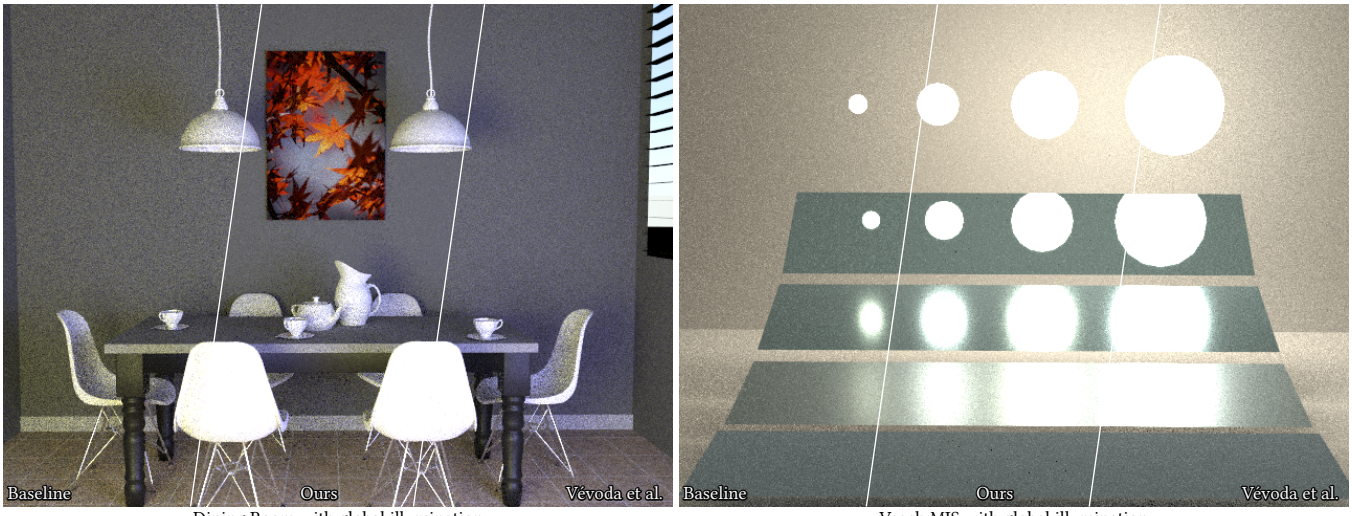
Revisiting controlled mixture sampling for rendering applications

QINGQIN HUA, Saarland University, Germany
PASCAL GRITTMANN, Saarland University, Germany
PHILIPP SLUSALLEK, Saarland University, Germany and DFKI, Germany



Dining Room with global illumination
Speedup: ours 1.90x× vs path tracer, 1.42× vs Vévoda et al.

Veach MIS with global illumination
Speedup: ours 1.48× vs path tracer, 1.41× vs Vévoda et al.

Fig. 1. Equal-time renderings from our main practical application: enhancing light selection with control variates (CVs). Our theoretical findings can be used to enhance prior, empirically motivated CVs [Vévoda et al. 2018], providing strong error guarantees and achieving significant speed-ups.

Monte Carlo rendering makes heavy use of mixture sampling and multiple importance sampling (MIS). Previous work has shown that control variates can be used to make such mixtures more efficient and more robust. However, the existing approaches failed to yield practical applications, chiefly because their underlying theory is based on the unrealistic assumption that a single mixture is optimized for a single integral. This is in stark contrast with rendering reality, where millions of integrals are computed—one per pixel—and each is infinitely recursive. We adapt and extend the theory introduced by previous work to tackle the challenges of real-world rendering applications. We achieve robust mixture sampling and (approximately) optimal MIS weighting for common applications such as light selection, BSDF sampling, and path guiding.

CCS Concepts: • **Computing methodologies** → **Rendering**; **Ray tracing**.

Additional Key Words and Phrases: ray tracing, global illumination, Monte Carlo, multiple importance sampling, control variates

Authors' addresses: Qingqin Hua, Saarland University, Saarland Informatics Campus, Saarbrücken, Germany, hua@cg.uni-saarland.de; Pascal Grittmann, Saarland University, Saarland Informatics Campus, Saarbrücken, Germany, grittmann@cg.uni-saarland.de; Philipp Slusallek, Saarland University, Saarland Informatics Campus, Saarbrücken, Germany and DFKI, Saarland Informatics Campus, Saarbrücken, Germany, philipp.slusallek@dfki.de.

0730-0301/2023/1-ART1
https://doi.org/10.1145/3592435

ACM Reference Format:
Qingqin Hua, Pascal Grittmann, and Philipp Slusallek. 2023. Revisiting controlled mixture sampling for rendering applications. *ACM Trans. Graph.* 1, 1, Article 1 (January 2023), 13 pages. https://doi.org/10.1145/3592435

1 INTRODUCTION

Efficient Monte Carlo rendering requires good importance sampling. Achieving that goal with a single sample distribution is all but impossible, hence combinations through mixture sampling or multiple importance sampling (MIS) [Veach and Guibas 1995b] are the norm.

But many challenges lie on the path to an effective combined algorithm. How should the individual densities be designed [Karlík et al. 2019]? How should the sampling ratios be chosen for a mixture model [Sbert et al. 2019; Pajot et al. 2010; Lu et al. 2013]? How should the budget be allocated for an MIS combination [Grittmann et al. 2022], and how should the MIS weights be chosen [Kondapaneni et al. 2019; Grittmann et al. 2019]?

Practical performance of a mixture or MIS combination can be very poor, if these challenges have not been met. Unfortunately, addressing all these questions is not exactly trivial. As an alternative, previous work has demonstrated that a control variate (CV) can be formed from the sampling techniques in a mixture, and that doing so improves the robustness of said mixture [Owen and Zhou 2000]. Throughout this paper, we refer to that approach as *controlled mixture sampling*. In rendering, that theory has only been applied to simple, low-dimensional problems [Fan et al. 2006]. Kondapaneni et al. [2019] have later shown that optimal MIS weights would

encode this same CV but also did not find practical applications beyond simple direct illumination.

In this paper, we extend the theoretical formulations of controlled mixture sampling to enable practical applications in common rendering scenarios. Specifically, we show how locally optimized, low-dimensional CVs can be used for recursive, infinite dimensional integrals, through spatial caching (Section 4). Then, we discuss general scenarios common in rendering applications, where such CVs can be beneficial (Section 5). We evaluate the practical performance on our main application to many light rendering and also briefly explore the potential in a path guiding setting (Sections 6 and 7).

Fig. 1 shows a sample of our results in the many lights application. There, we consistently outperform baseline path tracing as well as the empirically motivated CV of previous work [Vévoda et al. 2018]. Our rigorous theoretical foundation enables us to reap the same benefits as the previous ad-hoc CV and combine them with the advantages of optimal MIS weighting.

Code and data for this paper are at https://github.com/qingqhua/ControlledMixtureSampling.

2 PREVIOUS WORK

Mixture sampling and MIS are the corner stones of most rendering algorithms. Much work has been done to improve their performance via adaptation to the scene. Control variates (CVs) are orthogonal, and at first glance seemingly unrelated, variance reduction techniques that can also be adapted to the scene for better performance. Previous work has shown that mixtures / MIS and CVs work great in tandem, but have neither identified the scope of potential improvement in rendering, nor practical applications to non-trivial algorithms.

Mixtures and MIS. Importance sampling is essential for efficient Monte Carlo integration and used extensively in rendering applications. As designing a single probability density function (PDF) that works well for all possible cases is all but impossible, mixture sampling and multiple importance sampling (MIS) [Veach and Guibas 1995b] are used extensively. In simple forward path tracing, mixtures are used to sample multi-component BSDFs or to sample points on light sources [Pharr et al. 2016]. These two techniques are then typically combined via MIS. More elaborate algorithms introduce additional mixture components, such as path guiding [Vorba et al. 2019], or MIS techniques, like bidirectional samplers [Veach and Guibas 1995a; Lafortune and Willems 1993; Georgiev et al. 2012a; Hachisuka et al. 2012; Křivánek et al. 2014]. Mixtures and MIS are a crucial and ubiquitous component of any renderer and applying them effectively is an important research problem.

Adaptive PDFs. Light in the real world exhibits a wide and diverse range of fascinating effects. While beautiful to perceive, this complexity is the bane of efficient rendering. Designing PDFs that capture all these effects efficiently is extremely difficult. Therefore, much research has been done on adapting the densities to the scene at hand. Path guiding methods, e.g., construct PDFs based on observed samples to iteratively improve rendering performance for the scene that is currently rendered [Jensen 1995; Bashford-Rogers et al. 2012; Vorba et al. 2014; Müller et al. 2017, 2018; Rath et al. 2020; Hey and Purgathofer 2002; Reibold et al. 2018; Herholz et al. 2016, 2019; Ruppert et al. 2020; Schüßler et al. 2022]. These methods differ in the representations they use, the target densities they strive to learn, and how training and sampling are done. Finding the best solution for each of these components is still an ongoing research problem. Since solely relying on learned PDFs can result in unbounded variance and bias, guiding methods are generally combined with defensive PDFs such as BSDF sampling. That combination is typically done via mixture sampling, as MIS would incur exponential growth when applied repeatedly along a path. Some of these methods additionally use mixture models as their representation for the learned PDF, e.g., Vorba et al. [2014]; Ruppert et al. [2020]; Schüßler et al. [2022].

Adaptive sample counts. The performance of mixture sampling depends greatly on the mixture weights, i.e., the selection probabilities of each component. These, too, can be optimized on-the-fly based on statistics from the scene [Lu et al. 2013; Müller 2019]. Previous work has successfully used the same optimizations for an approximately optimal sample allocation for MIS [Sbert et al. 2019; Grittmann et al. 2022]. These methods closely relate to the problem of adaptively learning light selection probabilities [Vévoda et al. 2018; Donikian et al. 2006]: Sampling a point on a light is typically done through a mixture, where each component covers the area of one light. So, optimizing that probability is a special form of mixture weight optimization.

