Variance-Aware Multiple Importance Sampling

PASCAL GRITTMANN, Saarland University, Germany
ILIYAN GEORGIEV, Autodesk, United Kingdom
PHILIPP SLUSALLEK, DFKI and Saarland University, Germany
JAROSLAV KŘIVÁNEK, Charles University and Chaos Czech a. s., Czech Republic

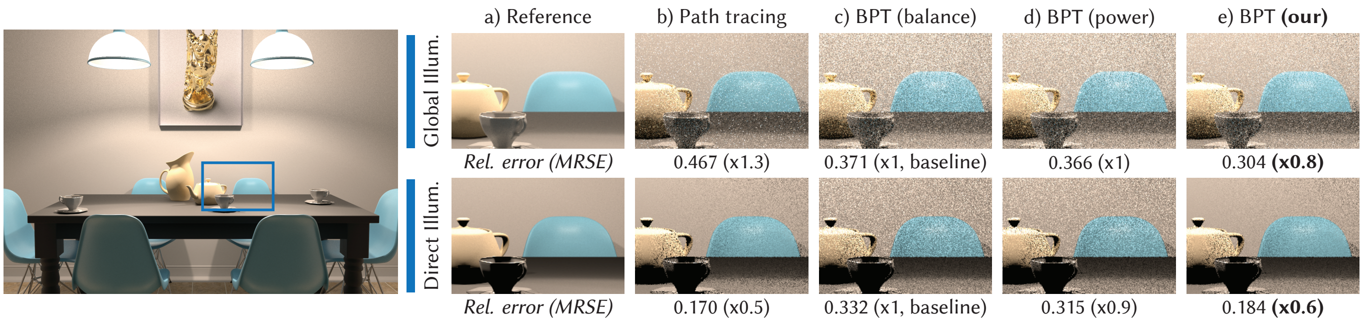

Fig. 1. Equal-time comparison of bidirectional path tracing (BPT) with different MIS heuristics. The balance (b) and power (c) heuristics perform visibly worse than using only the unidirectional path tracing samples that BPT includes (b). The error reduction in parentheses is w.r.t. the balance heuristic combination; lower is better. Our variance-aware balance heuristic significantly improves the result (e), especially the direct illumination component (bottom row).

Many existing Monte Carlo methods rely on multiple importance sampling (MIS) to achieve robustness and versatility. Typically, the balance or power heuristics are used, mostly thanks to the seemingly strong guarantees on their variance. We show that these MIS heuristics are oblivious to the effect of certain variance reduction techniques like stratification. This shortcoming is particularly pronounced when unstratified and stratified techniques are combined (e.g., in a bidirectional path tracer). We propose to enhance the balance heuristic by injecting variance estimates of individual techniques, to reduce the variance of the combined estimator in such cases. Our method is simple to implement and introduces little overhead.

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

Additional Key Words and Phrases: multiple importance sampling, bidirectional path tracing, stratification

ACM Reference Format:
Pascal Grittmann, Iliyan Georgiev, Philipp Slusallek, and Jaroslav Křivánek. 2019. Variance-Aware Multiple Importance Sampling. *ACM Trans. Graph.* 38, 6, Article 152 (November 2019), 9 pages. https://doi.org/10.1145/3355089.3356515

1 INTRODUCTION

Monte Carlo (MC) methods have become an indispensable tool in realistic rendering [Keller et al. 2015]. The performance of an MC estimator largely depends on its variance, which manifests itself as noise in the rendered image. One crucial tool for reducing this variance is multiple importance sampling (MIS). MIS is used to achieve robustness, such as in bidirectional path tracing [Veach and Guibas 1995a; Lafortune and Willems 1993] or vertex connection and merging [Georgiev et al. 2012a; Hachisuka et al. 2012], by combining multiple sampling techniques in the hope that at least one performs well in every scenario.

A key goal of MIS is that the combined algorithm is never (significantly) worse than any one of the sampling techniques alone. As shown in Fig. 1, this goal is not always achieved for bidirectional path tracing: The unidirectional path tracing samples alone (b) perform visibly better for the direct illumination component (bottom row) than the full bidirectional algorithm (c-d). As we discuss later, this effect is due to the combination of the light tracing samples, which are not stratified over the image plane, with the stratified camera samples. We show that the existing MIS heuristics cannot account for the variance reduction due to stratification and produce suboptimal weighting as a result.

The most commonly used MIS weighting heuristics are the balance heuristic and its variants [Veach and Guibas 1995b]. These heuristics are based solely on the effective densities of the sampling techniques, i.e., the product of sample count and sampling probability density. The effective density, however, does not always adequately represent the variance. Previous work has pointed out that low variance does not always manifest as high density, causing the existing MIS heuristics to perform poorly for defensive sampling applications [Owen and Zhou 2000; Georgiev et al. 2012b; Kondapaneni et al. 2019]. We point out that common variance reduction methods like sample (or image plane) stratification, Russian roulette, or splitting are also poorly handled by existing heuristics. These approaches modify either the number of samples or the probability

Authors' addresses: Pascal Grittmann, Saarland University, Saarbrücken, Germany, grittmann@cg.uni-saarland.de; Iliyan Georgiev, Autodesk, London, United Kingdom, iliyan.georgiev@autodesk.com; Philipp Slusallek, DFKI, Saarland University, Saarbrücken, Germany, slusallek@cs.uni-saarland.de; Jaroslav Křivánek, Charles University, Chaos Czech a. s. Prague, Czech Republic, jaroslav.krivanek@mff.cuni.cz.

density, or both in opposing ways. Hence, considering only the product of the two, i.e., the effective density, can be insufficient.

