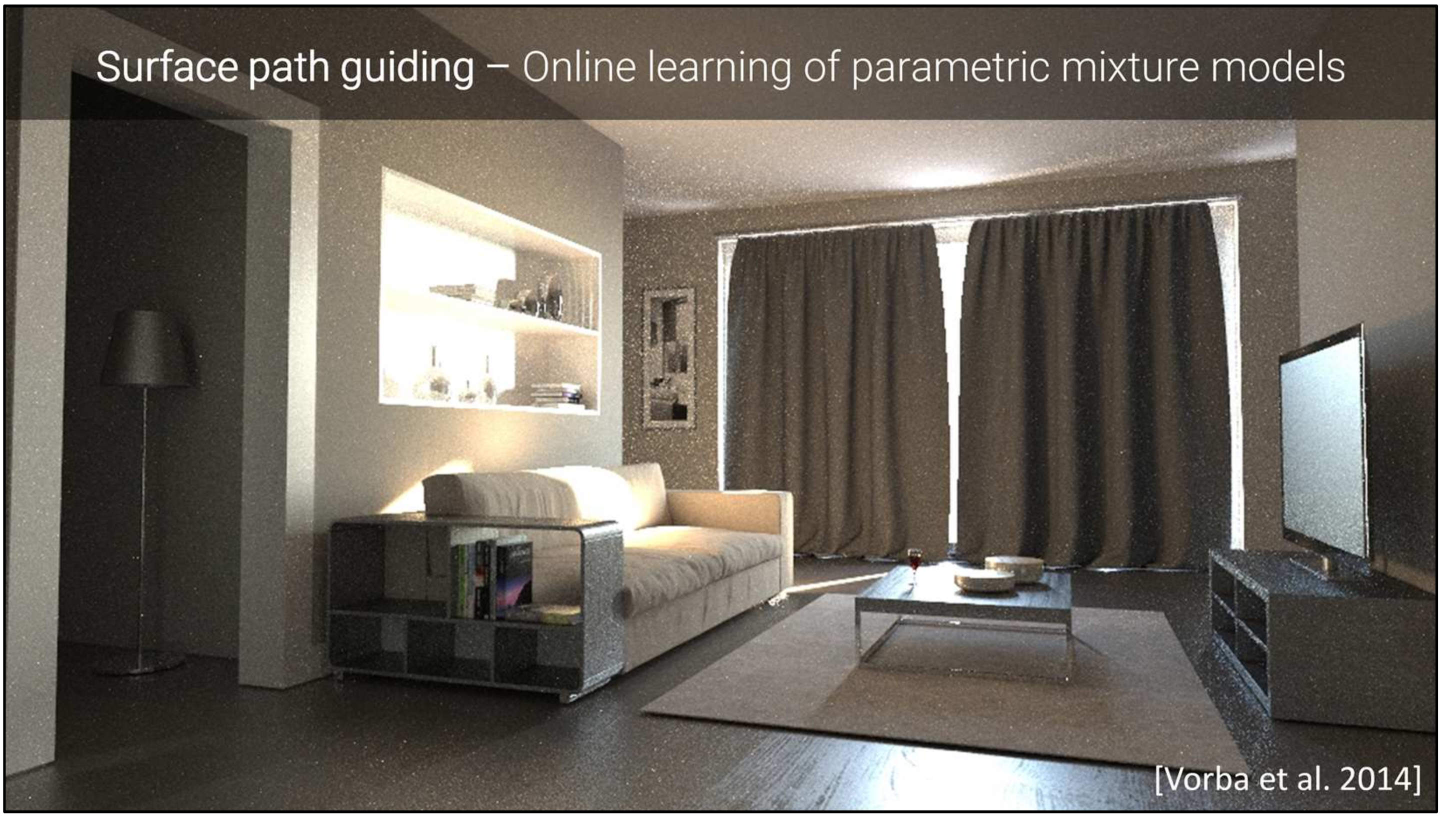
VOLUME PATH GUIDING

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Computer Graphics Charles University

DiSTRO

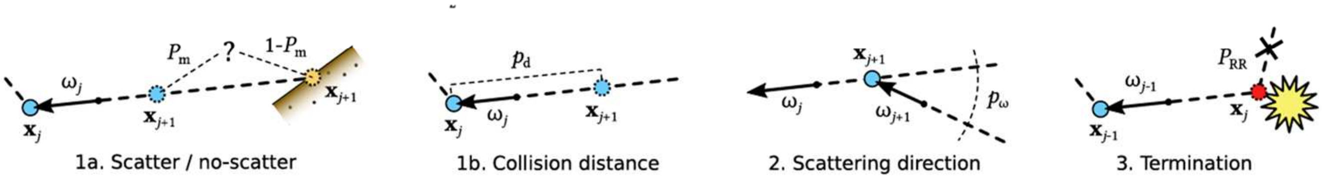
• Let me start by briefly showcasing our recent work on volumetric path guiding.



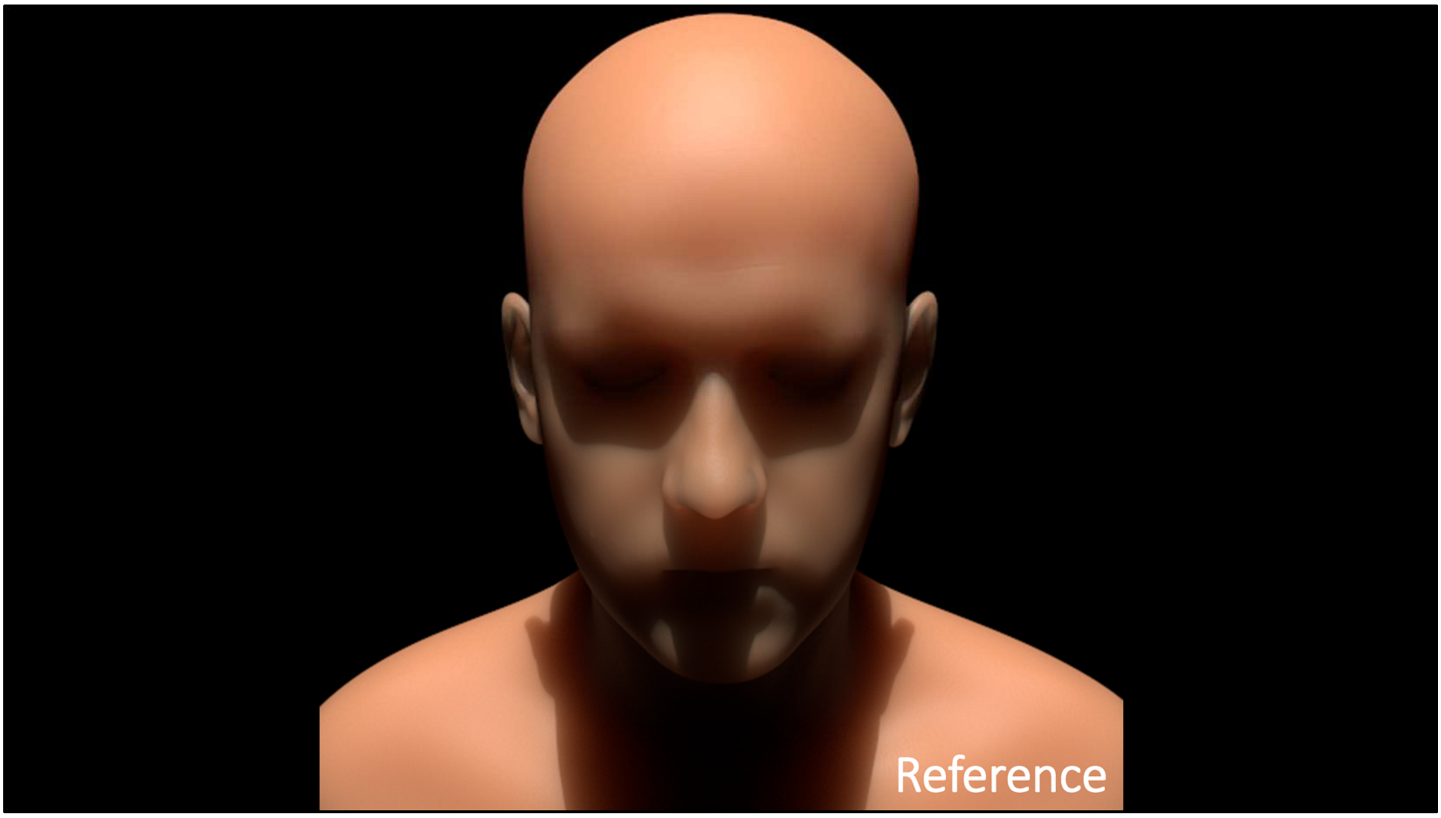
- The motivation, and a starting point, is our work on path guiding on surfaces [Vorba et al. 2014].

Volume path guiding

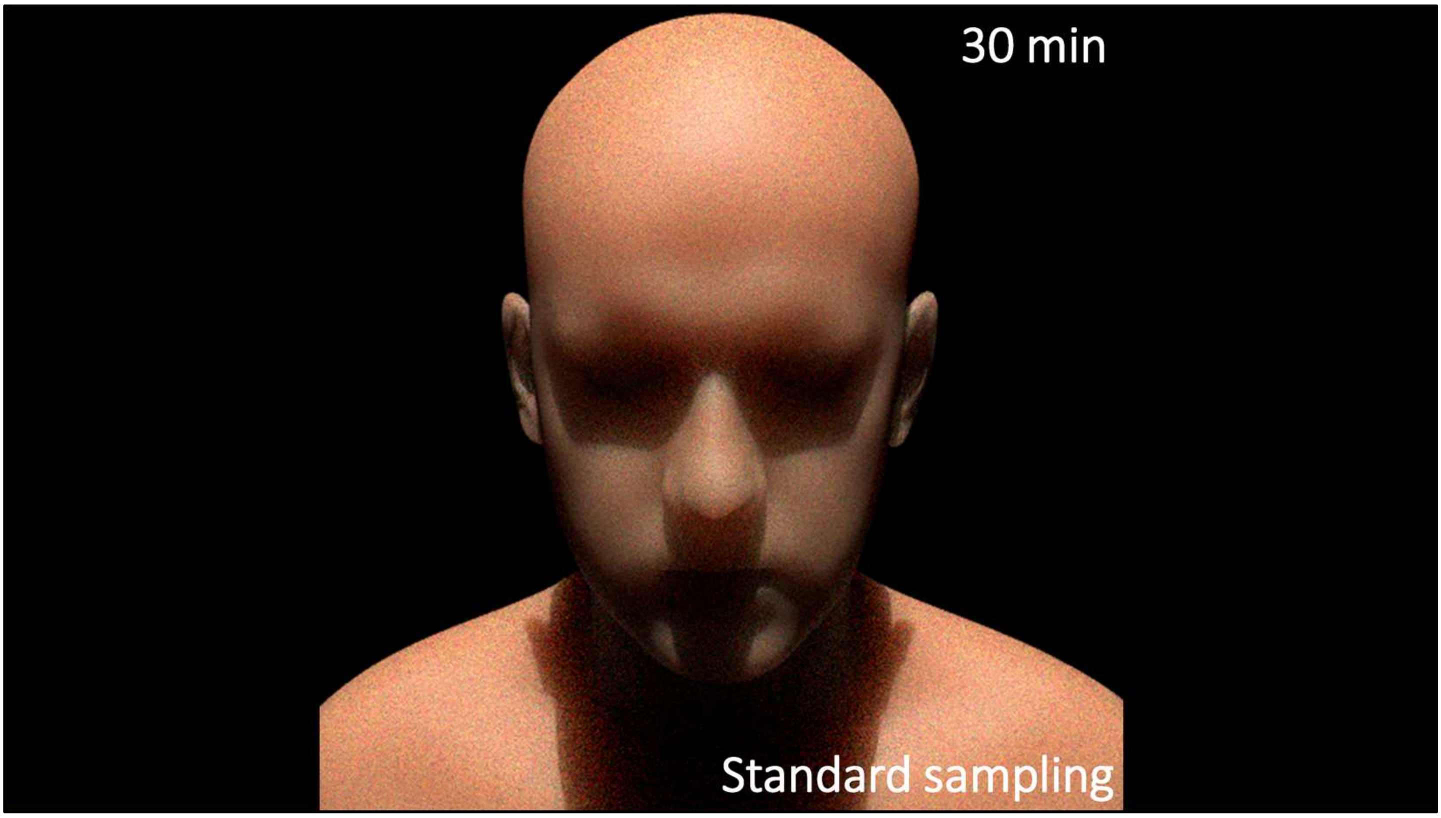
- All events importance sampled
- Product sampling for collision distance