Adaptive MIS weights. MIS and mixture sampling are closely related in that both combine a set of PDFs. But, where mixture sampling randomly picks one PDF to sample from, MIS always samples from all PDFs and utilizes weighting functions to achieve an unbiased estimate. The deterministic sampling reduces the variance [Veach 1997] and the weighting functions offer another way to further reduce that variance. While heuristic methods can be effective [Veach and Guibas 1995b; Georgiev et al. 2012b; Grittmann et al. 2021] adaptation to the scene has also proven a promising avenue to optimize the MIS weights [Kondapaneni et al. 2019; Grittmann et al. 2019]. Applied to pure MIS, our CVs are equivalent to an optimally weighted MIS estimator [Kondapaneni et al. 2019]. Hence, we can achieve (approximate) optimal MIS weighting in practical applications.

Control variates. Control variates (CVs) are another variance reduction method that can be used on its own or combined with (mixture / multiple) importance sampling. They operate by subtracting a function with known integral from the target integrand and then estimating only the difference. In rendering, applications for CVs include relighting [Rousselle et al. 2016] and direct illumination computations [Vévoda et al. 2018; Clarberg and Akenine-Möller 2008]. These utilize hand-crafted CVs for their specific applications. Adaptive construction of CVs is also possible. Müller et al. [2020] utilize deep learning for that and Salaün et al. [2022]; Crespo et al. [2021] perform regression in primary space to construct an adaptive (piecewise) polynomial CV.

CV and mixtures / MIS. Owen and Zhou [2000] showed that the performance of mixture sampling can be improved by adapting a CV formed by the PDFs. Their insights have been applied to rendering [Fan et al. 2006], but limited to simple, low-dimensional

problems. Kondapaneni et al. [2019] noted that the optimal MIS weights encode the exact same CV, revealing an interesting connection between the two concepts. We extend this theory to allow local optimization of CVs for infinite dimensional recursive integrals.

3 BACKGROUND

Images are rendered by computing a high-dimensional integral in every pixel [Pharr et al. 2016]. For that, Monte Carlo integration is used, i.e., the integrand f is sampled at n random positions x_k and the sample weights are averaged,

$$F = \int_{\mathcal{X}} f(x)\,\mathrm{d}x \approx \langle F \rangle = \frac{1}{n}\sum_{k=1}^{n} \frac{f(x_k)}{p(x_k)}. \tag{1}$$

The error in this estimate manifests as noise; its expected magnitude is given by the *variance*

$$\mathbb{V}\left[\langle F \rangle\right] = \mathbb{E}\left[\langle F \rangle^2\right] - F^2. \tag{2}$$

The variance is low if the probability density function (PDF) $p(x)$ closely matches the integrand.

3.1 Mixtures and MIS

Finding a single PDF that performs well is typically not possible, but multiple PDFs can be combined, either through mixture sampling or via multiple importance sampling (MIS).

3.1.1 Mixture sampling. A mixture PDF $p(x) = \sum_i c_i p_i(x)$ can be sampled by randomly picking a component PDF p_i with probability c_i, resulting in an estimator

$$\langle F \rangle_{\text{mix}} = \frac{1}{n}\sum_{k=1}^{n} \frac{f(x_k)}{\sum_i c_i p_i(x_k)}. \tag{3}$$

3.1.2 Multiple importance sampling. MIS instead always takes n_i samples from each component p_i and uses a *weighting function* $w_i(x)$ to obtain an unbiased estimator [Veach and Guibas 1995b]

$$\langle F \rangle_{\text{MIS}} = \sum_i \sum_{k=1}^{n_i} \frac{w_i(x_{i,k}) f(x_{i,k})}{n_i p_i(x_{i,k})}. \tag{4}$$

A provably good choice for the weighting function $w_i(x)$ is the balance heuristic [Veach and Guibas 1995b],

$$w_i(x) = \frac{n_i p_i(x)}{\sum_j n_j p_j(x)}. \tag{5}$$

MIS with the balance heuristic is closely related to mixture sampling: The estimate

$$\langle F \rangle_{\text{bal}} = \sum_i \sum_{k=1}^{n_i} \frac{f(x_{i,k})}{\sum_j n_j p_j(x_{i,k})} \tag{6}$$

is almost the same as a mixture estimator, except the components are not selected at random. This determinism reduces the variance [Veach 1997] but also mandates that there are at least as many samples as there are PDFs. That limits the utility of MIS in a recursive setting or when the number of PDFs is otherwise large.

3.2 Control variates

The method of control variates (CVs) introduces a correlated function $\lambda(x)$ with a known integral $\Lambda = \int_{\mathcal{X}} \lambda(x)\,\mathrm{d}x$. The original integral is computed by adding the known Λ to the difference integral,

$$F = \Lambda + F - \Lambda = \Lambda + \int_{\mathcal{X}} \left(f(x) - \lambda(x)\right)\,\mathrm{d}x. \tag{7}$$

If $\lambda(x)/p(x)$ closely correlates with $f(x)/p(x)$, then the difference integral will be small. Consequently, the CV estimator

$$\langle F \rangle_{\text{CV}} = \Lambda + \frac{f(x) - \lambda(x)}{p(x)} \tag{8}$$

has low variance. Zero variance is attained with maximum positive correlation, i.e., for any $\lambda(x) = f(x) + Cp(x)$, where $C \in \mathbb{R}$ is an arbitrary constant. The most intuitive optimum lies at $C = 0$, where the CV $\lambda(x) = f(x)$ is the integrand itself.

3.3 Control variates and mixtures

Owen and Zhou [2000] have shown that a CV formed from the components of a mixture PDF can achieve robust improvements over plain mixture sampling. Specifically, they consider the estimator

$$\langle F \rangle_{\text{mixCV}} = \sum_i \alpha_i + \frac{f(x) - \sum_i \alpha_i p_i(x)}{\sum_i c_i p_i(x)} \tag{9}$$

where the CV $\lambda(x) = \sum_i \alpha_i p_i(x)$ is a linear combination of the mixture components with coefficients α_i.

3.3.1 Optimal coefficients. The optimal coefficients minimize the variance,

$$\alpha_{\text{opt}} = \arg\min_{\alpha} \mathbb{V}\left[\langle F \rangle_{\text{mixCV}}\right]. \tag{10}$$

This is a convex optimization problem [Owen and Zhou 2000], so we can solve it by first expanding the variances into covariances,

$$\begin{aligned}\mathbb{V}\left[\langle F \rangle_{\text{mixCV}}\right] &= \mathbb{V}\left[\frac{f(x)}{p(x)}\right] \\ &+ \sum_{i,j} \alpha_i \alpha_j \mathrm{Cov}\left(\frac{p_i(x)}{p(x)}, \frac{p_j(x)}{p(x)}\right) - 2\sum_i \alpha_i \mathrm{Cov}\left(\frac{f(x)}{p(x)}, \frac{p_i(x)}{p(x)}\right),\end{aligned} \tag{11}$$

where we abbreviate notation by substituting $p(x) = \sum_i c_i p_i(x)$. Then, the partial derivatives,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\alpha_i}\mathbb{V}\left[\langle F \rangle_{\text{mixCV}}\right] &= \\ 0 + 2\sum_j \alpha_j \mathrm{Cov}\left(\frac{p_i(x)}{p(x)}, \frac{p_j(x)}{p(x)}\right) &- 2\mathrm{Cov}\left(\frac{f(x)}{p(x)}, \frac{p_i(x)}{p(x)}\right),\end{aligned} \tag{12}$$

are set to zero, yielding the system of equations

$$\begin{aligned}\sum_j \alpha_j \mathrm{Cov}\left(\frac{p_i(x)}{p(x)}, \frac{p_j(x)}{p(x)}\right) &= \mathrm{Cov}\left(\frac{f(x)}{p(x)}, \frac{p_i(x)}{p(x)}\right) \\ \Leftrightarrow \sum_j \alpha_j \left(\int_{\mathcal{X}} \frac{p_i(x)p_j(x)}{p(x)}\,\mathrm{d}x - 1\right) &= \left(\int_{\mathcal{X}} \frac{f(x)p_i(x)}{p(x)}\,\mathrm{d}x - F\right),\end{aligned} \tag{13}$$

where we expanded the covariance terms into their integral definitions, additionally using the fact that PDFs integrate to one. Because the variance is convex but not strictly so, this system has infinitely

many solutions [Kondapaneni et al. 2019]. Specifically, if $\alpha = \{\alpha_i\}$ is optimal, then so is the shifted set

$$\alpha' = \{\alpha_i + sc_i\} \tag{14}$$

for every real number s. This is because, when substituted in the estimator, any of these shifted coefficients simplify to the exact same estimate. The system can be made non-singular, and thus solvable, by imposing an additional constraint. E.g., forcing the coefficients to sum to the integral, $\sum_i \alpha_i = F$, yields

$$\sum_j \alpha_j \int_\mathcal{X} \frac{p_i(x)p_j(x)}{p(x)}\,\mathrm{d}x = \int_\mathcal{X} \frac{p_i(x)f(x)}{p(x)}\,\mathrm{d}x. \tag{15}$$

This can be shown by adding the constraint equation to all equations in the system (13).

3.3.2 Optimal MIS. Kondapaneni et al. [2019] note that the variance of an MIS estimator is convex in the weighting function $w_k(x)$. After computing the optimal weights, they show that these weights encode the exact same CV with a (deterministic) mixture computed by Owen and Zhou [2000]. Their findings allow us to treat MIS and mixture sampling with the same theoretical framework. Further, we can restrict ourselves to the balance heuristic as the weighting function, since we simultaneously achieve optimal MIS weighting.

4 ROBUST AND PRACTICAL CONTROL VARIATES

While the theoretical contributions of Kondapaneni et al. [2019]; Owen and Zhou [2000] are invaluable, their findings cannot be applied directly to non-trivial rendering tasks. In this section, we extend the theoretical framework such that we can optimize CV coefficients in spatial caches. Asides from affording an efficient practical implementation, this formulation also avoids the curse of dimensionality: Each CV is optimized for a local, low-dimensional integration problem: the reflected radiance at a point in the scene.

Conceptually, this is akin to path guiding methods [Vorba et al. 2019]. These subdivide the scene into spatial cells, each approximating the average incident radiance across all points within the same cell. Analogously, we compute CV coefficients that are optimal on average over all points in the same cell. Thereby, we can apply quasi-optimal MIS or CVs to high-dimensional global illumination rendering while only optimizing local, low-dimensional problems. Our theory, introduced in the following, ensures that spatial cells with incompatible data (e.g., boundaries) regress gracefully to the baseline; i.e., simple mixture sampling or balance heuristic MIS.