To this end, we propose to integrate variance estimates into the MIS weights. We derive a factor that accounts for the mismatch between the portion of variance that is considered by the balance heuristic and the actual variance. This factor is then used to increase the balance heuristic weight whenever the variance is smaller than the portion considered by the balance heuristic. We show that the proposed heuristic can be easily integrated even in complex algorithms, such as bidirectional path tracing, at a negligible cost. In the example shown in Fig. 1, this heuristic (e) offsets the negative effects of the balance and power heuristics (c,d) and yields variance on par with the – here almost optimal – unidirectional path tracing (b).

2 PREVIOUS WORK

Path tracing. A simple application of MC integration to light transport simulation is path tracing [Kajiya 1986]: tracing random light paths from the camera through the scene until they hit a light source. A path tracer is naturally stratified on the image plane as it renders each pixel individually. Its performance can be improved by next-event estimation, explicitly connecting the vertices along paths to the light sources. Next-event estimation and randomly intersecting lights via BSDF sampling are two examples of sampling techniques that can be combined via MIS.

Multiple importance sampling. MIS [Veach and Guibas 1995b] constructs an unbiased MC estimator by combining several sampling techniques. Samples from each technique are weighted based on their densities. The balance, power, cutoff, and maximum heuristics are provably good weighting functions: For all four of them, the variance of the combined estimator is always within certain bounds of the variance of the (unknown) optimal MIS combination. Recent work has revisited these bounds, pointing out that negative MIS weights (i.e., affine combinations rather than convex ones) can produce even lower variance [Kondapaneni et al. 2019]. In some applications, the original heuristics proposed by Veach and Guibas have been found to perform poorly [Georgiev et al. 2012b; Popov et al. 2015]. Specialized solutions have been proposed for these specific applications. We propose a more general approach that accounts for the mismatch between the portion of the variance that is minimized by existing heuristics and the entire variance.

Robustness in bidirectional sampling. In light transport, MIS is a crucial component of robust algorithms. Bidirectional path tracing [Veach and Guibas 1995a], for example, uses MIS to combine paths starting from the camera with paths traced from the lights in various ways. The efficient combination of multiple path sampling techniques is the key to the robustness of this algorithm under diverse scene and lighting configurations. It can also be combined with photon mapping [Georgiev et al. 2012a; Hachisuka et al. 2012] and further extended to participating media [Křivánek et al. 2014], again relying on MIS.

We show that the existing MIS weighting heuristics, specifically the power heuristic, yield sub-optimal results in such bidirectional methods. This is due to the combination of the unstratified light tracer with other techniques that are stratified over the image plane. The impact of variance reduction schemes like stratification has been discussed in the literature [Veach 1997; Hammersley and Handscomb 1968]. To the best of our knowledge, we are the first to point out that stratification is poorly handled by existing MIS heuristics. Our novel variance-aware heuristic reduces variance of MIS combinations that include stratified techniques.

Efficiency. Robustness does not necessarily imply overall efficiency. Combining many sampling techniques might be robust with respect to handling complex edge cases, but significant overhead is usually incurred by producing samples from techniques that are ineffective in a given situation. Several approaches adapt the number of samples taken from each technique, e.g., based on a variance analysis [Sbert et al. 2016, 2018; He and Owen 2014]. Other approaches alter the sampling densities based on the MIS weights of previously taken samples [Šik et al. 2016; Grittmann et al. 2018; Hachisuka et al. 2014]. For the latter applications to work well, it is particularly important that the MIS weights are close to optimal.

Defensive sampling. In cases where an almost optimal importance sampling pdf is available, a "defensive" (typically uniform) distribution can be mixed in via MIS [Owen and Zhou 2000; Hesterberg 1995] to prevent the potential extreme (or even unbounded) variance in regions where that pdf has low values [Vorba et al. 2014; Herholz et al. 2016]. Our method achieves significant improvements over the existing MIS heuristics for such defensive sampling combinations.

Optimal constant weights. In the MIS framework, the weights are functions of the samples. When they are constrained to be constant, i.e., not allowed to vary with the sample location, the MIS estimator simplifies to the well-known case of linear combination of estimators. In this case, the optimal (constant) weights are inversely proportional to the estimators' variances [Hammersley and Handscomb 1968; Rousselle et al. 2016; Sbert and Havran 2017]. In MIS, such constant weighting is typically – but not always – worse than using the heuristics of Veach [1997]. We strive to combine the best traits of the two approaches: the variance-awareness of the optimal constant weights with the flexibility of the MIS weighting heuristics. Our resulting weights outperform both approaches.