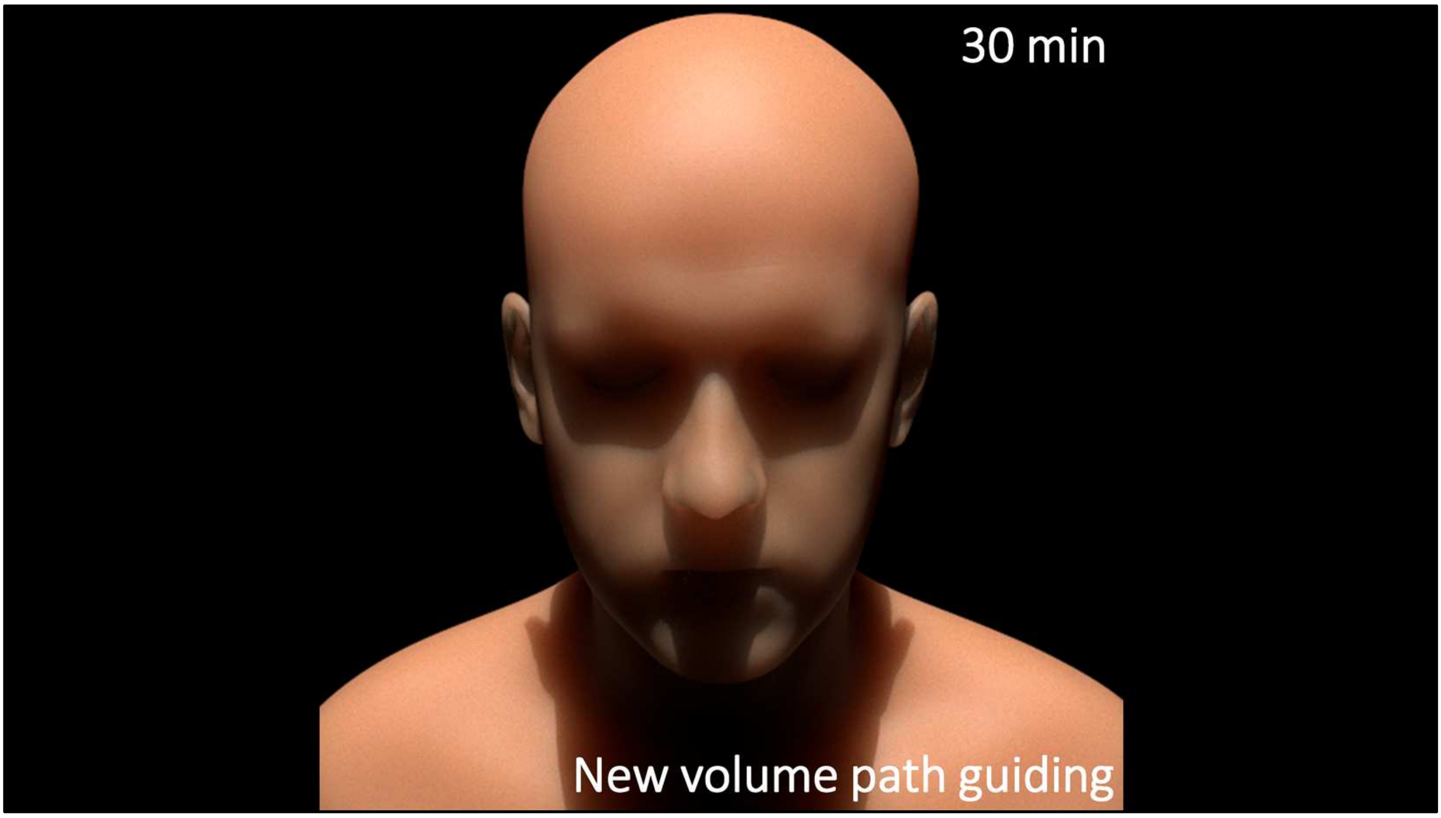
- In order to apply this idea to volumetric transport, all the various random decisions used when constructing a light transport path need to be appropriately importance sampled.
- This includes the selection of scattering distance along a ray, and the decision whether the scattering should occur in the volume or at the next surface interaction. These decisions are unique to volumetric light transport and do not appear in surface transport.
- Furthermore the decisions shared with surface transport include the choice of the scattering direction and random termination/splitting of the paths.



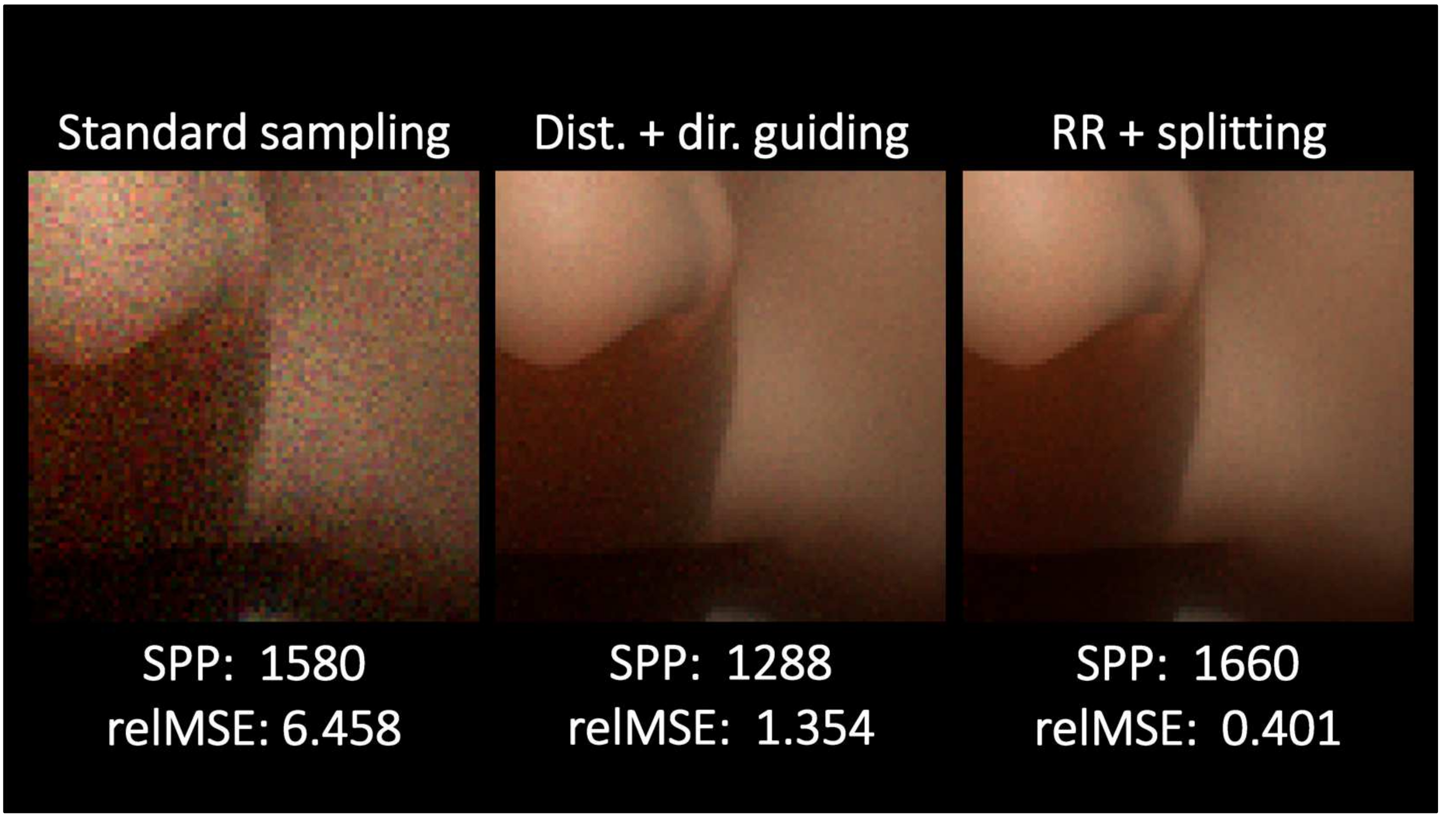
- Without giving any further details, let's have a look at some results.
- This is a homogeneous medium with scattering properties approximating these of a Caucasian skin.
- We use Monte Carlo path tracing to render the scene.



- With standard path sampling, we can see that even after 30 minutes of rendering, the image shows a significant amount of noise.



- Our volumetric path guiding based in the zero-variance sampling scheme yields a nearly clean image in the same time.