We start by generalizing previous theory to the common scenario where both MIS and mixture sampling are used simultaneously. Then, we extend the derivations to enable CV coefficients shared across many integrals (i.e., across a spatial cell). Finally, we show that local CV coefficients can be optimized such that they minimize the global image error.

4.1 Setup

Monte Carlo estimators in rendering applications commonly utilize a large number of sampling techniques, combined through both mixture sampling and MIS. A simple forward path tracer, e.g., mixes the per-component PDFs of the BSDF and uses MIS to combine the result with a mixture of per-light area sampling PDFs.

Formally, we consider a combination of M MIS techniques, each taking n_m samples from a mixture of T_m PDFs. That is, the effective density of the estimator is

$$p(x) = \sum_{m=1}^{M} n_m \sum_{t=1}^{T_m} c_{m,t} p_{m,t}(x), \tag{16}$$

where $c_{m,t}$ is the selection probability of the tth mixture component in the mth MIS technique.

Our goal is to improve the efficiency of that estimator via a CV

$$\langle F \rangle_{\mathrm{CV}} = \sum_i \alpha_i + \sum_{m=1}^{M} \sum_{k=1}^{n_m} \frac{f(x_{m,k}) - \sum_i \alpha_i \lambda_i(x_{m,k})}{p(x_{m,k})}, \tag{17}$$

where the λ_i are linear combinations of the sampling PDFs. Concretely, each

$$\lambda_i(x) = \sum_{k=1}^{N_i} \frac{n_k c_k}{\sum_{k'=1}^{N_i} n_{k'} c_{k'}} p_k(x) \tag{18}$$

is a mixture of N_i PDFs such that $\int \lambda_i(x)\,\mathrm{d}x = 1$. The mixture weights are proportional to the effective sample count of each PDF. This ensures *robustness*, because it implies that a set of non-zero coefficients exists such that

$$\frac{\sum_i \alpha_i \lambda_i(x)}{p(x)} = \sum_i \alpha_i, \tag{19}$$

i.e., the CV can cancel out. Thus, the original estimator can arise as the optimization outcome, guaranteeing that we will always improve upon it.

The setup of Owen and Zhou [2000] arises as a special case where only mixture sampling is used and $\lambda_i = p_i$. The setup of Kondapaneni et al. [2019] arises as a special case where the CV is formed from only the MIS techniques, i.e.,

$$\lambda_i(x) = \sum_{t=1}^{T_i} c_{i,t} p_{i,t}(x). \tag{20}$$

We extend these to allow other partitionings of the PDFs, like clustering of light sources, or grouping BSDF components into upper- and lower-hemisphere densities.

Independently of the setup, the optimal coefficients can be found by solving the linear system

$$\sum_j a_j \int \frac{\lambda_j(x)\lambda_i(x)}{p(x)}\,\mathrm{d}x = \int \frac{f(x)\lambda_i(x)}{p(x)}\,\mathrm{d}x. \tag{21}$$

This can be derived by following the same steps outlined in Section 3.3.1.

4.2 Optimal shared control variate

Computing the CV coefficients with the formulation discussed so far requires tracking multiple samples *for every integral*. Unfortunately, rendering applications compute (infinitely) many integrals. In the image-space setting considered by previous work [Kondapaneni et al. 2019; Fan et al. 2006; Salaün et al. 2022], we would have to compute one integral per pixel and per path length. In the spatial setting we aspire to, there is one integral for every point and outgoing direction.

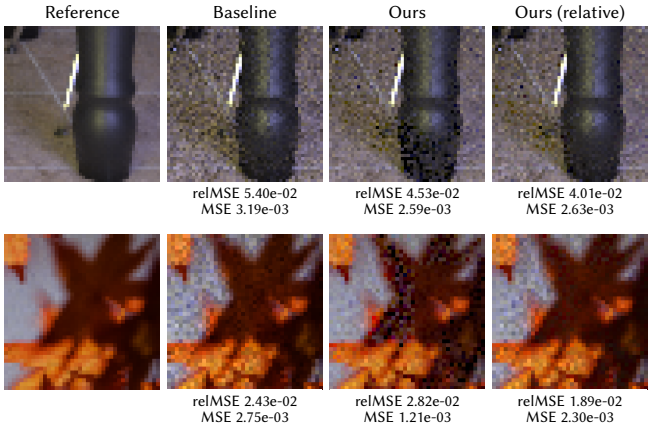


Fig. 2. Optimizing the CV for absolute variance ('Ours') produces visible artifacts if a spatial cell spans a high-contrast boundary. Optimizing for the relative variance ('Ours (relative)') resolves the issue. Note how optimizing the relative variance increases the mean squared error (MSE) but reduces the relative MSE (relMSE) of the image.

We overcome this challenge by optimizing the average variance. This seemingly simple step bears intricate challenges: not only does the integrand f vary but so may the PDFs—for instance, if the surface roughness changes between integrals. A spatial cell could, e.g., contain a textured surface with varying geometry. The points in that cell will then have different BSDF values but also different BSDF-based importance sampling PDFs. Mathematically, we consider a set of estimators $\mathcal{F} = \{\langle F_k \rangle\}$ each computing a different integral F_k with the same number of PDFs, but potentially different definitions or mixture weights.

A single set of coefficients can be shared across all these integrals. For that, we find the coefficients that minimize the sum of variances of all estimators,

$$\arg\min_{\alpha} \sum_k \mathbb{V}\left[\sum_{m=1}^{M}\sum_{i=1}^{n_{k,m}} \frac{f_k(x_{k,m,i}) - \sum_j \alpha_j \lambda_{k,j}(x_{k,m,i})}{p_k(x_{k,m,i})}\right]. \quad (22)$$

If the set $\mathcal{F}$ is continuous (e.g., for a spatial cache), the sum is replaced by an integral.

Fortunately, this averaging does not change the nature of the optimization problem but merely introduces an additional dimension that we sum (or integrate) over. Assuming that the $\langle F_k \rangle$ are mutually independent, we can apply similar steps as before to find the optimal coefficients

$$\sum_j a_j \sum_k \int \frac{\lambda_{k,j}(x)\lambda_{k,i}(x)}{p_k(x)}\,\mathrm{d}x = \sum_k \int \frac{f_k(x)\lambda_{k,i}(x)}{p_k(x)}\,\mathrm{d}x. \quad (23)$$

This result is convenient for practical use, as it means we can simply accumulate the estimates across all integrals.

While no longer optimal for every single integral F_k, these coefficients are optimal on average. Similarly, robust improvements over the plain MC estimator are no longer guaranteed per-integral, but they still hold on average. In practice, this means that care has to be taken when deciding what integrals to average over.

4.2.1 Relative variance. Averaging over incompatible integrals can be problematic. In practice, we found that the main challenge are high-contrast regions: if a cache contains very bright and very dark parts, the bright ones will dominate the coefficients, to the severe detriment of the darker ones.

A remedy is to optimize the sum of relative variances, i.e.,

$$\arg\min_{\alpha} \sum_k \frac{\mathbb{V}\left[\langle F_k \rangle_{\mathrm{CV}}\right]}{F_k^2}. \quad (24)$$

That way, brightness changes due to texture or illumination are no longer problematic. This trick has been successfully used in prior work [Rath et al. 2020, 2022; Grittmann et al. 2022]. The main obstacle is obtaining a surrogate $\tilde{F}_k \approx F_k$ for the ground truth. We follow previous work and use a denoised image from the first sample in each pixel.

Since F_k is a constant, our optimization is almost unchanged,

$$\sum_j a_j \sum_k \int \frac{\lambda_{k,j}(x)\lambda_{k,i}(x)}{\tilde{F}_k^2 p_k(x)}\,\mathrm{d}x = \sum_k \int \frac{f_k(x)\lambda_{k,i}(x)}{\tilde{F}_k^2 p_k(x)}\,\mathrm{d}x, \quad (25)$$

with the small yet crucial difference that each sample is divided by the surrogate ground truth.

Fig. 2 demonstrates the benefit of optimizing for the relative variance on an example. In this scene, some spatial cells contain high-contrast boundaries due to texture (bottom row) or illumination (top row) changes. When optimizing for the average variance, the CV coefficients almost exclusively favor the bright region, as it dominates the variance. Consequently, perceived error in the darker pixels increases. Optimizing for the relative variance distorts the optimization in favor of the darker pixels, resolving the issue.

4.3 Color

In an RGB renderer, each pixel is composed of three integrals—one for every color channel. In that case, we can optimize a separate CV coefficient for each color to reap variance reduction if color varies strongly, as done by Kondapaneni et al. [2019]. However, in a spectral renderer, this is no longer the case. There, infinitely many coefficients would have to be computed—one for every wavelength. Also, it may be desirable to reduce the overhead by using a single, monochromatic coefficient even in an RGB renderer.

For that, we can apply the same reasoning as for the spatial sharing. By simply averaging the estimates for our linear system, we can easily obtain coefficients that are optimal across all colors, or coefficients that are optimal for one frequency band.

4.4 Recursive integration

Previous work has considered the idealized case of a CV formed from the full PDFs $p(x)$. In rendering applications, the integrals are recursive and sampling is done incrementally. In other words, $p(x)$ is a high-dimensional quantity. Optimizing a CV in a high-dimensional space is quickly limited by the curse of dimensionality [Salaün et al. 2022]. We tackle this problem by optimizing local, low-dimensional CVs instead.

Viewed abstractly, rendering algorithms compute integrals

$$\int f(\overline{\mathbf{x}})\,\mathrm{d}\overline{\mathbf{x}} = \int g(x) \int h(x,y) \int i(y,z)\,\mathrm{d}z\,\mathrm{d}y\,\mathrm{d}x \quad (26)$$

that can be separated into a prefix x, a local point y, and a suffix z. Namely, the path leading to the point we consider (x), the direction at that point (y), and the path continuing in that direction to a light source (z). We consider an estimator

$$\langle F \rangle = \frac{g(x)}{p(x)} \left(\sum_t \alpha_t + \sum_{m=1}^{M} \sum_{k=1}^{n_m} \left(\frac{h(x,y)i(y,z)}{p(y|x)p(z|y)} - \frac{\sum_t \alpha_t \lambda_t(y|x)}{p(y|x)} \right) \right) \tag{27}$$

that forms a local CV with coefficients optimized over all prefixes. We show that the nested sampling can be ignored while the prefix can be accounted for easily.