Optimal MIS weights. The concurrent work of Kondapaneni et al. [2019] derives the optimal (non-constant) weights in the MIS framework, showing that MIS can be further improved by allowing the weights to be negative. They compute these weights by solving a linear system with coefficients given by integrals, each as complex as the original integration problem. Their method would become complex and costly in advanced algorithms featuring numerous sampling techniques, such as bidirectional path tracing and derived methods, where it has not yet been applied. While our method is not optimal, its simplicity can be appealing for complex use cases.

3 BACKGROUND

Monte Carlo integration. The integral of a function $f(x)$ over a domain Ω can be estimated via Monte Carlo integration:

$$\mu = \int_{\Omega} f(x)\,\mathrm{d}x \approx \sum_{i=1}^{n} \frac{f(X_i)}{np(X_i)}, \tag{1}$$

where X_i are independent random samples drawn according to some probability density function (pdf) $p(x)$. The quantity $np(x)$ is referred to as the *effective density*. An estimator with a single sample is called a *primary* estimator. Averaging over multiple samples from the same density, as in (1), forms a *secondary* estimator.

Multiple importance sampling. MIS combines several sampling techniques linearly as follows [Veach and Guibas 1995b]:

$$\int_\Omega f(x)\,\mathrm{d}x \approx \sum_{t\in T}\sum_{i=1}^{n_t} w_t(X_{t,i})\frac{f(X_{t,i})}{n_t p_t(X_{t,i})}. \tag{2}$$

Here, T denotes the set of sampling techniques, each drawing n_t samples, and $X_{t,i}$ is the ith sample from technique t. The variance of an MIS estimator, σ^2_{MIS}, has the following form [Veach 1997]:

$$\sigma^2_{\mathrm{MIS}} = \sum_{t\in T}\int_\Omega w_t^2(x)\frac{f^2(x)}{n_t p_t(x)}\,\mathrm{d}x - \sum_{t\in T}\frac{1}{n_t}\left(\int_\Omega w_t(x) f(x)\,\mathrm{d}x\right)^2. \tag{3}$$

Balance and power heuristics. A common choice for the weighting function $w_t(x)$ is the power heuristic [Veach and Guibas 1995b]:

$$w_{t,\mathrm{power}}(x) = \frac{(n_t p_t(x))^\beta}{\sum_{k\in T}(n_k p_k(x))^\beta}. \tag{4}$$

The balance heuristic, obtained when $\beta = 1$, *does not minimize the entire variance functional* (3). Instead, it only minimizes the sum of weighted second moments (the first term in equation (3)):

$$w_{t,\mathrm{balance}} = \arg\min_{w_t}\sum_{t\in T}\int_\Omega w_t^2(x)\frac{f^2(x)}{n_t p_t(x)}\,\mathrm{d}x. \tag{5}$$

The power heuristic with exponent $\beta = 2$ is often used for low-variance combinations, where the variance can be much smaller than the term minimized by the balance heuristic. The power heuristic, however, only amplifies the weighting of the balance heuristic, which sometimes reduces and sometimes increases the variance. We propose a heuristic that accounts for the portion of the variance that is ignored by the balance heuristic.

4 VARIANCE-AWARE MIS WEIGHTS

In this section, we introduce our novel MIS heuristic, analyze its variance, and discuss it on simple one-dimensional examples.

4.1 Motivation

Consider the unweighted contribution of a technique t to the variance (3) of an MIS combination:

$$\sigma_t^2 = \underbrace{\int_\Omega \frac{f^2(x)}{n_t p_t(x)}\,\mathrm{d}x}_{\mu_{2,t}} - \underbrace{\frac{\mu^2}{n_t}}_{r_t}. \tag{6}$$

For the sake of conciseness, we refer to $\mu_{2,t}$ as the second moment of t, when, in fact, it is the second moment of the *primary* estimator, divided by the number of samples n_t used by the *secondary* estimator.

The balance heuristic minimizes the sum of the weighted $\mu_{2,t}$ of all techniques (5), ignoring the residual terms r_t. Now, consider two cases: one where all techniques have high variance, and one where some technique t has low variance.

High variance. If all techniques have high variance, the balance heuristic is close to optimal. For the variance, σ_t^2, to be high, $\mu_{2,t}$ needs to be much larger than the residual term r_t. The portion of the combined variance neglected by the balance heuristic is thus relatively small. In other words, the closer the ratio $\mu_{2,t}/\sigma_t^2$ is to one, for all techniques, the closer the balance heuristic is to being optimal. In the extreme, if for all t we have $\mu_{2,t}/\sigma_t^2 = 1 \Leftrightarrow \mu_{2,t} = \sigma_t^2$, then the balance heuristic minimizes the variance of the combined estimator, i.e., it is optimal.

Low variance. If, however, some technique t has low variance, the balance heuristic will perform poorly. In this case, r_t will be large relative to $\mu_{2,t}$. The larger the ratio $\mu_{2,t}/\sigma_t^2$ is, for some t, the further away the balance heuristic will be from the optimal weights. In the extreme case, if the variance of technique t is zero, the optimal MIS weight would be $w_t(x) = 1$. The balance heuristic, however, considers an infinitely larger quantity $\mu_{2,t}$ and cannot achieve this optimal weighting.