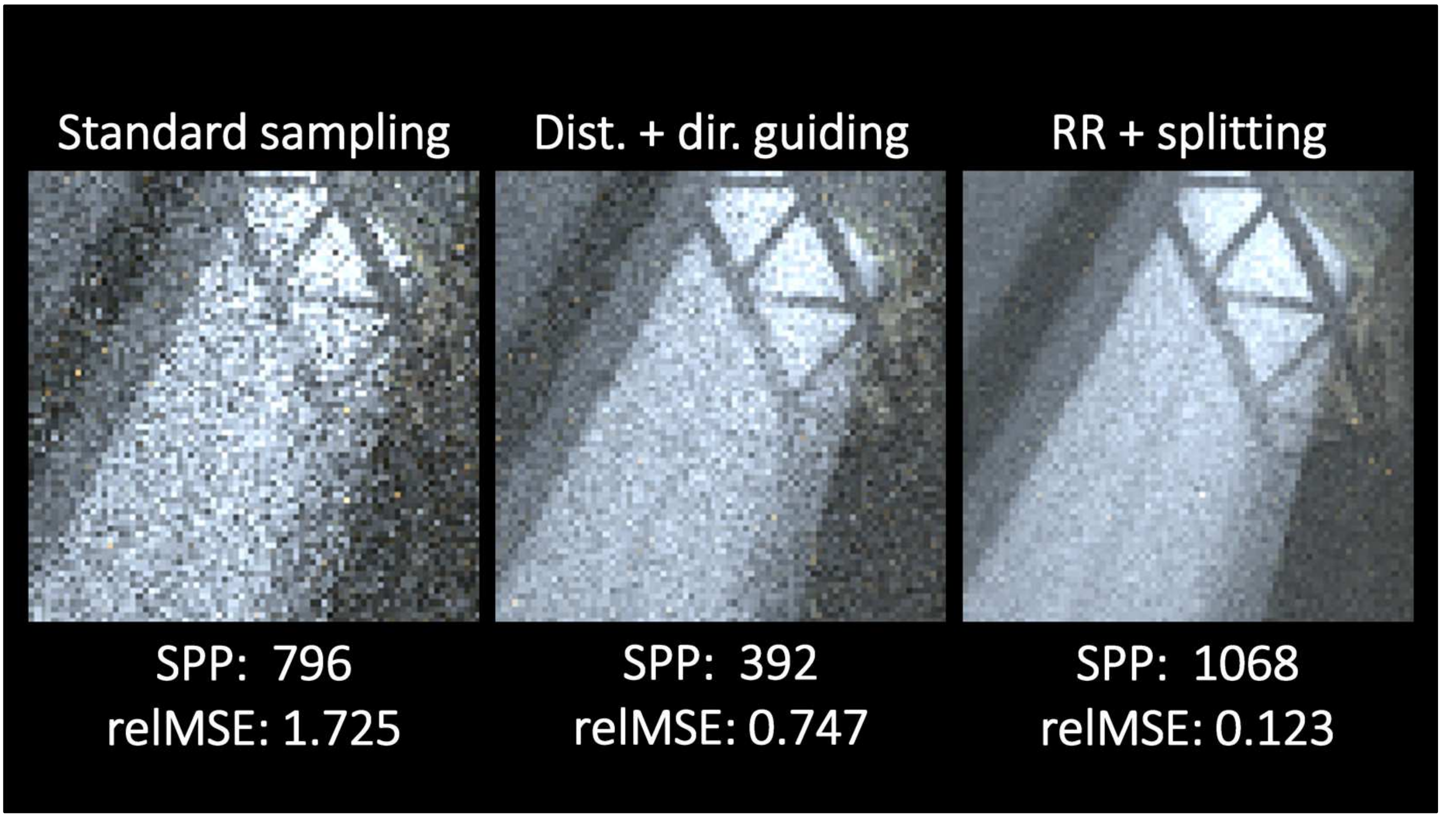


- This slide shows that the different random sampling decisions complement each other and together they yield the desired variance reduction.

- Here, we show the same technique applied to a very different scene – this time a natural history museum filled with thin haze and illuminated by light shafts.

- Again, the standard sampling shows a significant amount of noise…

- While the zero-variance-based path guiding provides a significant variance reduction.

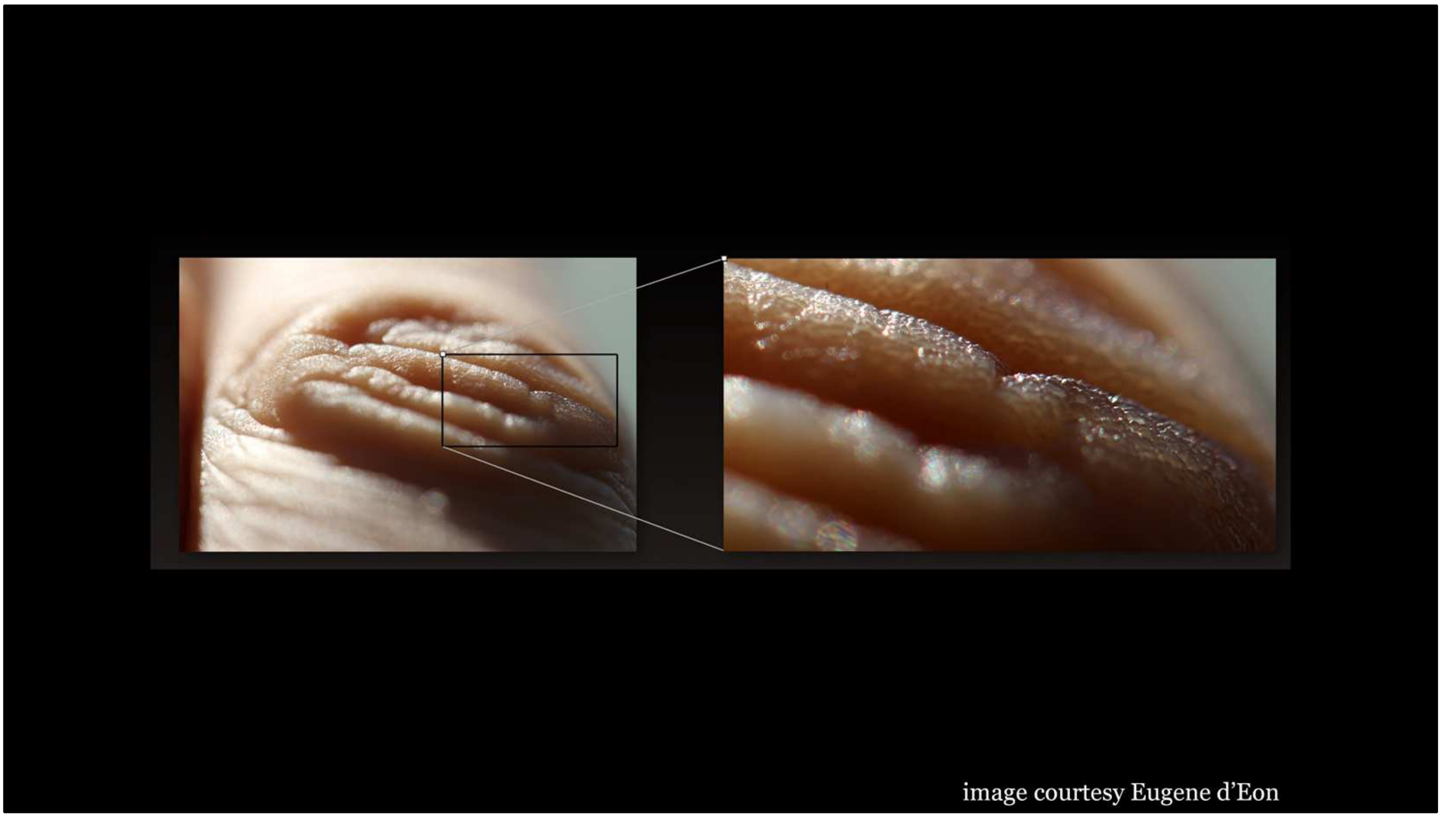


- Once again, we can see that the different random decisions add up to yield the final solution.

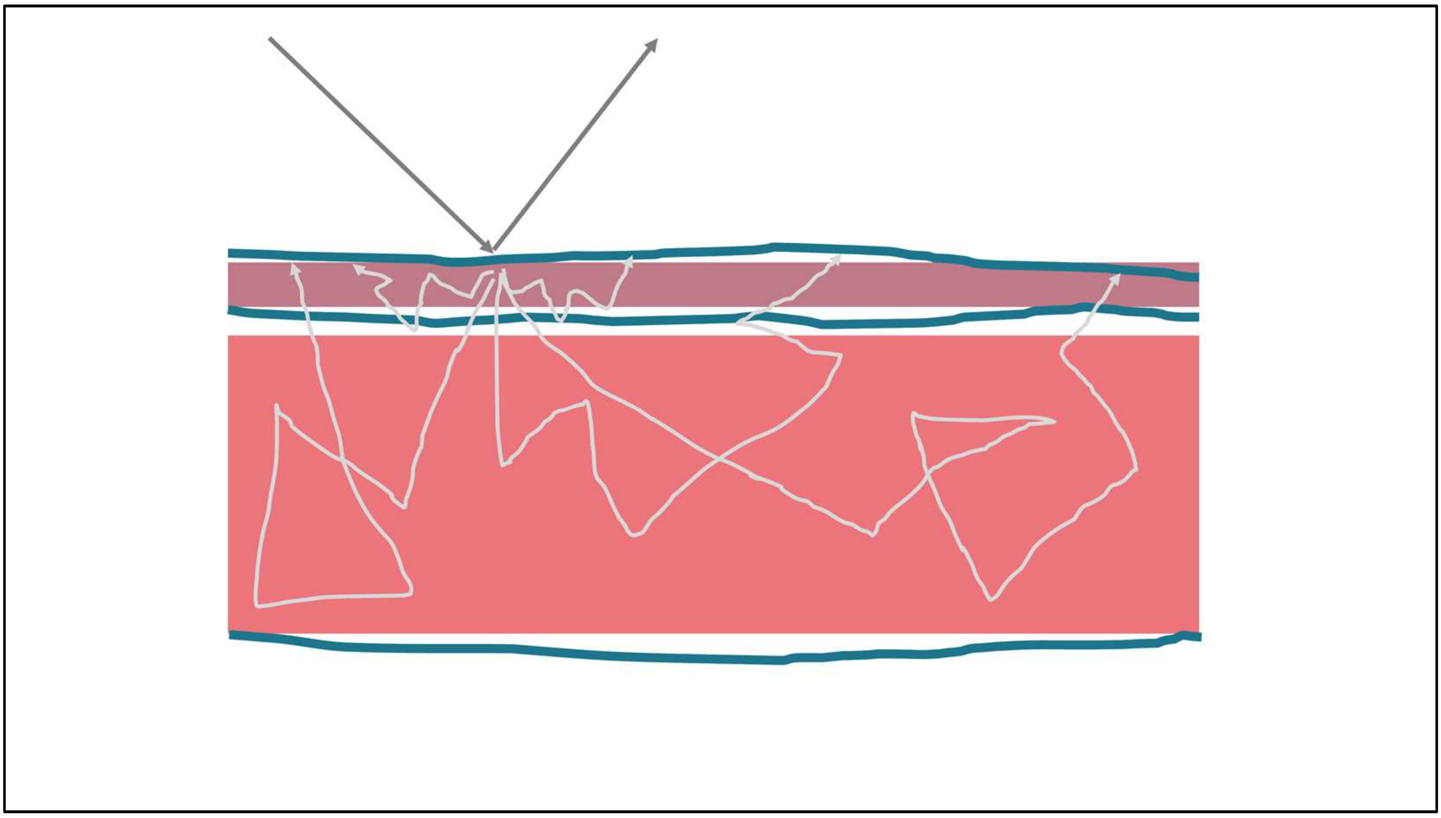
ZV-BASED SUBSURFACE SCATTERING

with Eugene d’Eon

• Let us now discuss path guiding based on an approximate solution obtained through an analytical solution in a canonical case.
• This work was done with Eugene d’Eon when I visited Weta Digital in 2013.
• It was presented as a talk at SIGGRAPH 2014 [Křivánek and d’Eon 2014].

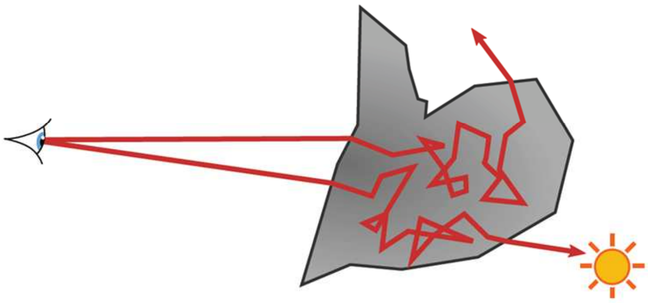


- Our primary application is subsurface scattering, notably in the human skin.