Repeating similar steps as before—i.e., expanding the variance into covariances, taking the derivatives, and setting them to zero—we obtain the linear system

$$\begin{aligned} &\sum_j \alpha_j \left(\int \frac{g^2(x)}{p(x)} \frac{\lambda_i(y|x)\lambda_j(y|x)}{p(y|x)} - \int \frac{g^2(x)}{p(x)} \,\mathrm{d}x \right) = \\ &\int \frac{g^2(x)}{p(x)} \frac{h(x,y)I(y)\lambda_i(y|x)}{p(y|x)} - \int \frac{g^2(x)}{p(x)} h(x,y)I(y) \end{aligned} \tag{28}$$

We can simplify this solution further by shifting the system to compute a different optimum along the line of infinitely many optima,

$$\begin{aligned} &\sum_j \alpha_j \int \frac{g^2(x)}{p(x)} \int \frac{\lambda_i(y|x)\lambda_j(y|x)}{p(y|x)} \,\mathrm{d}y \,\mathrm{d}x \\ &= \int \frac{g^2(x)}{p(x)} \int \frac{\lambda_i(y|x)h(x,y)I(y)}{p(y|x)} \,\mathrm{d}y \,\mathrm{d}x. \end{aligned} \tag{29}$$

Appendix A proves that this system yields optimal coefficients. Note that the variance of the nested estimator is irrelevant as only the expectation $I(y) = \int i(z|y)\,\mathrm{d}z$ affects the result. This is not true, however, for the prefix sampling $p(x)$. Fortunately, in practice, the latter can be accounted for by simply multiplying the squared throughput weight onto each sample.

5 VARIANCE-REDUCTION POTENTIAL

Optimized PDF-based CVs have been used in rendering before [Kondapaneni et al. 2019; Fan et al. 2006]. Though these prior works have reported improvements, it is not clear when, how, and why exactly the CVs are beneficial. In this section, we present and discuss rendering-specific scenarios where CVs offer improvements.

The rendering results in the following are produced by training a path guiding distribution [Herholz and Dittebrandt 2022] with $100k$ samples and computing CV coefficients from the same set of samples. The CV is formed from the individual mixture components of the path guiding distribution and the BSDF. The images are rendered by mixing path guiding with 50% BSDF sampling. In all cases, the scene is a diffuse plane illuminated by an HDR image. Results in more realistic settings are discussed in Section 7.

5.1 Color noise

An advantage of CVs over importance sampling is that the former can handle color. Consider the setup shown in Fig. 3. The visible portion of the environment map has roughly constant brightness, yet the color varies between blue and green. Importance sampling cannot eliminate this variation in color—if, e.g., a sample is taken from the blue region, then only the blue contribution can be estimated. If

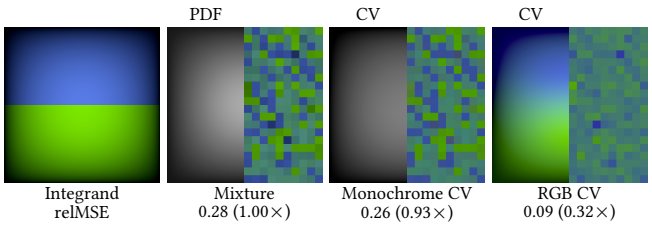


Fig. 3. Controlled mixture sampling can reduce color noise. Here, a diffuse plane is illuminated by an HDR with strong color differences, as shown on the left. A mixture-based path guiding method and BSDF sampling are combined to estimate the irradiance. The plain mixture has strong color noise; applying a monochromatic CV provides minor noise reduction, and a per-color CV can reduce the error threefold.

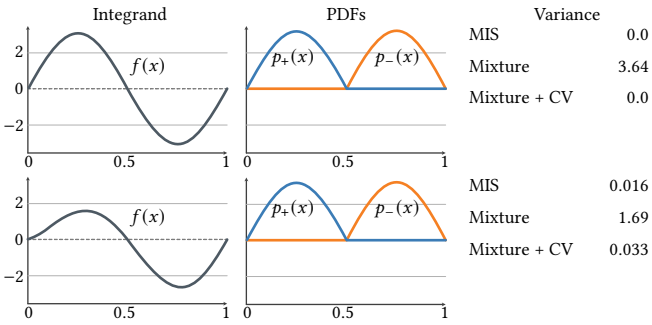


Fig. 4. 1D examples for signed integration problems. In an idealized case (first row) using perfect PDFs, both, MIS and mixture sampling with CVs yield zero variance. If the integrand is not matched perfectly (bottom row), mixture sampling with a CV can still achieve almost the same variance as MIS.

the CV coefficients are optimized per color channel, contribution from all colors is always included in the estimate. In this extreme case, variance is reduced threefold in that case. In contrast, if a monochrome CV is employed, only minor speed-ups of 7% can be achieved.

5.2 Positivisation

Most integrands in rendering are positive, though there are exceptions; e.g., differentiable rendering. Via pure importance sampling, it is impossible to remove the variance due to the sign changes. As a remedy, antithetic sampling [Zeltner et al. 2021] or positivisation [Owen and Zhou 2000] can be used. That is, instead of taking a single sample, two (or more) samples are taken: one from the positive part $p_+(x)$ and one from the negative part $p_-(x)$. Fig. 4 illustrates this concept on a simple sine curve.

Positivisation / antithetic sampling can be thought of as forming an MIS estimator from the positive and negative parts, instead of a simple mixture PDF [Owen and Zhou 2000]. Therefore, this scheme shares the same drawback as MIS: recursive application is expensive. Previous work [Zeltner et al. 2021] have ameliorated the issue through path replay [Vicini et al. 2021]. That makes the overhead linear in the path length, but does not remove it entirely and comes with its own challenges.

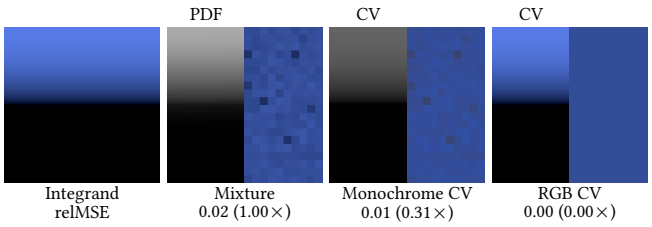


Fig. 5. CVs can restore the zero-variance property of a mixture component, even if that component is only rarely sampled. Here, the uniform sky is sampled perfectly by BSDF sampling and mixing in guiding increases the variance. A monochrome CV eases the problem by always matching the correct average brightness, while a full-color CV achieves zero variance.

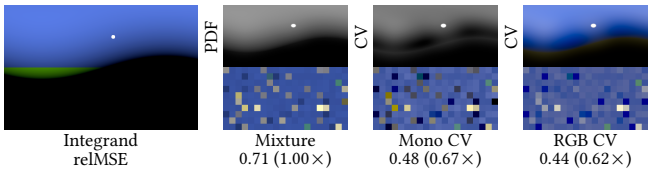


Fig. 6. CVs can combat oversampling. Here, the darker sky is oversampled by the combination of guiding and BSDF sampling. While the sky is captured perfectly, there is significant variance due to the sun in the plain mixture result. The CV rectifies this by always including contributions from the sun, as reconstructed through the path guiding PDFs.

Optimized CVs could offer an alternative here. Akin to the reduction of color noise, the CV always includes contribution from both, the negative and positive components. In the simplified example of a sine curve, we thus achieve zero-variance with a single sample. Testing the utility of this approach in, e.g., a differentiable renderer is left for future work.

5.3 Low-variance and oversampling

Mixture and MIS combinations exhibit two well-known issues: they perform poorly when one PDF is (almost) optimal [Veach and Guibas 1995b], and they often oversample unimportant regions [Karlík et al. 2019]. Including a CV can resolve or ease both problems.

5.3.1 Low-variance. Many applications combine a (close to) optimal PDF with a suboptimal one. Often, this is due to defensive sampling: a learned PDF may sometimes be a perfect match, but to avoid unbounded variance and bias, it is always combined with a defensive density. For instance, path guiding is always mixed with defensive BSDF sampling, even if the guiding approach strives to learn the full product with the BSDF [Herholz et al. 2016].

It is well-known that such a combination sacrifices some of the benefits of the low-variance technique. When combined through MIS, it is possible to restore the zero-variance outcome through well-chosen MIS weights [Veach and Guibas 1995b; Grittmann et al. 2019; Kondapaneni et al. 2019]. This is not the case for mixture sampling. There, the optimal technique might sometimes not be sampled at all.

For mixture sampling, Owen and Zhou [2000] address the problem by optimizing CVs based on the sampling PDFs. They proved that the variance of the resulting estimator is at most σ_t^2/c_t, where t is the best technique and c_t and σ_t^2 are its selection probability and variance, respectively. As our theory is based on theirs, we also reap those benefits.

Fig. 5 demonstrates this on a simplified example. BSDF sampling has zero variance in this case, but the mixture with path guiding considerably worsens the result. The monochrome CV reduces this error threefold, but, as it returns a grayscale approximation of the integral for poorly sampled pixels, it cannot achieve zero variance. Switching to a per-color CV enables us to perfectly fit the CV to the integrand, hence yielding zero variance despite the fact that we are using mixture sampling.

5.3.2 Oversampling. Oversampling occurs when multiple PDFs strive to match the entire integrand, but most only succeed in the low-contribution regions. A common example is direct lighting from a sun-and-sky model, as shown in Fig. 6. There, BSDF sampling almost perfectly matches the sky but neglects the bright sun. At the same time, guiding achieves a crude match of the entire integrand. When combined, the darker sky will receive far too many samples.

Previous work has proposed remedies for tabulated PDFs [Karlík et al. 2019] and iterative refinement schemes [Rath et al. 2020]. CVs can additionally ease the oversampling problem, with or without these enhancements.