4.2 Variance-aware balance heuristic

Above we observed that the following *variance factor* indicates how close the balance heuristic is to being optimal:

$$v_t = \frac{\mu_{2,t}}{\sigma_t^2}. \tag{7}$$

If v_t is close to one for all techniques, the balance heuristic is almost optimal. A large v_t indicates that the variance of technique t is much lower than the term considered by the balance heuristic, therefore the balance heuristic weight for t should be increased.

As noted in Section 2, weighting based solely on (estimated) variances is also possible. Such constant weights, however, can perform even worse than the balance heuristic [Veach 1997]. We will use the ratio (7) to combine the best of both approaches.

We propose to adjust the balance heuristic weights based on this variance factor, yielding our *variance-aware balance heuristic*:

$$w_{t,\mathrm{our}}(x) = \frac{v_t n_t p_t(x)}{\sum_k v_k n_k p_k(x)}. \tag{8}$$

The three key advantages of this approach are:

(1) the weight of techniques with low variance is increased,
(2) the weighting functions are close to the provably good balance heuristic when all techniques have high variance, and
(3) they yield the optimal constant weight when combining techniques with equal effective densities, but different variance.

We will now discuss and justify each of these points individually.

(1) Low-variance combinations. The variance factor v_t contains the technique's variance in the denominator, thus the weight is increased when the variance is low. In the limit,

$$\lim_{\sigma_t^2\to 0} v_t = \infty. \tag{9}$$

Our variance-aware weights thus handle low-variance problems in a manner that is more likely to achieve an improvement than, e.g., the power heuristic, especially when a low-variance technique also has a low effective density, as we will discuss in the next sections.

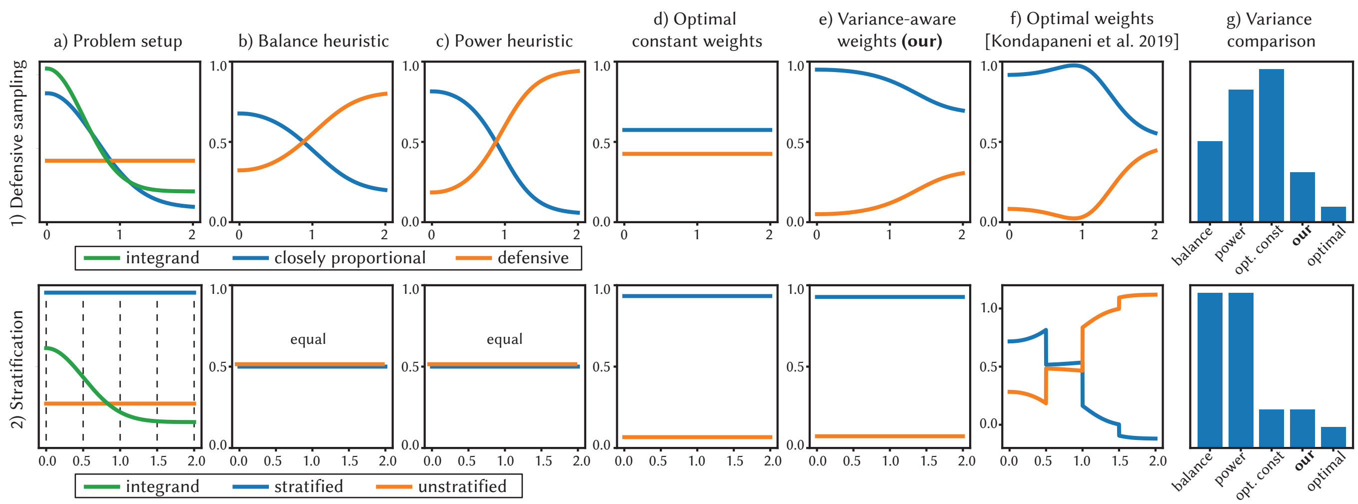

Fig. 2. Two integration problems, each combining samples from two techniques (orange and blue): defensive sampling (top row) and stratification (bottom row). Columns (b-f) compare different MIS weighting heuristics. The power heuristic can amplify the sub-optimal weighting of the balance heuristic, and variance-based weighting (i.e., using optimal constant weights) can be too coarse. Our heuristic is closest to the optimal weights (f).

(2) High-variance combinations. Since the second moment is no smaller than the variance, v_t is always greater than or equal to one. Due to the division by the variance, v_t falls off rapidly towards one with increasing variance. That is,

$$\mu_{2,t} \geq \sigma_t^2 \;\Rightarrow\; v_t \geq 1 \qquad \text{and} \qquad \lim_{\sigma_t^2 \to \infty} v_t = 1. \tag{10}$$

This property effectively maintains the error guarantees of the balance heuristic in high-variance cases. It also implies that variance estimates do not have to be computed for techniques that are expected to have a high variance: v_t can be approximated by one.