Classical random walk

- Tends to get lost in the medium

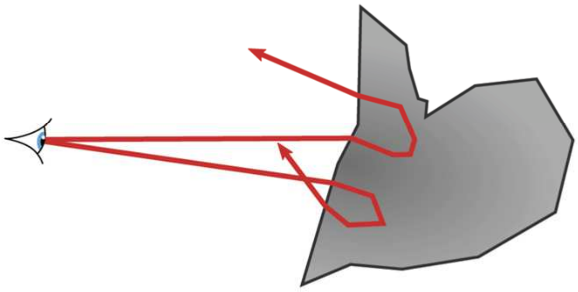


MC METHODS FOR VOLUMETRIC LIGHT TRANSPORT – ZERO VARIANCE-BASED SAMPLING SCHEKEMS (A.K.A. PATH GUIDING)

- When one applies classic path tracing for calculating subsurface scattering in the skin, the path sampling procedures are unaware of the fact that we are interested in calculating the solution at the surface boundary – instead, they tend to wander around in the medium without making a relevant contribution to the image.

Goal

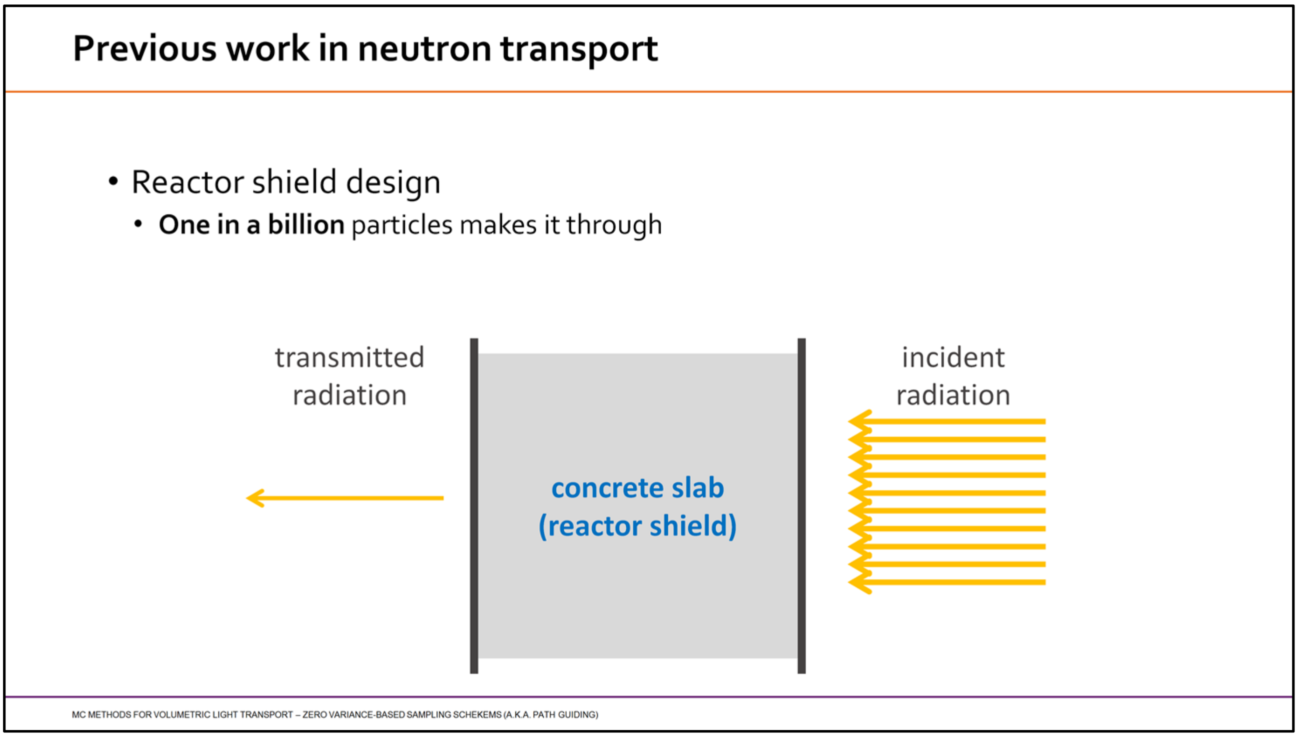
- Guide paths toward the boundary



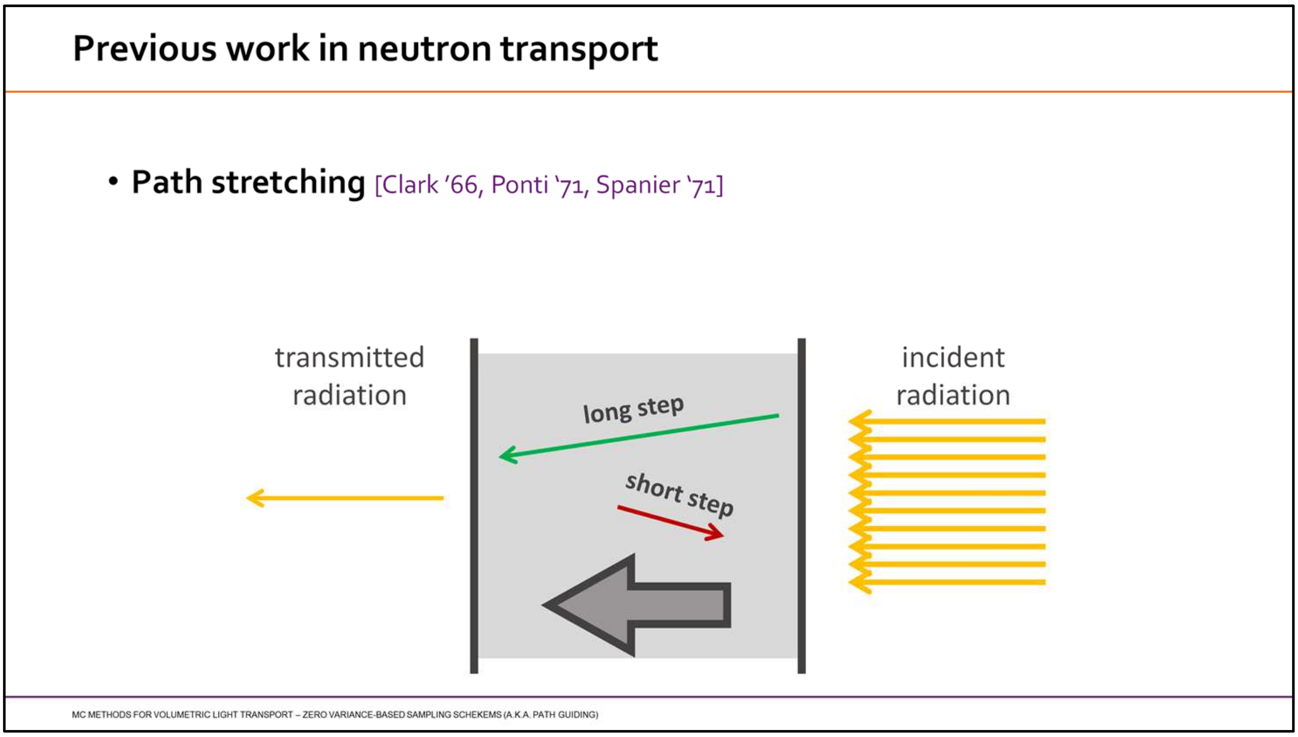
MC METHODS FOR VOLUMETRIC LIGHT TRANSPORT – ZERO VARIANCE-BASED SAMPLING SCHEKEMS (A.K.A. PATH GUIDING)

- Our goal here is to inform the path sampling procedure that it should preferably guide the paths outside of the medium.

- To solve this issue, we turn our attention to the neutron transport literature…



- … where similar problems are encountered in the reactor shield design calculations.
- The neutron transport literature refers to this class of problems as “deep penetration problems”.
- By design, only a tiny fraction of the incident radiation is allowed to pass through a reactor shield.
- For example, in a blind MC simulation only one in a billion particles would make it through, which makes the classic MC simulation totally hopeless.



- One of the approaches to solve this issue is the so-called "path stretching".
- The idea is to advance the particles toward the outside by artificially stretching the sampled distance whenever the particle in a MC simulation points toward the exterior, and to shrink it when the particle is directed to the interior. To compensate for this, one needs to adjust the weight of the particle, because its behavior no longer follows the laws of physics.
- Path stretching has originally been derived heuristically and relied on an ad-hoc parameter (the 'strength' of the stretching). While it often worked great, it could actually deteriorate the result (increase variance) when the stretching parameter wasn't set judiciously.

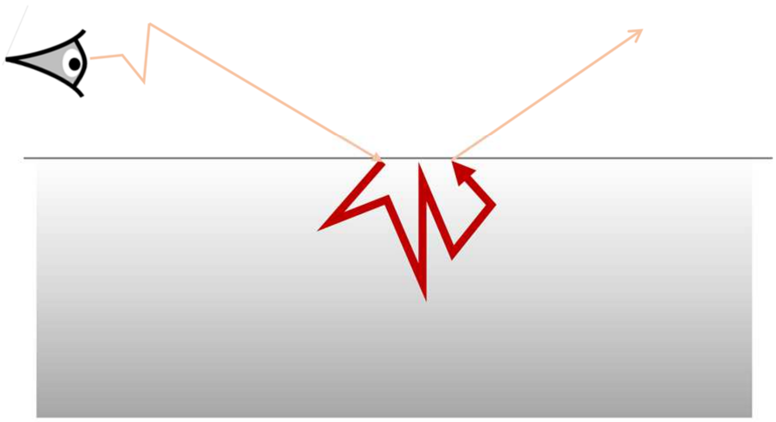
Zero-variance random walk theory

- **[Dwivedi '81]**
  - Optimal **path stretching**
  - **Idea:** If you **approximate the solution**, you can use this to **guide sampling**
  - Specific application of the **zero variance-based sampling**

- This is where the theory of zero-variance random walks comes into play.
- Dwivedi [1981] was the first to apply the theory of zero-variance random walks to deep penetration problems.
- He has shown how the heuristic path stretching automatically follows from the theory, while giving a clear answer to the parameter setting.
- He was also the first to show that to robustly reduce variance, the path stretching needs to be combined with an appropriate angular sampling.
- While he conceived his work with reactor shield design problems in mind, we show that it can be adopted to subsurface light transport, and we propose some further improvements.