In the example shown in Fig. 6, the optimized CV adds approximated contributions from the undersampled regions (sun and bright parts of the sky) to every sample. At the same time, it subtracts contribution form the oversampled darker parts of the sky. The net result is a 40% faster rendering. Here, variance is dominated by brightness differences due to the sun, so monochromatic coefficients can also provide significant speed-ups.

6 IMPLEMENTATION

We assess the practical utility of our formulation by implementing it in a guided path tracer with next event estimation. We use the Intel® Open Path Guiding Library [Herholz and Dittebrandt 2022] which is based on the guiding approach of Ruppert et al. [2020]. To ease comparison, we pre-train the guiding distribution once and use the same one for all methods.

The guiding PDF is mixed with a BSDF PDF using a mixture weight of 0.5. Both of these PDFs are themselves mixtures. Direct lighting is computed through an MIS combination of this mixture with next event estimation. For the latter, we uniformly select one light source and uniformly sample its area. If an environment map is present, an additional sample is taken for it, using an MIS compensated [Karlík et al. 2019] PDF.

To aid comparability, we disabled path guiding for our main results in Section 7.2. The exploratory application to guiding (Section 7.5), in turn, is done without next event estimation and hence without the per-light control variate. We include statistics for all techniques enabled (with guiding and per-light control variate) in supplemental.

6.1 Optimizing the CV

The CV is trained by first rendering 8 samples per pixel with plain mixture sampling / MIS. During this training phase, we accumulate Monte Carlo estimates of the coefficients in the linear system we need to solve. After this training, the systems are solved and the remaining iterations are rendered using the optimized CV. We average the training image with the subsequent iterations.

To compute the approximate pixel value needed to optimize the relative error, we denoise the image rendered by the first sample per pixel. The overhead of this denoising (around 100ms with our denoiser) is insignificant at our render times (60s) and can be shared with other methods that need the same surrogate reference, like improved target functions [Rath et al. 2020] or Russian roulette [Rath et al. 2022]. Further, in a progressive rendering setting, repeated denoising may be performed anyway [Firmino et al. 2022]. Thus, we do not include the denoising cost in the overhead of our method.

The coefficients for our CV are stored in a simple (sparse) regular grid. We manually specify the resolution for each scene. We chose that approach for simplicity; we recommend that practical applications use adaptive subdivision schemes like those common in path guiding methods.

6.2 Handling per-light PDFs

The number of PDFs in our application is dominated by the number of light sources. A simple means to reduce that number is clustering. But even then, there may be thousands of clusters. Naïvely computing and solving our system for this many PDFs is very costly. Therefore, we derived a specialized solver for this case.

For that, we note that the per-light PDFs do not overlap each other, as each one only samples one light. Hence, the inner products

$$\int \frac{p_i(x)p_j(x)}{p(x)}\,\mathrm{d}x = 0 \tag{30}$$

are zero for all $i \neq j$. If we only used these non-overlapping PDFs, the system would be diagonal,

$$\alpha_i \int \frac{p_i^2(x)}{p(x)}\,\mathrm{d}x = \int \frac{p_i(x)f(x)}{p(x)}\,\mathrm{d}x. \tag{31}$$

But what if additional PDFs are present? In that case, we can still exploit the non-overlapping nature of the many per-light PDFs. Consider the setup where the CV is computed from n non-overlapping and m overlapping techniques. We can substitute the first n rows of the linear system—the ones corresponding to the non-overlapping techniques—into the remaining m rows. That yields an $m\times m$ system for the coefficients of the overlapping techniques:

$$\sum_{j=1}^{m} \alpha_{n+j}\left(A_{n+i,n+j} - \sum_{k=1}^{n}\frac{A_{i,n+k}A_{k,n+j}}{A_{k,k}}\right) = B_{n+i} - \sum_{k=1}^{n}\frac{A_{i,n+k}B_k}{A_{k,k}}, \tag{32}$$

where $A_{i,j}$ and B_i denote the matrix and vector elements, respectively, of the ith row and jth column in the original system. Substituting the result into the first n equations provides us with the remaining coefficients,

$$\alpha_i = \frac{B_i - \sum_{k=1}^{m} A_{i,n+k}\alpha_{n+k}}{A_{i,i}}. \tag{33}$$

The asymptotic cost of these operations is linear in the number of lights / clusters; both, in terms of computation effort and in terms of memory consumption.

7 EVALUATION

In this section, we discuss the practical performance of our method applied to direct lighting computation in path tracing. We show that considerable speed-ups can be had for full global illumination, where the theory of Kondapaneni et al. [2019]; Owen and Zhou [2000] is not applicable. Then, we show the benefits of our rigorous theory over the previous, empirically motivated approach of Vévoda et al. [2018]. We conclude with a discussion of the main weaknesses of any learned CV approach: the overhead and the error in the coefficient estimation.

7.1 Global illumination

An important benefit of our formulation over the original theory [Owen and Zhou 2000; Kondapaneni et al. 2019] is that we side-step the curse of dimensionality through local CV optimization. Hence, speed-ups can be achieved for arbitrary path lengths and full global illumination without an excessive growth in overhead.

Fig. 7 demonstrates this by comparing our results to baseline path tracing. Each column in the figure shows the contribution of one path length; the leftmost column shows direct lighting only, the rightmost shows only thrice-bounced indirect illumination. While the speed-ups are greatest for direct lighting, considerable improvements can be achieved even for long indirect paths.

7.2 Comparison to Vévoda et al. [2018]

To gauge the benefit of our rigorous theory, we compare its practical results to the ad-hoc CV of Vévoda et al. [2018]. Both methods use the same data structure, rendering process, and number of training samples. The CV of Vévoda et al. [2018] is formed by only the light sampling techniques, and the coefficient of the ith light / cluster PDF is the MIS weighted contribution of that cluster to the spatial cell.

Vévoda et al. [2018] empirically found this CV to perform well but did not find theoretical guarantees for its performance. Our theory can yield these: the coefficients they used arise as a special case of our optimization when applied only to the MIS weighted integral of the next event estimation technique. This proves that their ad-hoc formulation is never worse than the baseline. However, further improvements can be had whenever BSDF sampling is beneficial.

Fig. 8 shows this on three representative cases. Despite the higher overhead of our method, we achieve equal or better results than the simpler formulation of Vévoda et al. [2018] in all scenes.

The benefit is greatest in the Dining Room scene. It is illuminated by multiple large area lights and an HDR background, hence BSDF sampling performs well. Including it in the CV can triple the performance in some parts of the scene, compared to Vévoda et al. [2018].

The RGB Sofa features a scenario where CVs are extremely worthwhile: The RGB lamp in the corner emits many different colors but has constant brightness. Since the surfaces are mostly diffuse and the lights quite small, our method does not yield improvements over

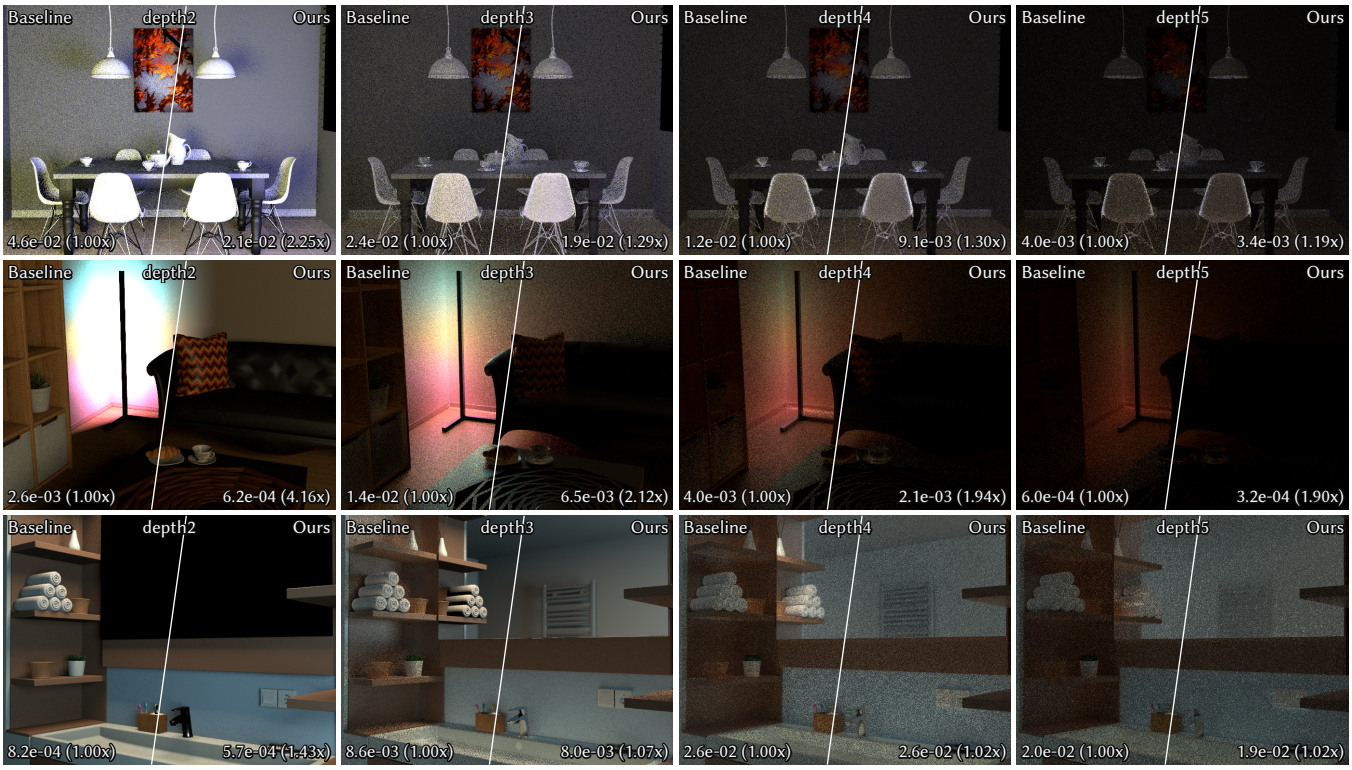


Fig. 7. Equal-time (60s) comparison per path length between ours and baseline path tracing. The numbers in parentheses are the speed-up compared to the baseline, higher is better. We achieve consistent speed-ups for both, direct and indirect illumination. The benefit of our method is greatest if there is much color variation (second scene) or large light sources (first scene).