(3) Optimality with equal effective densities. The property making our method particularly well suited to combinations of stratified and unstratified techniques (as in bidirectional path tracing) is that it produces the optimal constant weight when the effective densities of all techniques are (almost) the same. With $n_t p_t(x) \approx n_k p_k(x)$ for all pairs of techniques t and k, the weights simplify to

$$w_t(x) \approx \frac{v_t}{\sum_k v_k} = \left(\sum_k \frac{v_k}{v_t}\right)^{-1} = \left(\sum_k \frac{\sigma_t^2}{\sigma_k^2}\right)^{-1}, \tag{11}$$

where the last equality holds because equal effective densities imply equal second moments.

4.3 Discussion in 1D

We now empirically validate our heuristic on simple one-dimensional examples, modeled after the practical settings discussed in Section 6. The source code can be found in the supplemental material.

Figure 2 shows results from two integration problems, one per row. Column (a) shows the problem setup, i.e., the integrand and the sampling densities of two techniques. Columns (b-f) then compare the MIS weights of different heuristics, and column (g) compares their variance.

Defensive sampling. In the first row, the blue density almost matches the integrand, but its low tail induces excessive variance. To ameliorate this, this 'almost-proportional' distribution is combined with a defensive, uniform one. One sample is taken from each.

While the balance and power heuristics do not perform well, they produce lower error than solely variance-based weighting (i.e., the optimal constant weights). Our method combines the strengths of both methods, getting close to the optimal weights derived by Kondapaneni et al. [2019], at lower cost and implementation complexity.

Stratified sampling. The second row combines an unstratified estimator, taking n samples from a uniform density, with a stratified estimator, taking one sample from n uniform densities. Here, the blue line (a) consists of n densities normalized within their corresponding strata. The overall number of samples is the same for both estimators, as are their effective densities, hence Veach's heuristics yield equal weights. Here, our weights are identical to the optimal constant ones.

The optimal weights of Kondapaneni et al. are particularly expensive in this case. The simpler heuristics can treat the stratified variant as one technique, despite it consisting of multiple densities, each non-zero only over the corresponding stratum. The optimal weights can only be obtained by treating these per-stratum densities as separate sampling techniques. Therefore, the cost of computing the optimal weights increases quadratically with the number of techniques and the number of strata, while our method is linear in the number of techniques and independent of the number of strata.

In a bidirectional path tracer, the stratified path tracing samples and the unstratified light tracing samples can have similar effective densities. The two examples in Fig. 3 show that in such a setup, the balance heuristic can perform worse than simply averaging all samples from all techniques. The first row shows the same setup as the second one in Fig. 2. Here, the balance heuristic is equivalent to simply averaging all samples. In the second row, the effective

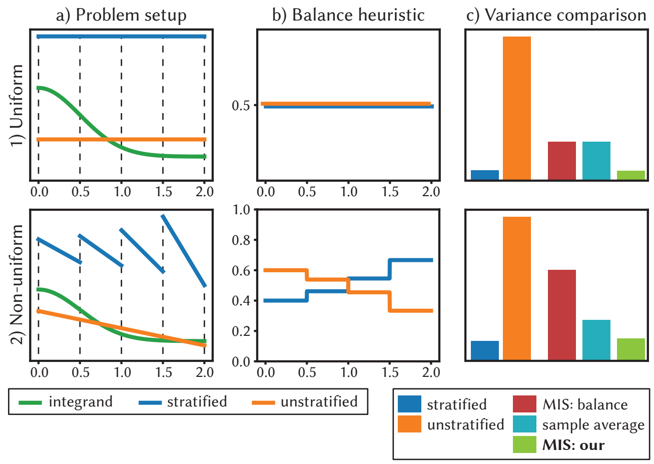

Fig. 3. Two integration problems where a stratified and an unstratified estimator are combined. Column (a) plots the integrand and sampling densities. The unstratified estimator takes n samples from the orange density, while the stratified estimator takes one sample within each of n strata (dashed lines), using the per-stratum normalized densities plotted in blue. Column (b) shows that the effective densities, and hence the balance heuristic weights, are almost identical in both cases. The variance comparison (c), however, reveals that not only is the balance heuristic combination worse than the stratified samples alone, it can also be worse than assigning equal weight to both techniques ('sample average').

densities are slightly different. Despite the unstratified technique having a much higher variance, the balance heuristic assigns a higher weight to it than to the stratified technique, in some regions. The result is worse than simply averaging all samples. In both cases shown in Fig. 3, our new variance-aware weighting heuristic retains the lower variance of the stratified technique. By being close to the optimal constant weights, the variance with our method is, by definition, always better than simply averaging the samples from all techniques.

5 IMPLEMENTATION

To show its suitability to light transport problems, we have implemented our proposed weighting heuristic in two MIS applications: bidirectional path tracing, which combines techniques with and without image plane stratification, and a defensive sampling application for direct illumination. Both have been implemented in PBRT [Pharr et al. 2016], sharing the code for estimating and utilizing the variance factors v_t introduced in the previous section. The source code can be found in the supplemental material.