Setup

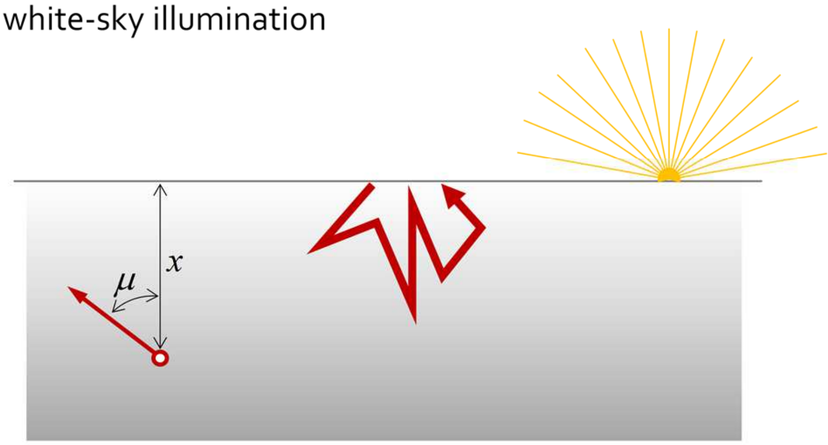
- (Unidirectional) path tracing



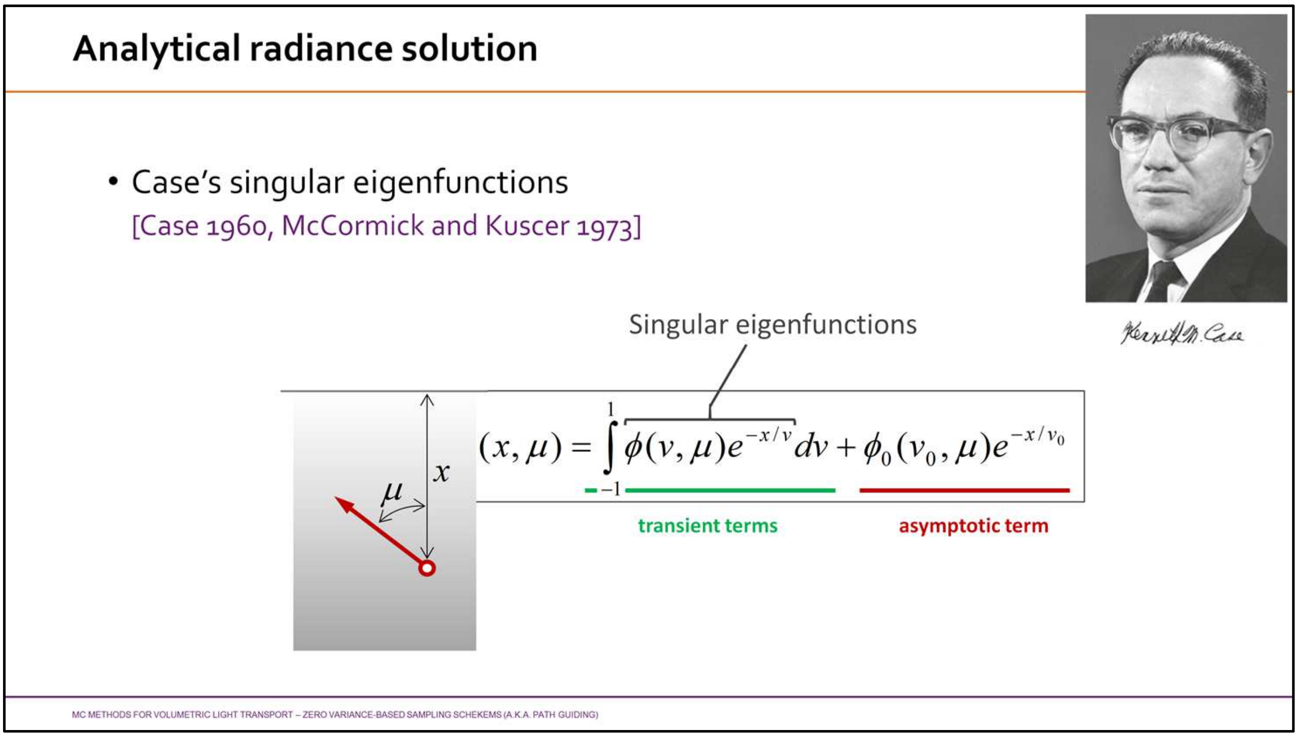
- We apply the technique in a unidirectional path tracer.
- The zero-variance-based random walk is used only for the part of the path under the surface.
- The rest of the path is not affected at all and follows the same rules as in a regular path tracer.

Assumptions

- Flat, semi-infinite medium
- Uniform, white-sky illumination



- Since we do not know what the path tracer will encounter after escaping from the surface, we assume, for the sake of constructing the zero-variance walk, white-sky illumination, uniform in the directional and spatial domains.
- This is equivalent to saying that ‘escape’ from the medium is our only source of importance that guides the random walk.
- Note that this is not a necessary step: the general theory (‘Caseology’ on the next slide) permits knowing (in theory) the full directional radiance distribution inside some volume due to a particular light source(s) outside of it. In this much more complicated case, the approximate internal radiance solution would guide the subsurface sub-paths not only to try to escape the medium, but also tend to positions which permit angle selections that then leave the medium in a direction that tends to hit the light. However, this would be quite complex in practice so we decided not to pursue this option.

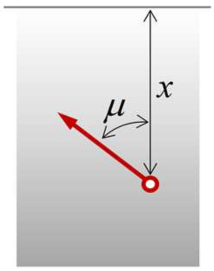


- In order to be able to build upon the zero-variance theory in constructing the random walk, we need an approximation of the radiance solution under the surface.
- Note that the Monte Carlo estimator that we construct is unbiased irrespective of the accuracy of this approximate solution; the accuracy only affects the variance reduction but not the unbiasedness of the resulting estimator.
- For this reason, we choose to approximate the true sub-surface distribution of radiance using a solution for a half-space with flat boundary.
- In this setting, the medium can be specified by the (optical) depth under the surface, *x*, and the direction cosine υ.

- The exact half-space solutions that we use can be derived in several ways, including what is known as Caseology, named after Kenneth Case who was on the theoretical division at Los Alamos building the first bombs. In 1960 Case studied in detail the spectrum of the transport operator in the plane-parallel case, and derived an expansion into the discrete asymptotic diffusion mode, and the continuous spectrum of singular eigenfunctions - so named because the angular distributions (the phi-functions on the slide) for the transient terms are singular. This study of the structure of exact solutions in transport is incredibly insightful and we are probably the first to exploit it directly in a rendering technique.
- Each eigenfunction of the spectrum has the same form: the spatial term that only depends on the depth under the surface is a simple exponential, with different eigenfunctions having different decay rates. The angular term then corresponds to integrating the spatial term along a line from a given depth in a given direction.

Approximate solution

- Drop transient terms

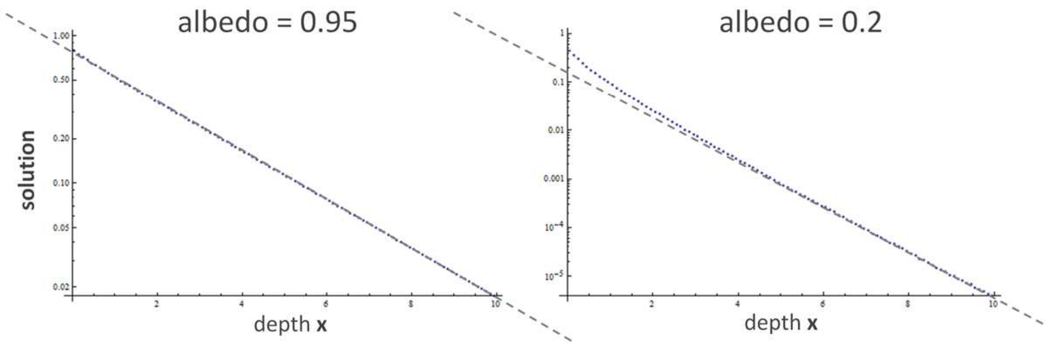


$$L(x,\mu) = \underbrace{\phi_0(v_0,\mu)e^{-x/v_0}}_{\text{asymptotic term}}$$

- Since the complete solution is rather complex – being given by an integral over the spectrum – Dwivedi choose to only use the discrete (asymptotic) mode of the spectrum, disregarding its continuous part.
- This is a reasonable assumption because the spatial part of the discrete mode has a slower decay rate than the rest of the spectrum. Therefore, as one moves away from the boundary, this terms dominates all the other terms and *asymptotically* becomes the correct solution.

Approximate solution

- Only the asymptotic term

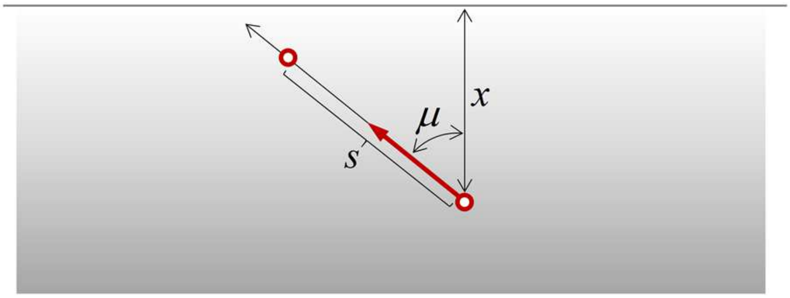


Distance sampling

$$p(s) \propto \underbrace{\exp(-s\sigma)}_{\text{transmittance}} \cdot \underbrace{\exp(-x/v_0)}_{\text{solution}}$$

$$\propto \exp(-s\sigma') \qquad \sigma' = \sigma(1-\mu/v_0)$$

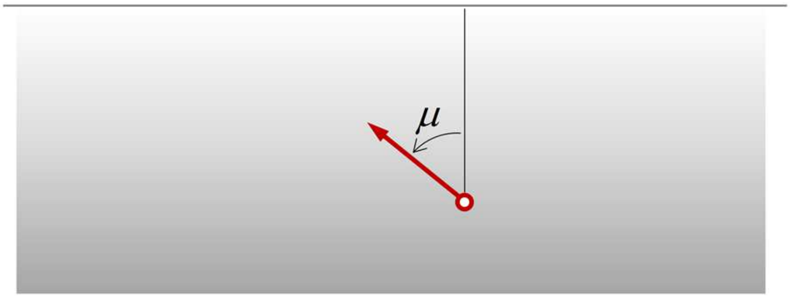
path stretching!



- We follow Dwivedi and use only the discrete, asymptotic term as an approximation of the true solution under the surface for constructing the random walk.
- Given a particle at a state (x, u) under the surface, we want to randomly sample its next state (x', u').
- This is decomposed into sampling the next collision distance $\underline{x}'$ along the line from x in the direction u, followed by sampling the new particle direction u'.
- When sampling the collision distance, the zero-variance theory tells us that we should do this with a pdf proportional to the product of the spatial part of the transport kernel (i.e. transmittance = exponential) and the solution (which we assume to be the Case's discrete mode = exponential).
- This yields an extremely convenient result, where we sample from a simple exponential pdf, where the transport coefficient is modified from its true value using the equation shown on the slide.
- Note that this is exactly equivalent to the idea of path stretching: for directions toward the boundary the sampled distance will be stretched, whereas it will be shortened for directions away from the boundary.
- Note that the v_0 parameter is a constant here, because it only depends on the single scattering albedo of the medium.