Vévoda et al. [2018], but both achieve twice as fast renderings as the baseline by reducing the color noise.

Finally, the Bathroom is a case where CVs do not offer much benefit. The scene is illuminated by a few medium-sized area lights and features many glossy surfaces. As shown in Fig. 7, both CV methods can achieve small improvements in the direct illumination component, but little in the indirect illumination. But, importantly, due to the robustness guarantees of our theory, results are not worse than the baseline.

7.3 Overhead and estimation error

Our theory guarantees robust and consistent improvements over the baseline. Unfortunately, to truly ensure these guarantees in practice, we would need perfect estimates of all required quantities. And we would need to be able to obtain those at no extra cost. Naturally, that is not the case in reality.

An extreme case is our variant of Veach's iconic MIS test scene. It features strong glossy highlights due to around 10k emissive triangles forming the illuminating spheres. The combination of both results in high overhead and high estimation error in the coefficients. Both, our method and the previous ad-hoc CV fare worse than the baseline in this challenging scenario.

7.3.1 Overhead. Table 1 compares the equal-sample and equal-time performance of our method to assess the impact of the overhead for the per-light control variate application. In theory, further speed-ups of our method are possible. For instance, the RGB Sofa and Bathroom scene yields 10% lower error in equal-sample comparisons than the CV of Vévoda et al. [2018]. In our current implementation, that gain is offset entirely by the overhead. Hence, reducing the cost of our method is a worthwhile endeavor.

The overhead is due to three main sources. First, the estimates for the linear systems must be accumulated. The exact cost of that depends on the number of occupied spatial cells, and other scene-dependent factors. On average, it increased render times by around 5% in our tests. Second, the linear system in each spatial cell must be solved. The cost of that again depends on the spatial resolution but also on the number of PDFs. We found this time to be negligible, ranging from 1ms to 50ms in our tests. Third, to obtain the PDFs of zero-contribution paths, additional ray tracing operations are required. If a BSDF sample yields zero contribution, we must still determine what point on the light this direction intersects to compute all PDFs. In our scenes, this caused around 10% more ray tracing operations.

7.3.2 Estimation error. Asides from the overhead, the practical benefits are also limited by the estimation quality of the CV coefficients. Each coefficient is computed by accumulating a large number of Monte Carlo estimates and then solving a linear system based on those. High variance in these estimates can result in coefficients that

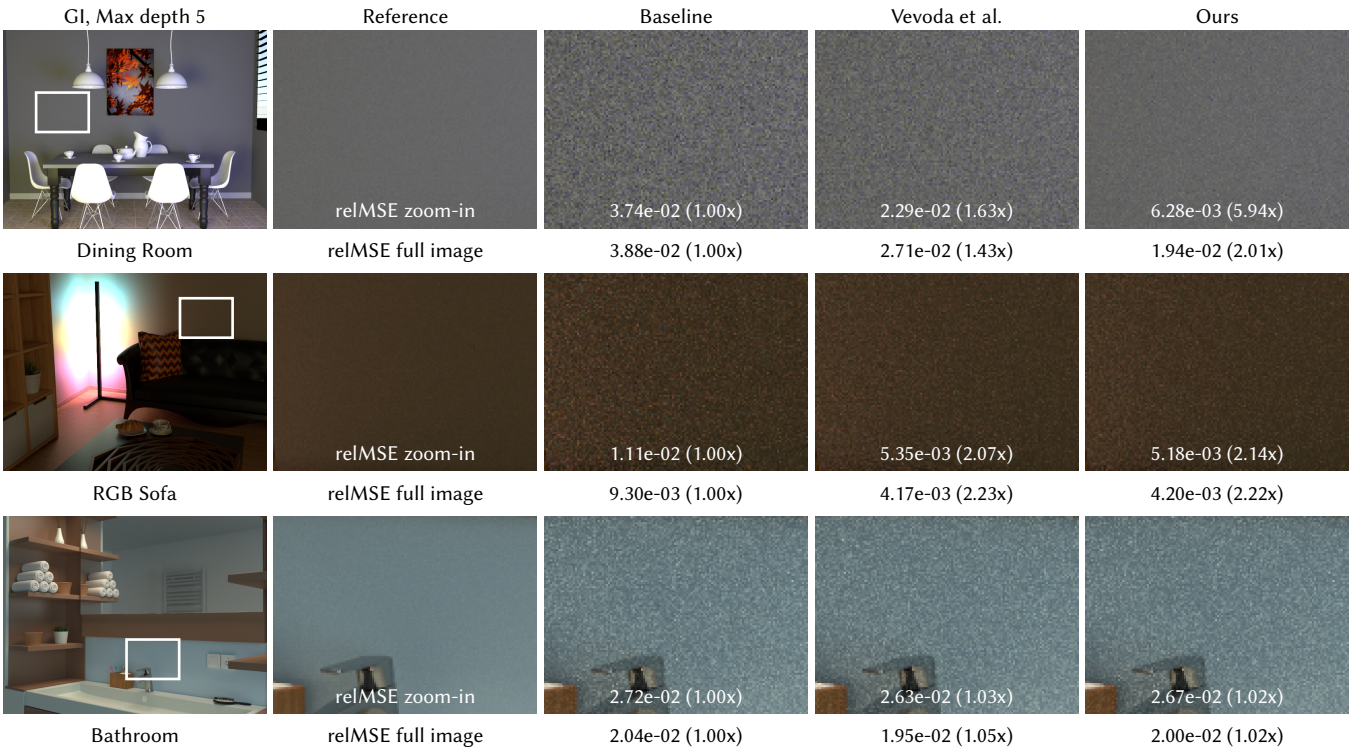


Fig. 8. Equal-time (60s) comparison in full global illumination between ours, Vévoda et al. [2018] and baseline path tracing. The numbers in parentheses are the speed-up compared to the baseline, higher is better. Our method achieves the same or higher performance than the previous ad-hoc approach.

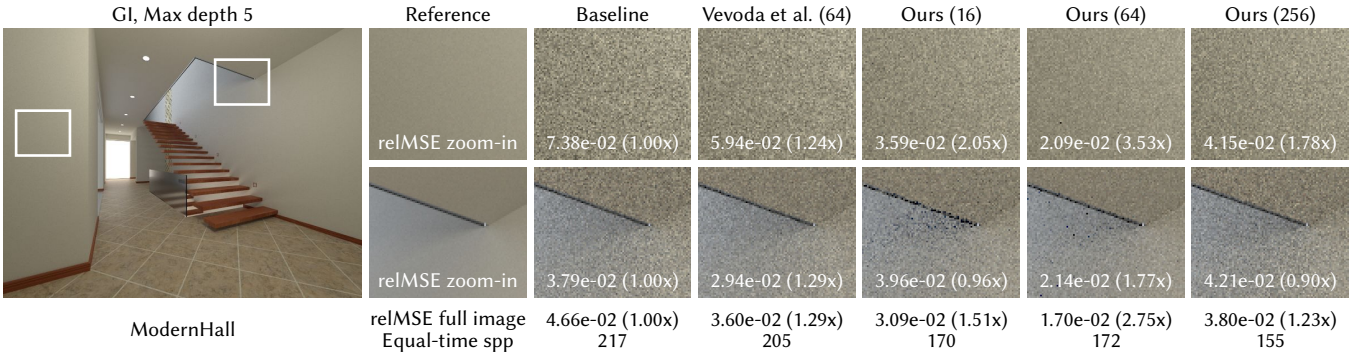


Fig. 9. Equal-time (60s) comparison between baseline path tracing, Vévoda et al. [2018], and ours with different grid resolutions (16, 64, and 256). The first row of zoom-ins shows a region with an almost constant integrand, while the second row features discontinuities in illumination, material, and geometry. Large spatial cells (e.g., 16) perform okay on the flat surface, but not in the discontinuous region. Smaller cells (e.g., 64) perform better, but if cells are too small (e.g., 256), performance is hampered by noisy estimates for the CV coefficients and a higher computational cost.

are severely off. This is a limitation of every on-the-fly adaptation method, like path guiding or adaptive sampling.

With our method, there are two key parameters to control the estimation error in the coefficients: the number of training samples and the size of the spatial cells. Naturally, good choices for these are trade-offs between estimation error, cost, and accuracy.

Fig. 9 demonstrates the impact of different grid resolutions. The same scene is rendered in equal-time with a 16×16, 64×64, and 256×256 grid. Lower resolutions mix incompatible data, finer ones have more noise in the CV estimates and a higher computational overhead. We found that inaccurate CV coefficients due to noise often manifest as negative-valued outliers, like the black pixels in

Table 1. Statistics of the speed-up and overhead (average of 10 runs). Equal-sample comparisons use 128spp, equal-time images are rendered for 60s.

Ours vs baseline			
	Equal-sample	*Equal-time*	*Time overhead*
Dining Room	2.20×	1.90×	15.46%
RGB Sofa	2.17×	2.03×	11.72%
Bathroom	1.03×	1.01×	13.65%
Veach MIS	1.78×	1.60×	16.82%
Modern Hall	3.52×	3.28×	15.27%
Ours vs Vévoda et al.			
	Equal-sample	*Equal-time*	*Time overhead*
Dining Room	1.63×	1.41×	15.31%
RGB Sofa	1.11×	0.98×	15.42%
Bathroom	1.00×	0.92×	17.12%
Veach MIS	1.64×	1.41×	27.30%
Modern Hall	2.62×	2.32×	18.26%

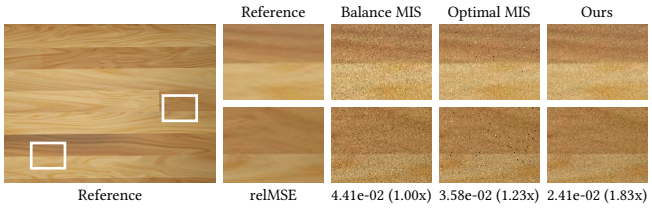


Fig. 10. Equal sample (8spp) comparison between balance MIS, optimal MIS, and ours. Our theory can also be used to more robustly compute optimal MIS weights through spatial averaging, removing the salt-and-pepper noise exhibited by Kondapaneni et al. [2019].