Procedure. Our method is easily implemented on top of progressive rendering algorithms. To utilize variance estimates, rendering is split into two main stages. In the first stage, we estimate the image with a few samples per pixel (spp), using the standard balance or power heuristics in all MIS calculations; our implementation takes 1 spp. Based on these samples, we estimate the variance of each unweighted technique and compute v_t. In the second stage, any further rendering iterations use our variance-aware heuristic. By estimating the v_t factors in the first stage, and only using them in the second stage, we ensure that the result is unbiased [Kirk and Arvo 1991].

Handling initial samples. Samples from the first stage are not wasted. There are three options to proceed with the rendering result of the first stage: Simply average with the second stage (noisy if the balance heuristic performs poorly), average but weigh based on the v_t factors (biased [Kirk and Arvo 1991]), or keep only those pixels where no technique has a variance factor above a certain threshold. We chose the last approach, with a threshold of $\max_t v_t < 2$.

Estimating v_t. Theoretically, we would like to compute a v_t factor per technique in every pixel. To allow our method to work with a small number of samples, even just one per pixel, we instead divide the image into equal-sized tiles (8×8 in our tests). For each tile, we compute the sample variance and sample mean of each technique. The result are per-tile v_t factors that are used for all pixels within the tile. This approach is simpler and cheaper than approximating v_t for every pixel. Its downside is that the sample variance increases if the tile contains a discontinuity, effectively reverting to the balance heuristic for such tiles.

6 RESULTS AND DISCUSSION

In this section we evaluate the performance of our method in two rendering applications: bidirectional path tracing and defensive sampling for light source selection. The full-size images of all results can be found in the supplemental material, along with the source code and scripts to reproduce them. The tests were performed on a workstation with an Intel i7-4790 processor and 32 GB of RAM.

We compare our method to the optimal MIS weights of Kondapaneni et al. [2019] only for the simpler defensive sampling application. Incorporating these weights into a bidirectional path tracer is a non-trivial task. Aside from the cost of maintaining multiple large matrices for every pixel, the optimal weights also have to consider paths with zero contribution, which requires major incisions into the light transport and material logic. We therefore do not attempt to compare their method to ours on the full bidirectional path tracer. Such a comparison would be interesting to assess the margin for further improvements but it is beyond the scope of this work.

To compare results numerically, we use the mean relative squared error (MRSE) metric. The MRSE of an image is computed by dividing the squared error of each pixel by the reference pixel value, then taking the average of the result among all pixels. This avoids error values being dominated by bright pixels, which the commonly used root mean squared error metric (RMSE) is susceptible to.

6.1 Bidirectional path tracing

Most techniques in bidirectional path tracing are stratified over the image plane, with one exception – light tracing, which connects light paths to the camera via shadow rays. Not being restricted to sampling within a given pixel is the very reason why light tracing is efficient at rendering caustics, focused indirect illumination, and even some types of direct illumination (e.g., lights close to surfaces, inside volumes, or with peaked emission profiles). As we observed before, however, the classical MIS heuristics cannot capture the variance reduction due to stratification in the other techniques, in

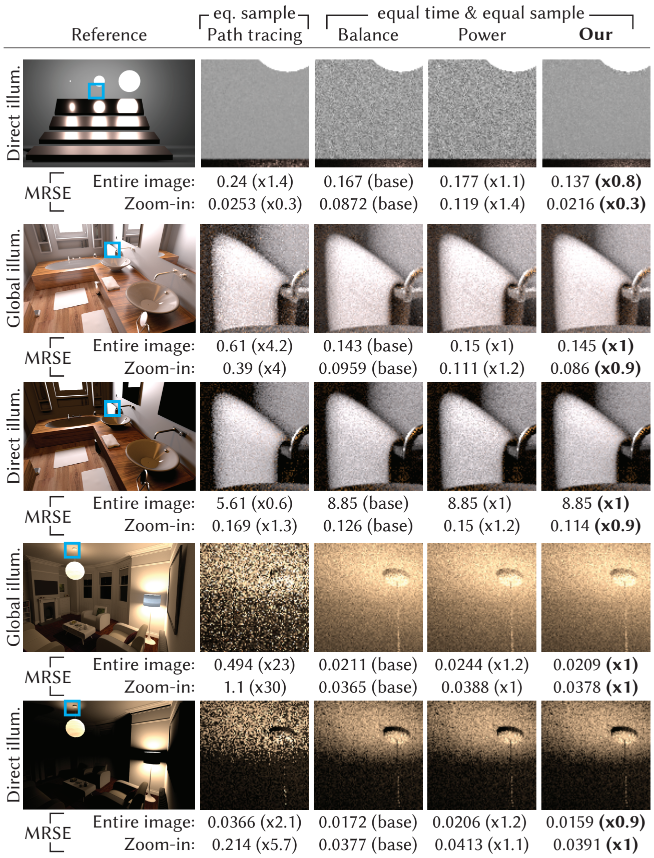

Fig. 4. Equal-time (and equal-sample) comparisons of BPT using three different MIS heuristics. Our heuristic improves low-variance cases while, in contrast to the power heuristic, never being significantly worse than the balance heuristic overall.