Directional distribution

$$\boxed{p(\mu) \quad \propto \quad \frac{\nu_0}{\nu_0 - \mu}}$$

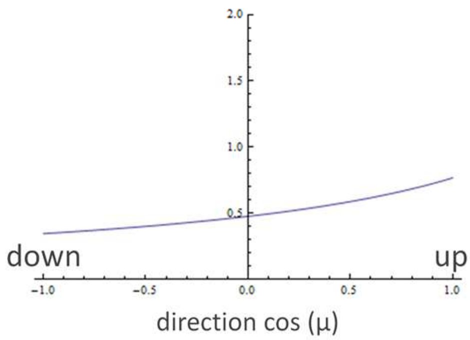


- The work of Dwivedi was the first to point out that path stretching must be accompanied by a corresponding modification of the direction sampling in order to achieve a robust variance reduction.
- This follows naturally from the zero-variance theory: we need to sample the direction from the product of the angular part of the transport kernel (=phase function) and the angular part of the true solution.
- In our work, we assume isotropic media with a constant phase function, so we only need to sample from a pdf proportional to the angular part of the solution.
- This can be derived simply by integrating the spatial part of the solution (=exponential) along lines of different direction.
- Furthermore, in deriving the angular solution, Dwivedi assumes that the spatial solution extends beyond the boundary surface, so he does not need to treat the direction toward the boundary as a special case.
- This yields an extremely elegant solution for the angular part, which is independent of the depth under the surface, and can be easily normalized and used for sampling in a closed form.
- Note that because we only assumed 'escape' as our goal, the 'guided' direction selection only involves modifying the selection of the direction cosine *u*: the azimuthal angle selections at each step are chosen uniformly. This is one feature that would change if, say, you knew the source of light outside the medium came from a particular direction. In these cases, more advanced deterministic solutions in the interior of the medium would be required, and works of Jakob et al. [2014] (*A Comprehensive Framework for Rendering Layered Materials*) and d'Eon [2014] (*A Dual-Beam 3D Searchlight BSSRDF*) could potentially both be used to produce such importance functions for path guiding (for the plane-parallel directional light source case, or the point source near a subsurface surface case).

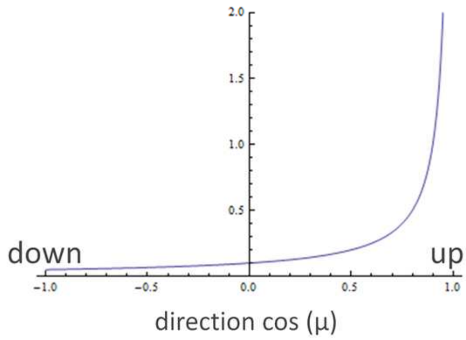
Directional distribution

$$p(\mu) \quad \propto \quad \frac{v_0}{v_0 - \mu}$$

albedo = 0.95

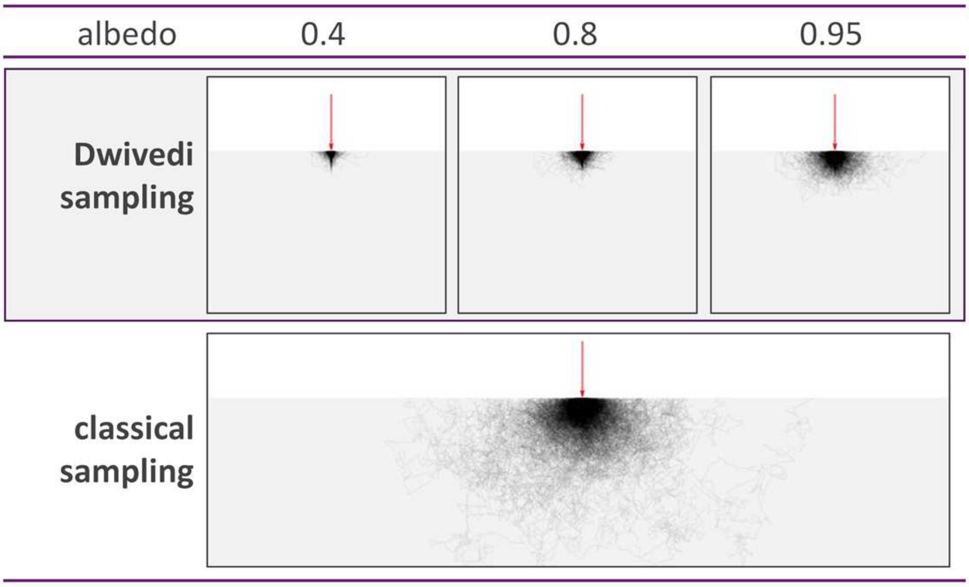


albedo = 0.2

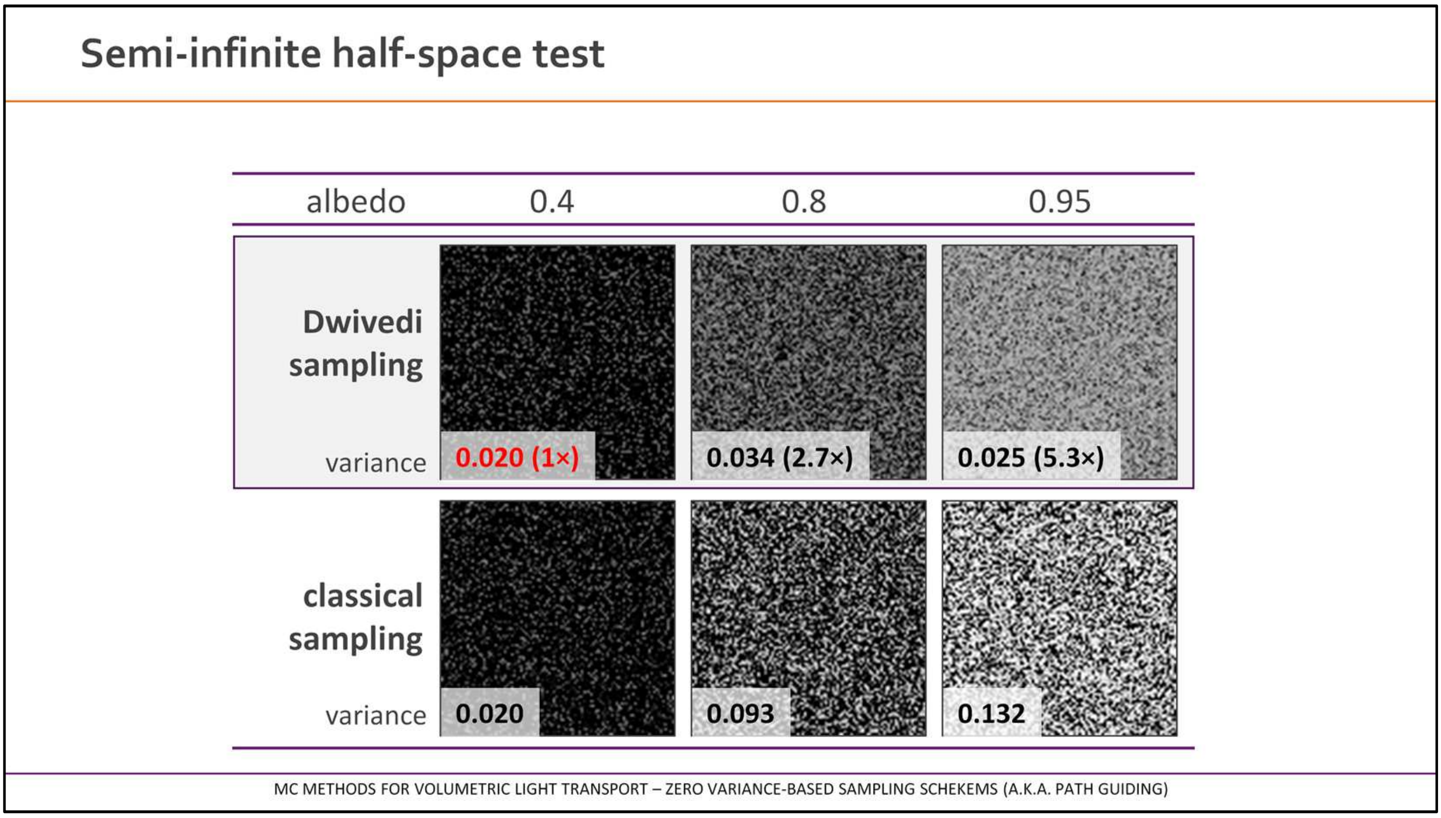


- These plots show the resulting angular pdf (the horizontal axis corresponds to the direction cosine and the vertical axis is the pdf value).
- We can see that for low absorption (high albedo) the solution is mostly uniform, which corresponds to the well-known fact that the diffusive multiple scattering in low-absorption media leads to a solution that is mostly uniform in directions.
- For high absorption, we can see a pronounced peak of the distribution with strong preference to sampling directions toward the boundary.

Semi-infinite half-space test



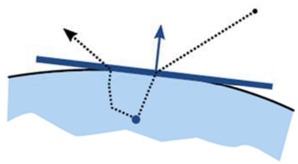
- Before using the method in rendering, we have tested it in a simple simulator of light transport in a half-space.
- On this slide, we demonstrate the effect that the new method has on the trajectory of light paths under the surface.
- In the classical sampling the paths tend to wander quite far from the point of entry, while the new sampling scheme concentrates most of the sampling effort around the entry point. This effect is more pronounced for higher absorption levels (lower albedos) because in this regime, the diffusive multiple scattering has only low effect on the final result.



- The images on this slide correspond to rendering subsurface scattering on the flat, index-matched boundary of a semi-infinite medium under white-sky illumination.
- These are exactly the assumptions used to build the zero-variance walk, so it is the best case for what we can achieve.
- We can observe substantial variance reduction for low absorption, i.e. higher albedos (0.95 roughly corresponds to human skin).
- There is little, if any, variance reduction for lower albedos. This is due to the fact that the assumed shape of the solution does not match the true shape (see slide 27).
- However, when we take the computation time into account, we obtain a substantial net improvement in efficiency even in those cases because the resulting sub-surface walks are much shorter on average.

- Note that in the case that there is no absorption (albedo = 1), the classical subsurface walk is already zero variance: you just keep sampling (with whatever sample weight you had as you entered) until finally you exit: with that same weight no matter where you went, so there is nothing to improve upon in that case (unless, as mentioned above, you knew more about where light sources might enter the medium).

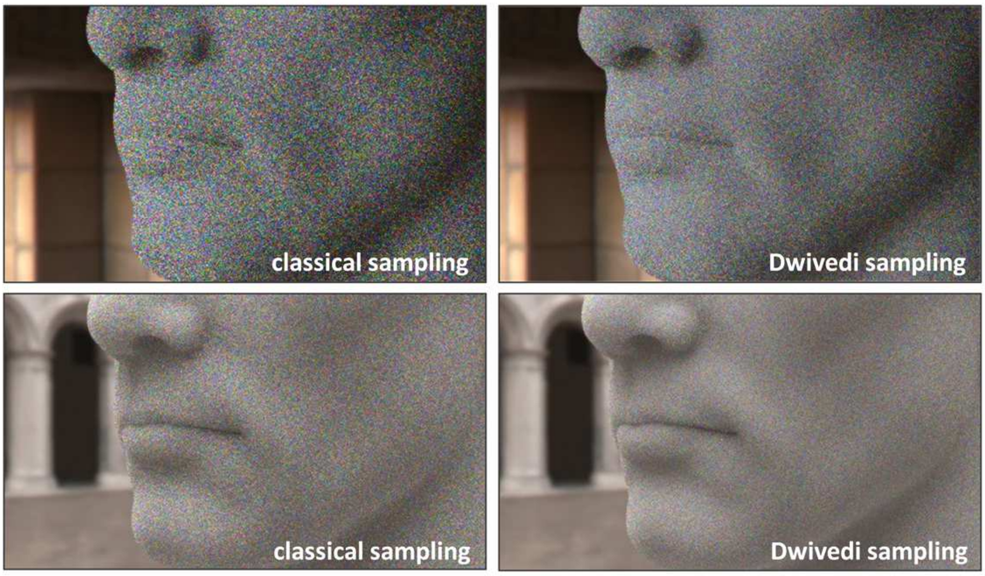
Use in rendering



- To use the method in rendering, we assume that the half-space is aligned with the surface normal at the point where the path enters under the surface.