"Ours 64." Full renderings at different grid resolutions are included in the supplemental.

7.4 Comparison to Kondapaneni et al. [2019]

When applied to only direct illumination and for a CV formed from only the MIS techniques, our method is almost identical to the progressive optimal MIS estimator of Kondapaneni et al. [2019]. The sole difference is that they optimize the weight / CV for each pixel integral, while we compute coefficients that perform well on average over multiple pixels.

Fig. 10 shows equal-sample results from the defensive sampling application discussed by Kondapaneni et al. [2019]. Despite averaging over a textured surface, our method achieves virtually identical speed-ups as optimal MIS. Unlike the per-integral optimization, our approach can make do with far fewer training samples, thus we do not suffer from the salt-and-pepper noise encountered by Kondapaneni et al. [2019]. Furthermore, computing fewer coefficients means lower overhead, both in terms of computation as well as memory. Therefore, our theory also constitutes one step forward towards practical optimal MIS weighting.

7.5 Path guiding

Our theoretical findings are general and can be applied in many different contexts. One such example is the sampling of indirect illumination. Our main application formed the CV for only the direct lighting at each hit point. We have also briefly tested applying our theory to indirect illumination sampling.

Specifically, we can form a CV from the mixture density of BSDF components and path guiding PDFs. Our initial findings in that context are shown in Fig. 11. There, we render a glossy box illuminated by a sun and sky model and a large area light. No next event estimation is performed. Thus, path guiding is instrumental to capture the direct illumination from the sun. At the same time, guiding significantly worsens the quality on the glossy surfaces, where BSDF sampling is better suited. Our CV, applied at each vertex during the random walk, achieves noticable variance reduction in this equal-sample test.

Unfortunately, the gains due to our CV are mostly offset by the associated overhead in this application. Further, indirect illumination is prone to severe variance and outliers, even with path guiding. Successfully optimizing a CV requires careful adaptation of the spatial structure and handling of outliers. Further research is required to address these challenges.

8 LIMITATIONS AND FUTURE WORK

The main limitations of our method regard the implementation and the overhead. Beyond improving these, future work could look into applications to bidirectional rendering, or jointly optimizing the CV along with the guiding target functions, the mixture weights, and the Russian roulette probabilities.

Implementation. Implementing the CV computations is challenging for two reasons. First, the PDFs of all individual components must be obtained—provided we aim to unleash the full potential of the CV. In existing code bases, these are not always trivial to access (efficiently). Second, the PDFs must be computed also for samples that have zero contribution [Kondapaneni et al. 2019]. Since renderers often use early-exits when the contribution is zero, it can be tedious (and costly) to obtain all these PDFs.

Overhead. These PDF computations are also a key factor in the overhead. The remainder of the overhead is mainly due to the memory required to estimate and store the coefficients. That cost can be amortized by sharing data structures with an existing adaptive sampling method such as path guiding. Therefore, our CVs are most effective when used jointly with path guiding, or learned light selection, in a renderer with well-optimized PDF computations.

Spatial subdivision. The practical performance of our CV depends greatly on the spatial subdivision. If cells are too large, too many incompatible integrals are mixed and no improvements can be had. If cells are too small, the coefficients contain too much error. In the extreme, a far too fine subdivision can lead to artifacts akin to the salt-and-pepper noise encountered by Kondapaneni et al. [2019]. The question of effective spatial subdivision is one we share with path guiding methods. In that context, multiple approaches and heuristics exist, but a truly satisfactory answer has yet to be found.

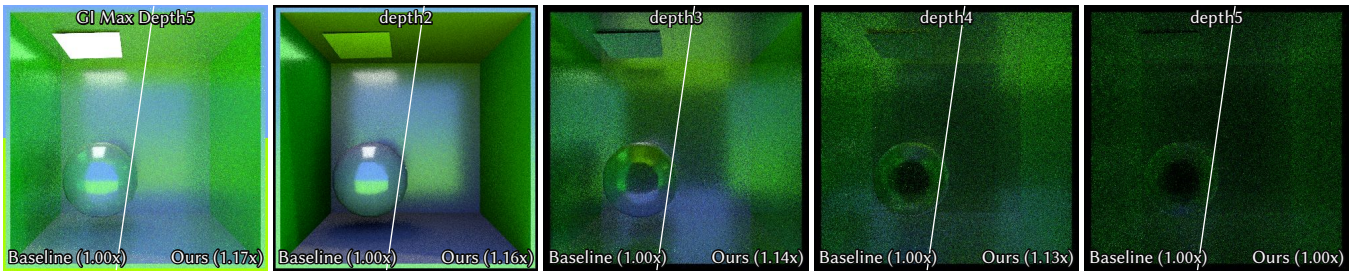


Fig. 11. Equal-sample results from applying our local CVs to guided path tracing without next event estimation. The CV is evaluated at every bounce along the path, hence direct and indirect illumination benefit from it. In this example, improvements can be had by optimizing the mixture sampling in the BSDF model and by compensating for poor guiding quality on glossy surfaces.

Selectively using the CV. Because our CVs only yield improvements in some cases, efficiency can be greatly improved by finding a way to automatically disable the CVs if their variance reduction is not worth the overhead. This could be done by additionally tracking efficiency estimates [Grittmann et al. 2022] using a heuristic to approximate the overhead due to the CV. Alternatively, a simpler approach could be to disable the CV if the computed coefficients are close to canceling out, i.e., if they are approximately proportional to the effective sample counts of the corresponding PDF, $\alpha_i \approx s n_i c_i$. It may be feasible to design a heuristic that reliably detects this case.

Bidirectional methods and correlation. Beyond these performance improvements, it would also be interesting to apply our theory to bidirectional methods. A key challenge there is sample correlation: techniques like photon mapping or splitting produce correlated samples, for which MIS with the balance heuristic is known to fare poorly [Jendersie and Grosch 2018; Grittmann et al. 2021]. Optimizing a CV that takes this correlation into account could be a solution to that problem.

Target functions. The main use-case of our optimized CVs is in conjunction with adaptive methods such as path guiding. There, it would also be interesting to investigate what would be the optimal target functions that a guiding implementation should learn if it is combined with our CVs. Existing work on optimized target functions for plain path guiding [Rath et al. 2020] could be a starting point for future work in this direction.

Mixture weights and path termination. Similarly, He and Owen [2014] showed that the variance is jointly convex in the CV coefficients and the mixture weights. This theory has not yet been applied to rendering and could yield further efficiency gains. Also, a closely related question is that of path termination. When our CVs are used, paths that are terminated through Russian roulette return the sum of CV coefficients instead of zero. In that context, there are two open questions to answer: (1) What is the best average return value for Russian roulette in a spatial cache? (2) Can we optimize the termination probability jointly with the CV coefficients?

9 CONCLUSION

In this article, we revisit the theoretical formulations of controlled mixture sampling and optimal MIS weighting introduced in prior work. These previous methods are not readily applicable to most practical problems in rendering, due to the curse of dimensionality. We show that this obstacle can be overcome by extending the theory to allow local, spatially shared, low dimensional control variates that are optimized to minimize the variance of a full global illumination image. Our theoretical findings are general and, in principle, applicable to almost any problem in rendering. The practical utility is only limited by the overhead and implementation effort. In our application to light selection, we achieve considerable equal-time speed-ups over previous, empirically motivated control variates. We can reap the benefits of PDF-based control variates and optimal MIS weighting in practical applications with full global illumination.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their insightful remarks and Wig42 for sharing the test scene Dining Room.

REFERENCES

Thomas Bashford-Rogers, Kurt Debattista, and Alan Chalmers. 2012. A significance cache for accelerating global illumination. In *Comput. Graph. Forum*, Vol. 31. Wiley Online Library, 1837–1851.

Petrik Clarberg and Tomas Akenine-Möller. 2008. Exploiting visibility correlation in direct illumination. In *Computer Graphics Forum*, Vol. 27. Wiley Online Library, 1125–1136.

Miguel Crespo, Adrian Jarabo, and Adolfo Muñoz. 2021. Primary-space adaptive control variates using piecewise-polynomial approximations. *ACM Transactions on Graphics (TOG)* 40, 3 (2021), 1–15.

Michael Donikian, Bruce Walter, Kavita Bala, Sebastian Fernandez, and Donald P Greenberg. 2006. Accurate direct illumination using iterative adaptive sampling. *IEEE Transactions on Visualization and Computer Graphics* 12, 3 (2006), 353–364.

Shaohua Fan, Stephen Chenney, Bo Hu, Kam-Wah Tsui, and Yu-chi Lai. 2006. Optimizing control variate estimators for rendering. In *Computer Graphics Forum*, Vol. 25. Wiley Online Library, 351–357.

Arthur Firmino, Jeppe Revall Frisvad, and Henrik Wann Jensen. 2022. Progressive Denoising of Monte Carlo Rendered Images. In *Computer Graphics Forum*, Vol. 41. Wiley Online Library, 1–11.

Iliyan Georgiev, Jaroslav Křivánek, Tomáš Davidovič, and Philipp Slusallek. 2012a. Light transport simulation with vertex connection and merging. *ACM Trans. Graph.* 31, 6 (2012), 192–1.

Iliyan Georgiev, Jaroslav Křivánek, Stefan Popov, and Philipp Slusallek. 2012b. Importance caching for complex illumination. In *Comput. Graph. Forum*, Vol. 31. Wiley Online Library, 701–710.

Pascal Grittmann, Iliyan Georgiev, and Philipp Slusallek. 2021. Correlation-Aware Multiple Importance Sampling for Bidirectional Rendering Algorithms. *Comput. Graph. Forum (EG 2021)* 40, 2 (2021).

Pascal Grittmann, Iliyan Georgiev, Philipp Slusallek, and Jaroslav Křivánek. 2019. Variance-Aware Multiple Importance Sampling. *ACM Trans. Graph. (SIGGRAPH Asia 2019)* 38, 6 (Nov. 2019), 152:1–152:9.