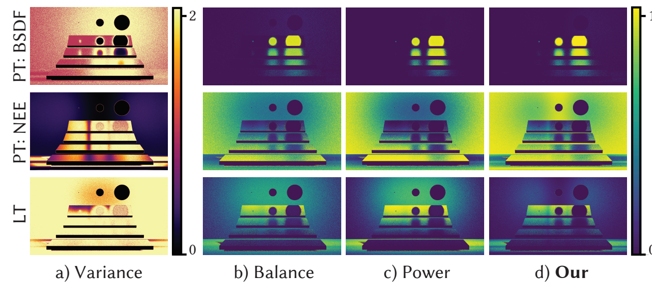

Fig. 5. Comparison of the average per-pixel MIS weights (b-d) to the relative variance of three techniques (a): path tracing with BSDF sampling (PT: BSDF), with next-event estimation (PT: NEE), and light tracing (LT).

this case over the image plane. Figure 1 shows how this can result in excessive noise due to the unstratified light tracer in image regions where the variance of the stratified techniques is low.

Figure 4 compares our heuristic to the balance and power heuristics on three scenes. We show results for full global illumination and for direct illumination alone. The numbers under each row provide the MRSE and the ratio of that error to the balance-heuristic error (in parentheses), across both the entire image and the corresponding zoom-in. Note that the comparisons are not only equal-time but also equal-sample, since our method does not introduce measurable computational overhead. The 'path tracing' images have been produced from the subset of next-event estimation samples in BPT (without the need to apply MIS). Therefore, ideally, the noise in the full MIS combinations should never be higher than those samples alone.

The first row in Fig. 4 shows an extreme case: The error with the balance heuristic is four times larger than the path-tracing samples alone. The MIS-weight comparison in Fig. 5 shows the reason for this behavior: The figure compares the variance of the three techniques (a) for direct illumination (BSDF samples, next-event estimation, and light tracing) to the per-pixel average MIS weights of the three heuristics. Next-event estimation (middle row) has close to zero variance on the wall behind the light sources. This low variance is partially due to the stratification on the image plane, which is ignored by the balance and power heuristics when combining with the unstratified light tracer. Therefore, the balance heuristic assigns an excessive weight to the light tracer. The power heuristic amplifies the issue further. Our approach accounts for the variance reduction due to stratification and maintains the lower error of path tracing.

The second, Bathroom scene in Fig. 4 is a case where no sampling technique has particularly low variance. We achieve small local improvements, as seen in the zoom-ins, while the power heuristic again worsens the results. Importantly, our overall error is not (significantly) worse than that of the balance heuristic.

Finally, the Living room scene shows a case where the unstratified light tracer performs best. Here, our method also achieves small local improvements while not performing worse than the balance heuristic overall. In contrast, the power heuristic produces 20% larger error than the balance heuristic.

Even with coarse 1 spp variance estimates (without any filtering, denoising, or regularization), our method has never performed significantly worse than the balance heuristic in our experiments. The shape of the v_t factors, which quickly fall off to one (i.e., yielding the balance heuristic) as the variance increases, plays an important role in achieving this robustness.

6.2 Comparison to optimal weights

Kondapaneni et al. [2019] illustrated their optimal weights on a simple defensive sampling example: They combine two techniques for light selection in direct illumination computation: selecting the light based on estimates of the unoccluded contribution cached in a regular grid (the 'almost proportional' one) [Pharr et al. 2016], and uniformly (the defensive technique). We implemented our heuristic for the same application so as to compare it to the optimal weights.

The results for the Staircase scene are shown in Fig. 6; the supplemental material contains additional results. The structure of the figure is the same as in Fig. 4, where again the error numbers in the parentheses are relative to the balance heuristic.

For the comparison to the approach by Kondapaneni et al., we used their favored 'direct' estimator which has lower variance and less overhead but also a small amount of bias. The comparisons in

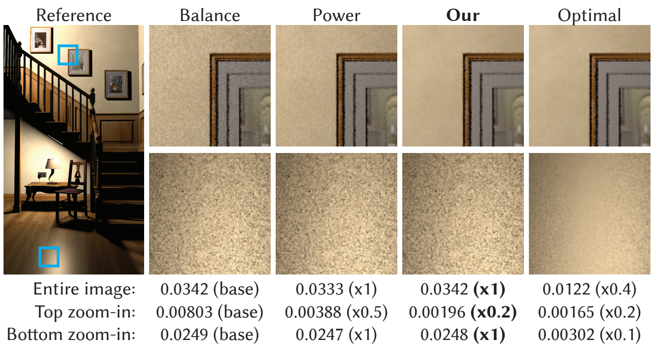

Fig. 6. Equal-sample comparison (8 per technique) for defensive sampling. Our results are close to optimal MIS weights [Kondapaneni et al. 2019] whenever one technique is significantly better than the rest (top zoom-in).