Use in rendering – IBL

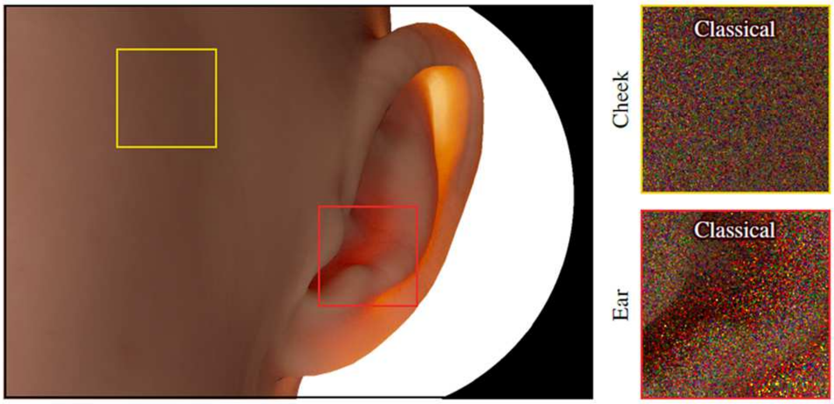
equal-time comparison, 100 samples per pixel, MIS 25% classical + 75% Dwivedi



- Despite the fact that the importance function driving the sampling assumes uniform hemispherical illumination, the modified path sampling lowers variance even when the illumination is nonuniform.
- The images rendered with classical sampling use 100 samples/pixel while in our results we trace about 50% more samples/pixel in the same time.
- While the speedup is a profitable side-effect, most of the variance reduction is due to the sampling pdfs closely approximating the zero-variance sampling scheme.
- As on the previous slide, the subsurface medium has a single scattering albedo of 0.943 for all wavelengths and index-matched smooth boundaries.
- The individual rows show results for different environment maps.

Better application to rendering

- [Meng et al. 2016]



MC METHODS FOR VOLUMETRIC LIGHT TRANSPORT – ZERO VARIANCE-BASED SAMPLING SCHEKEMS (A.K.A. PATH GUIDING)

- The follow-up work of Meng et al. addresses the problems encountered when using the above method in media that significantly differ from a semi-infinite half-space, such as the ear shown on the slide.
- In addition to aligning the half-space with the point of entry (PoE), they consider aligning it with the normal of a point on the “the other side” of the medium (CP).
- Additionally, the consider the direction to a dominant light source (Il).
- Combining all these techniques yields a more robust solution.

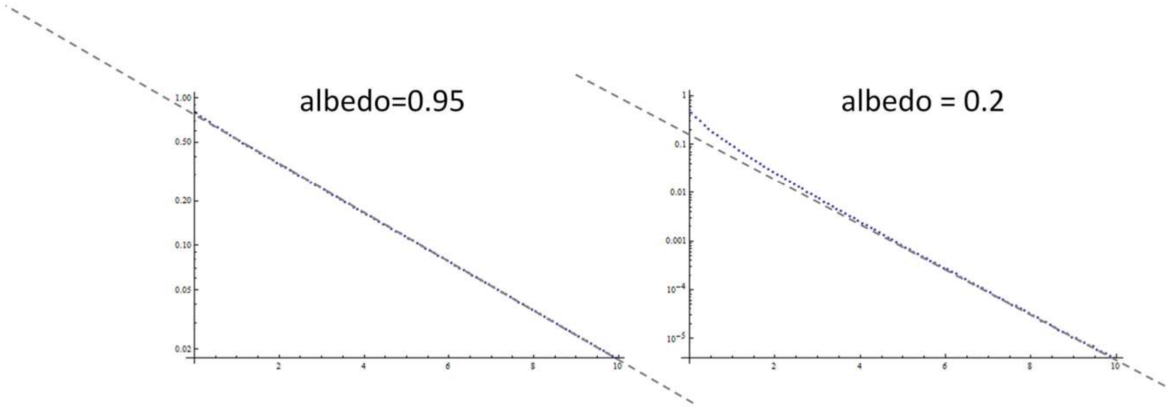
- The zero-variance-based guided subsurface scattering – as described so far – is, to our knowledge, used in most of the production renderer in VFX.

CAN WE DO BETTER?

- How could the above ZV-based path guiding described so far be further improved?

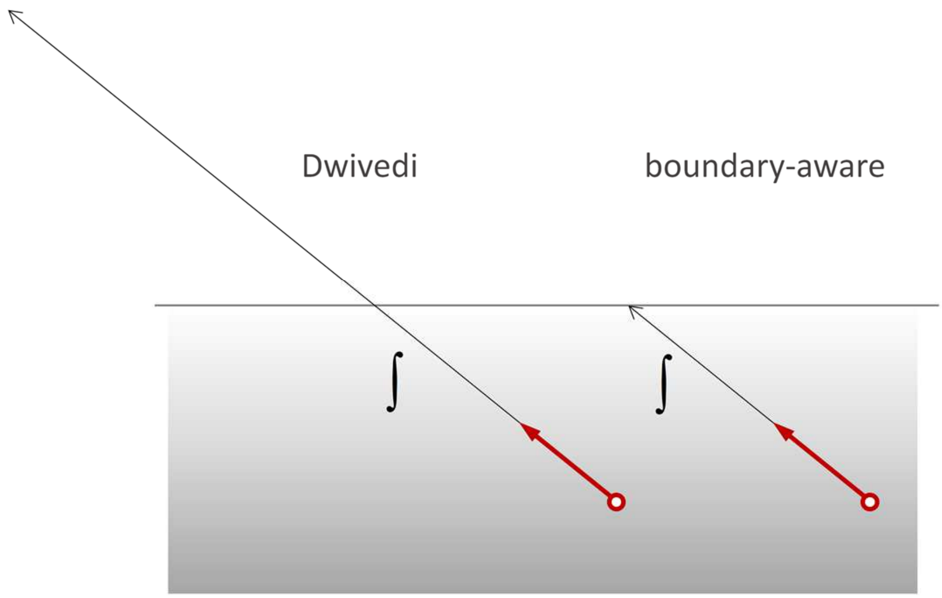
Exponential radiance falloff

- … does not hold for low scattering



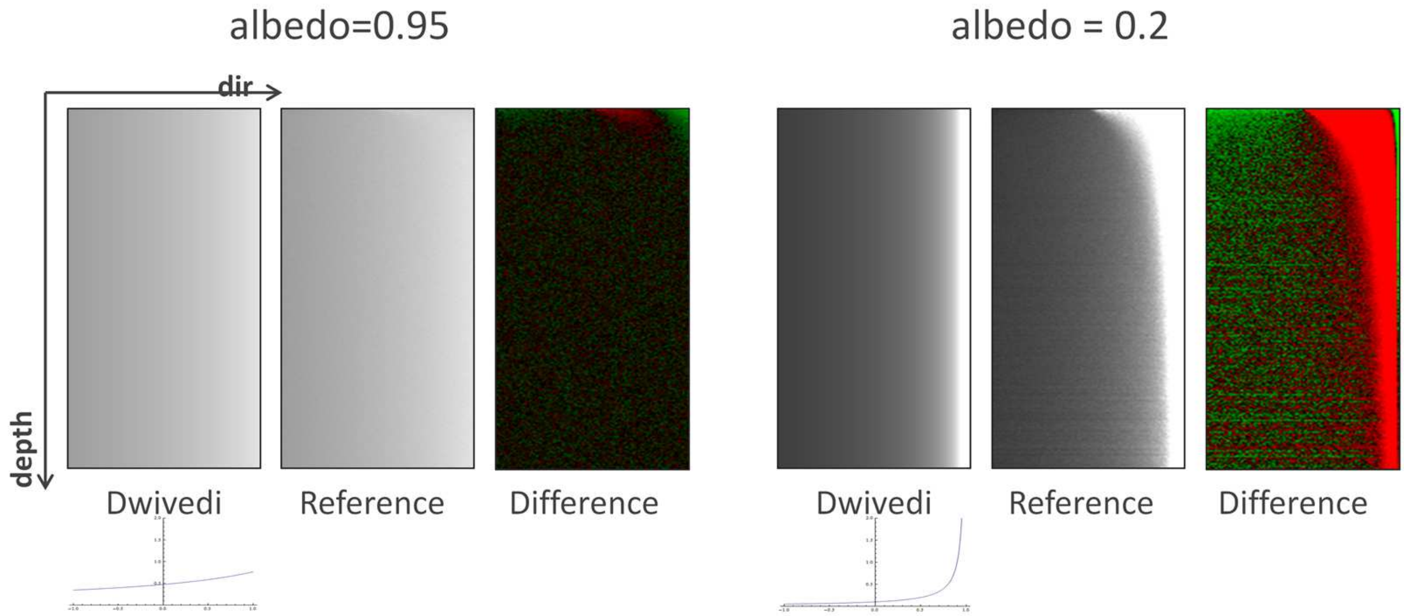
- There are a few reasons for which the Dwivedi sampling scheme still yields a substantial amount of variance, despite being derived from a *zero-variance* scheme.
- First, the assumed solution inside of the medium is a poor approximation when absorption is high.

Directional distribution



- Second, the Dwivedi scheme assumes that the exponential importance solution extends all the way outside of the medium, when in fact, the importance is constant (equal to one) for any point outside of the medium (remember that 'escape' is our only source of importance and it does not matter to us where exactly we escape from the medium).

Radiance distribution



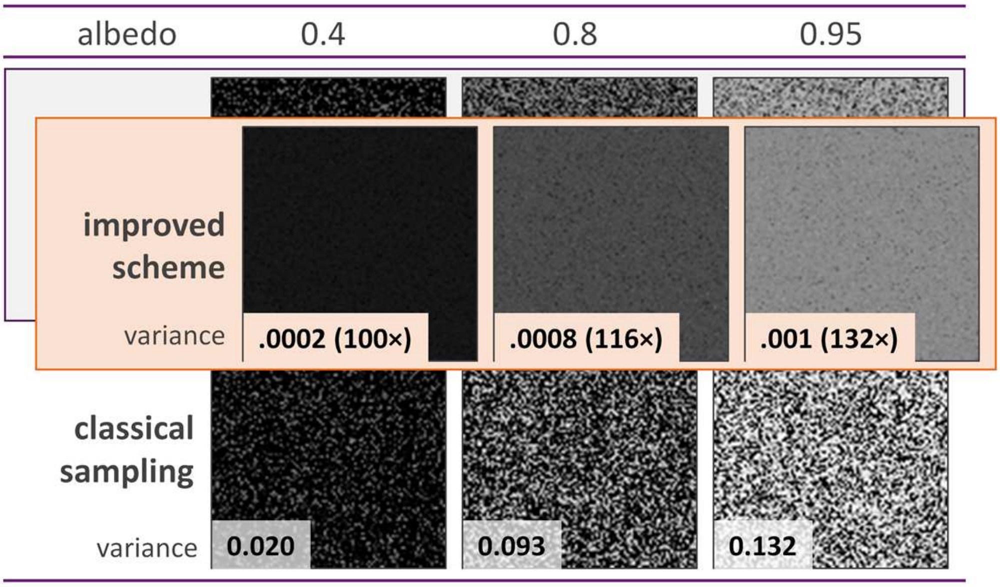
- We can see in these plots that there can be a significant discrepancy between the directional distributions assumed by Dwivedi and those obtained by a reference MC simulation, especially just below the surface. This is a consequence of both considering only the discrete asymptotic mode of the true solution and ignoring the boundary.