Pascal Grittmann, Ömercan Yazici, Iliyan Georgiev, and Philipp Slusallek. 2022. Efficiency-Aware Multiple Importance Sampling for Bidirectional Rendering Algorithms. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2022)* 41, 4, Article

80 (jul 2022), 12 pages. https://doi.org/10.1145/3528223.3530126
Toshiya Hachisuka, Jacopo Pantaleoni, and Henrik Wann Jensen. 2012. A path space extension for robust light transport simulation. *ACM Trans. Graph. (TOG)* 31, 6 (2012), 191.
Hera Y He and Art B Owen. 2014. Optimal mixture weights in multiple importance sampling. *arXiv preprint arXiv:1411.3954* (2014).
Sebastian Herholz and Addis Dittebrandt. 2022. Intel® Open Path Guiding Library. https://www.openpgl.org.
Sebastian Herholz, Oskar Elek, Jiří Vorba, Hendrik P. A. Lensch, and Jaroslav Křivánek. 2016. Product Importance Sampling for Light Transport Path Guiding. *Comput. Graph. Forum* 35 (2016), 67–77.
Sebastian Herholz, Yangyang Zhao, Oskar Elek, Derek Nowrouzezahrai, Hendrik P. A. Lensch, and Jaroslav Křivánek. 2019. Volume Path Guiding Based on Zero-Variance Random Walk Theory. *ACM Trans. Graph.* 38, 3, Article 25 (June 2019), 19 pages.
Heinrich Hey and Werner Purgathofer. 2002. Importance Sampling with Hemispherical Particle Footprints. In *Proceedings of the 18th Spring Conference on Computer Graphics* (Budmerice, Slovakia) *(SCCG '02)*. ACM, 107–114.
Johannes Jendersie and Thorsten Grosch. 2018. An Improved Multiple Importance Sampling Heuristic for Density Estimates in Light Transport Simulations.. In *EGSR (EI&I)*. 65–72.
Henrik Wann Jensen. 1995. Importance Driven Path Tracing using the Photon Map. In *Rendering Techniques*.
Ondřej Karlík, Martin Šik, Petr Vévoda, Tomáš Skřivan, and Jaroslav Křivánek. 2019. MIS Compensation: Optimizing Sampling Techniques in Multiple Importance Sampling. *ACM Trans. Graph. (SIGGRAPH Asia '19)* 38, 6 (2019), 12 pages.
Ivo Kondapaneni, Petr Vévoda, Pascal Grittmann, Tomáš Skřivan, Philipp Slusallek, and Jaroslav Křivánek. 2019. Optimal Multiple Importance Sampling. *ACM Trans. Graph. (SIGGRAPH 2019)* 38, 4 (July 2019), 37:1–37:14.
Jaroslav Křivánek, Iliyan Georgiev, Toshiya Hachisuka, Petr Vévoda, Martin Šik, Derek Nowrouzezahrai, and Wojciech Jarosz. 2014. Unifying Points, Beams, and Paths in Volumetric Light Transport Simulation. *ACM Transactions on Graphics (Proceedings of SIGGRAPH)* 33, 4 (July 2014). https://doi.org/10/f6cz72
Eric P. Lafortune and Yves D. Willems. 1993. Bi-Directional Path Tracing. 93 (Dec. 1993), 145–153.
Heqi Lu, Romain Pacanowski, and Xavier Granier. 2013. Second-Order Approximation for Variance Reduction in Multiple Importance Sampling. In *Computer Graphics Forum*, Vol. 32. Wiley Online Library, 131–136.
Thomas Müller. 2019. "Practical Path Guiding" in Production. In *ACM SIGGRAPH Courses: Path Guiding in Production, Chapter 10* (Los Angeles, California). ACM, New York, NY, USA, 18:35–18:48. https://doi.org/10.1145/3305366.3328091
Thomas Müller, Markus H. Gross, and Jan Novák. 2017. Practical Path Guiding for Efficient Light-Transport Simulation. *Comput. Graph. Forum* 36 (2017), 91–100.
Thomas Müller, Brian McWilliams, Fabrice Rousselle, Markus Gross, and Jan Novák. 2018. Neural importance sampling. *arXiv preprint arXiv:1808.03856* (2018).
Thomas Müller, Fabrice Rousselle, Alexander Keller, and Jan Novák. 2020. Neural control variates. *ACM Transactions on Graphics (TOG)* 39, 6 (2020), 1–19.
Art Owen and Yi Zhou. 2000. Safe and Effective Importance Sampling. *J. Amer. Statist. Assoc.* 95, 449 (2000), 135–143.
Anthony Pajot, Loic Barthe, Mathias Paulin, and Pierre Poulin. 2010. Representativity for robust and adaptive multiple importance sampling. *IEEE transactions on visualization and computer graphics* 17, 8 (2010), 1108–1121.
Matt Pharr, Wenzel Jakob, and Greg Humphreys. 2016. *Physically based rendering: From theory to implementation*. Morgan Kaufmann. https://www.pbr-book.org/3ed-2018/
Alexander Rath, Pascal Grittmann, Sebastian Herholz, Petr Vévoda, Philipp Slusallek, and Jaroslav Křivánek. 2020. Variance-Aware Path Guiding. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2020)* 39, 4 (July 2020), 151:1–151:12. https://doi.org/10.1145/3386569.3392441
Alexander Rath, Pascal Grittmann, Sebastian Herholz, Philippe Weier, and Philipp Slusallek. 2022. EARS: Efficiency-Aware Russian Roulette and Splitting. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2022)* 41, 4, Article 81 (jul 2022), 14 pages. https://doi.org/10.1145/3528223.3530168
Florian Reibold, Johannes Hanika, Alisa Jung, and Carsten Dachsbacher. 2018. Selective guided sampling with complete light transport paths. In *SIGGRAPH Asia 2018 Technical Papers*. ACM, 223.
Fabrice Rousselle, Wojciech Jarosz, and Jan Novák. 2016. Image-space control variates for rendering. *ACM Transactions on Graphics (TOG)* 35, 6 (2016), 1–12.
Lukas Ruppert, Sebastian Herholz, and Hendrik PA Lensch. 2020. Robust fitting of parallax-aware mixtures for path guiding. *ACM Transactions on Graphics (TOG)* 39, 4 (2020), 147–1.
Corentin Salaün, Adrien Gruson, Binh-Son Hua, Toshiya Hachisuka, and Gurprit Singh. 2022. Regression-based Monte Carlo integration. *ACM Transactions on Graphics (TOG)* 41, 4 (2022), 1–14.
Mateu Sbert, Vlastimil Havran, and László Szirmay-Kalos. 2019. Optimal Deterministic Mixture Sampling.. In *Eurographics (Short Papers)*. 73–76.
Vincent Schüßler, Johannes Hanika, Alisa Jung, and Carsten Dachsbacher. 2022. Path Guiding with Vertex Triplet Distributions. In *Computer Graphics Forum*, Vol. 41. Wiley Online Library, 1–15.
Eric Veach. 1997. *Robust Monte Carlo methods for light transport simulation*. Stanford University PhD thesis.
Eric Veach and Leonidas Guibas. 1995a. Bidirectional estimators for light transport. In *Photorealistic Rendering Techniques*. Springer, 145–167.
Eric Veach and Leonidas J Guibas. 1995b. Optimally Combining Sampling Techniques for Monte Carlo Rendering. In *SIGGRAPH '95*. ACM, 419–428.
Petr Vévoda, Ivo Kondapaneni, and Jaroslav Křivánek. 2018. Bayesian online regression for adaptive direct illumination sampling. *ACM Trans. Graph. (TOG)* 37, 4 (2018), 125.
Delio Vicini, Sébastien Speierer, and Wenzel Jakob. 2021. Path Replay Backpropagation: Differentiating Light Paths using Constant Memory and Linear Time. *Transactions on Graphics (Proceedings of SIGGRAPH)* 40, 4 (Aug. 2021), 108:1–108:14. https://doi.org/10.1145/3450626.3459804
Jiří Vorba, Johannes Hanika, Sebastian Herholz, Thomas Müller, Jaroslav Křivánek, and Alexander Keller. 2019. Path Guiding in Production. In *ACM SIGGRAPH 2019 Courses* (Los Angeles, California) *(SIGGRAPH '19)*. ACM, New York, NY, USA, Article 18, 18:1–18:77 pages. https://doi.org/10.1145/3305366.3328091
Jiří Vorba, Ondřej Karlík, Martin Šik, Tobias Ritschel, and Jaroslav Křivánek. 2014. On-line Learning of Parametric Mixture Models for Light Transport Simulation. *ACM Trans. Graph. (Proceedings of SIGGRAPH 2014)* 33, 4 (2014).
Tizian Zeltner, Sébastien Speierer, Iliyan Georgiev, and Wenzel Jakob. 2021. Monte Carlo Estimators for Differential Light Transport. *Transactions on Graphics (Proceedings of SIGGRAPH)* 40, 4 (Aug. 2021). https://doi.org/10.1145/3450626.3459807

A SOLVING FOR SHIFTED COEFFICIENTS

Given two systems of equations,

$$\sum_j a_j A_{i,j} = B_i \quad \text{and} \quad \sum_j a'_j (A_{i,j} + C_A) = B_i + C_B, \tag{34}$$

we show that there is a constant C such that both solutions lie on the line of all optimal solutions, i.e.,

$$\exists C : a_j = a'_j + n_j c_j C. \tag{35}$$

We can show the above by first substituting B_i,

$$\sum_j a_j A_{i,j} = \sum_j a'_j (A_{i,j} + C_A) - C_B \tag{36}$$

and then substituting the desired relationship,

$$\sum_j a'_j A_{i,j} + C \sum_j n_j c_j A_{i,j} = \sum_j a'_j A_{i,j} + C_A \sum_j a'_j - C_B. \tag{37}$$

The coefficients have the form

$$A_{i,j} = \int \lambda_i(x) \frac{\lambda_j(x)}{p(x)} \, \mathrm{d}x. \tag{38}$$

Therefore, if our conditions to ensure robustness are fulfilled, the weighted sum

$$\sum_j c_j n_j A_{i,j} = \int \lambda_i(x) \, \mathrm{d}x = 1 \tag{39}$$

can be simplified to obtain

$$C = C_A \sum_j a'_j - C_B \tag{40}$$

Proving that the constant C exists and hence both systems yield optimal coefficients. Thus, we can apply arbitrary constant shifts to the LHS and RHS of any system of equations that solves our problem.