Improved sampling

- Take boundary into account
- Better radiance approximation
  - Matching 1^{st} and 2^{nd} moments of the true solution

- In our work on improving the Dwivedi scheme, we build on the paper by Hoogenboom [2008] who describes precisely how to construct a truly zero-variance walk for a half-space that considers the correct probability of escaping from the medium at any given step.
- To approach the zero variance ideal more closely, we employ a better approximation of the solution inside the half-space, based on matching the 1^{st} and 2^{nd} moments of the true solution.
- Another important aspect is that we explicitly take the boundary into account. This has a significant effect on the probability of escaping the medium and on the shape of the angular distribution.

Semi-infinite half-space test



- With this improved scheme, we are able to achieve a variance reduction of two orders of magnitude compared to classical sampling.
- So far, we have only tested the scheme in our synthetic half-space simulator. Application to rendering is an ongoing work.

Future work

- Boundary (Fresnel, rough)
- Anisotropic scattering

- Much work remains to be done.

- First, in our tests, we have assumed index-matched boundary. Changing the roughness or Fresnel ratio at the boundary does actually change the magnitude of the asymptotic term, but not its decay rate. Because the Dwivedi sampling discards the transient terms and renormalize, the magnitude of the asymptotic part goes away, and the Dwivedi scheme is then completely invariant to the boundary conditions.
- However, in our improved approach, changing the boundary condition - and therefore the magnitude of the solution - has a huge effect. This is because in the improved method we strictly separate the solution approximation inside of the medium (an *appropriately scaled* exponential), and outside (constant, unit source of importance). So one can say that Dwivedi's invariance to boundary conditions is only a lucky 'artifact' of some major simplifications in his scheme.

- Another open issue is anisotropic scattering. As the phase function changes away from isotropic, more and more discrete asymptotic terms appear. You can pick the largest one and still apply Dwivedi's scheme with success (we've tried HG with g = 0.75 and still get a nice variance reduction).

- Of course, we still need to apply our improved scheme in rendering.

- And finally, we believe there is much space for exploring other applications of the zero-variance theory in rendering problems.

Conclusion

- Zero variance schemes – solid framework for variance reduction
- Requires (approximate) solution
  - **Learning** from MC samples (Machine/statistical learning techniques)
  - **Analytical approximation**

Thank you

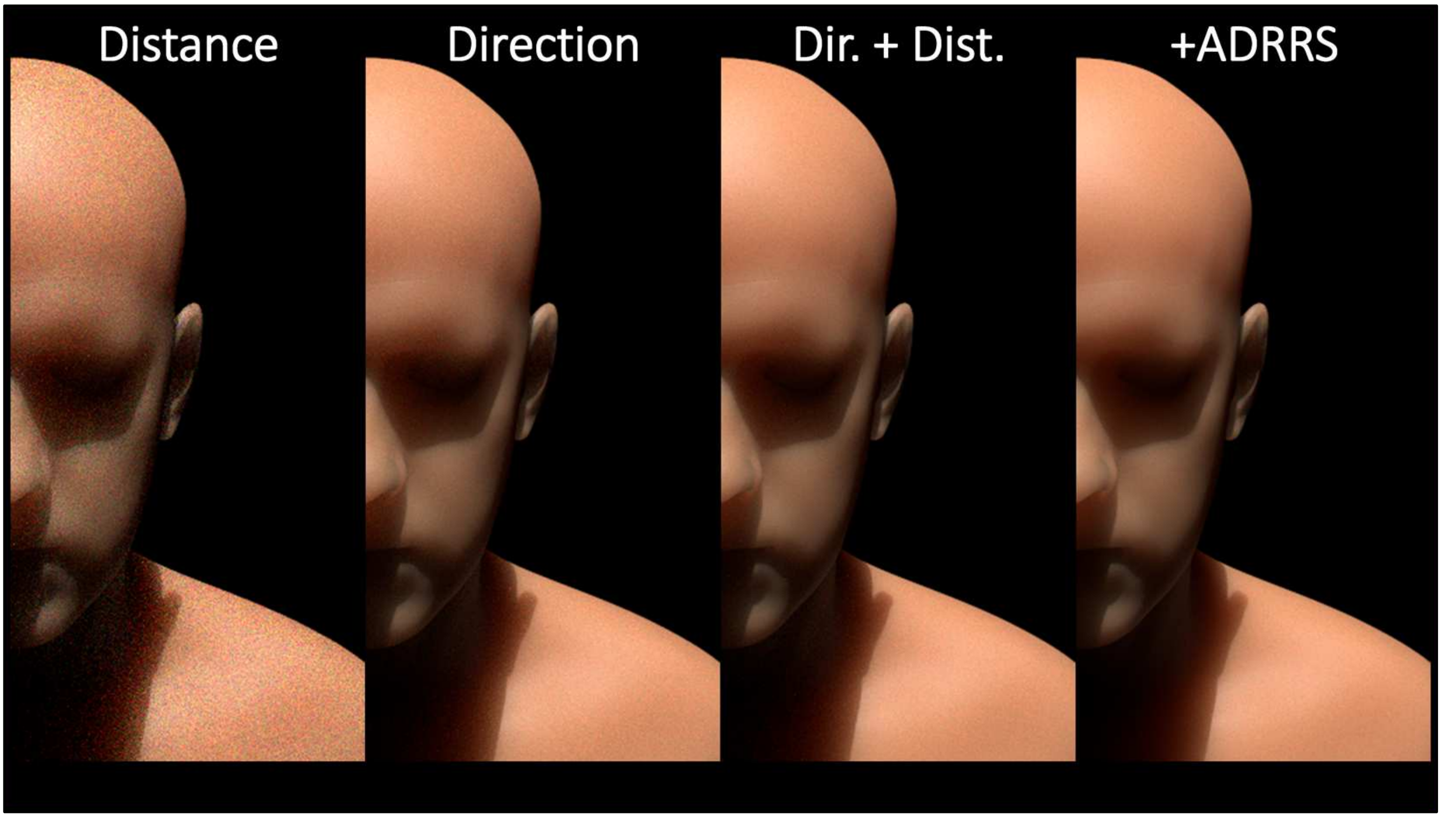
- Acknowledgments
  - Czech Science Foundation (16-18964S)
  - EU ITN No 642841 – "DISTRO"

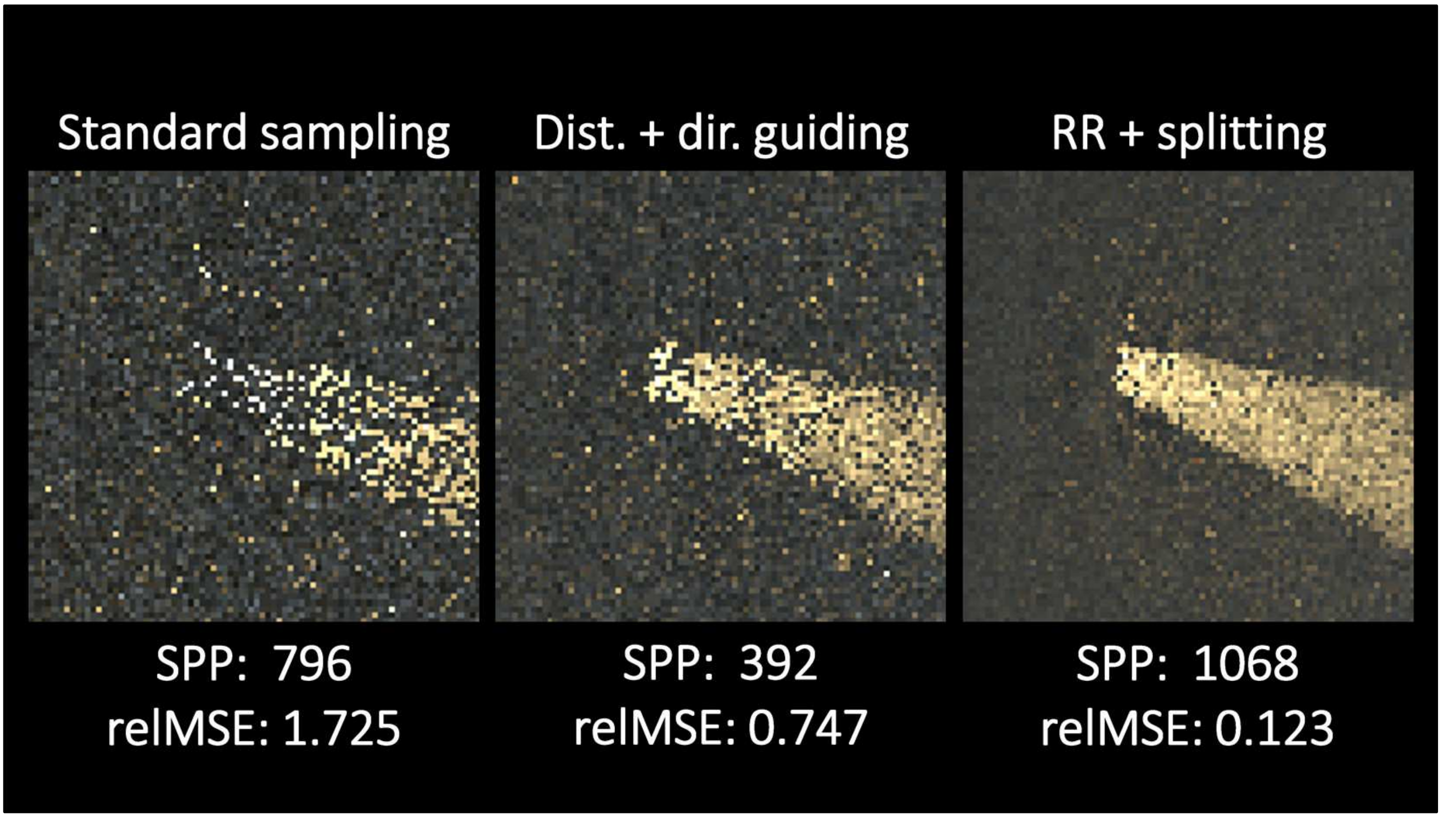
Computer Graphics Charles University

DiSTRO

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 642841.

BACKUP SLIDES


Distance
Direction
Dir. + Dist.
+ADRRS


Standard sampling
Dist. + dir. guiding
RR + splitting
SPP: 796
relMSE: 1.725
SPP: 392
relMSE: 0.747
SPP: 1068
relMSE: 0.123

Application to rendering