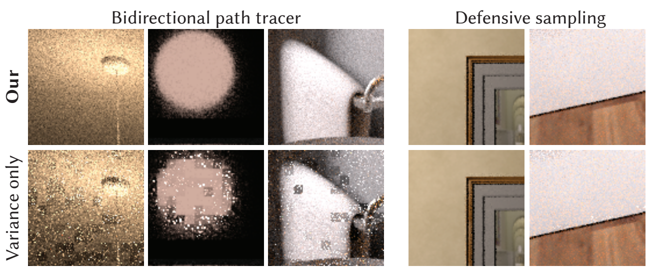

Fig. 7. Incorporating the variance estimates into the balance heuristic (top row), as we propose, greatly improves the robustness compared to using these estimates on their own (bottom row).

this case are equal-sample, using just eight samples per technique (and per pixel). For our heuristic, the comparison is also equal-time to the balance and power heuristics. The approach by Kondapaneni et al. is 5 − 10% slower per sample on our test scenes.

The Staircase scene contains two different cases where significant improvements over the balance and power heuristics are possible: A low-variance problem (top row) and an example where the *negativity* of the optimal weights can reduce the error, despite all techniques having high variance (bottom row). For the low-variance problem, our result is close to that of Kondapaneni et al.'s direct estimator. However, our heuristic cannot be negative and therefore cannot improve on the balance heuristic in the second case (where the optimal weights are negative).

In conclusion, our method and the approach of Kondapaneni et al. complement each other. Their method is provably optimal, yet challenging to apply in practice. Our method is simpler to implement, but only ensures that the combined estimator is never worse than the best technique alone, in an equal-sample comparison.

6.3 Comparison to weighting with variance estimates

Figure 7 compares our variance-aware MIS weights (top row) to classical variance-based weighting (bottom row) on zoom-ins from various scenes. In some cases, especially the simpler defensive sampling application, using variance estimates alone can produce results similar to our weights. Noise in the variance estimates, however, will

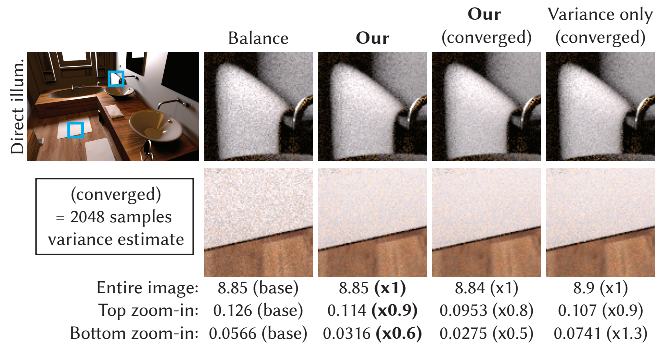

Fig. 8. Comparison to using variance estimates computed with 2048 samples per pixel. Computing more accurate v_t factors can improve the results further, but at higher cost. Even the converged variances alone perform worse than our approach with coarse estimates.

then manifest as visible artifacts in the image. Our method prevents these artifacts due to the injection into the balance heuristic and the lower bound of one (i.e., $v_t \geq 1$). While the variance estimates are far too noisy to be usable on their own, incorporating them into the balance heuristic, as we propose, yields improvements without any visible artifacts.

We also compare the performance when using accurate variance estimates (computed from 2048 samples). The results for direct illumination on the Bathroom scene are shown in Fig. 8. Using the accurate variance estimates alone performs worse than using those same estimates with our approach.

6.4 Overhead

Computation-wise, the overhead of our method is negligible. However, storing the v_t factors can require a significant amount of memory if the number of combined techniques is very large.

In the bidirectional path tracer, our implementation computes the v_t factors for path lengths up to five – a total of 25 techniques. Since we store v_t every 8×8 pixels, this requires roughly two bytes per pixel on average – not much in the context of a typical renderer.

The memory footprint can be reduced further. Figure 9 shows the v_t factors for the direct and one-bounce indirect illumination in the Bathroom scene. A large fraction of these v_t factors are almost one. Therefore, applying compression would significantly reduce the memory requirements. We leave such optimizations for future work as in our benchmarks the overhead has been insignificant.

7 LIMITATIONS AND FUTURE WORK

In addition to the downside of relying on estimated quantities, some care has to be taken when applying our heuristic to biased techniques. In this section we discuss such limitations and outline possible improvements. Lastly, we introduce other promising applications for our heuristic.

7.1 Limitations

Error in variance estimates. Weighting using estimated quantities can worsen the results when these estimates are highly inaccurate. We have not encountered any such issues in our experiments, despite

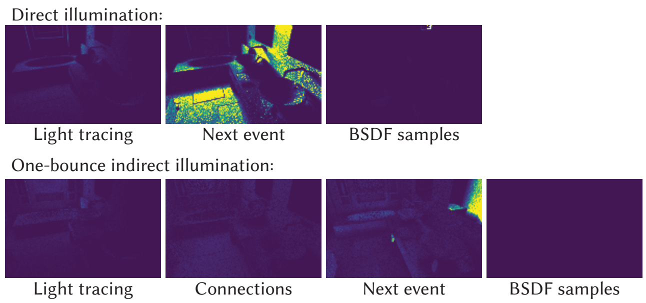

Fig. 9. The estimated v_t factors for the bidirectional path tracer techniques in the Bathroom scene. The majority of pixels receive a factor close to one.

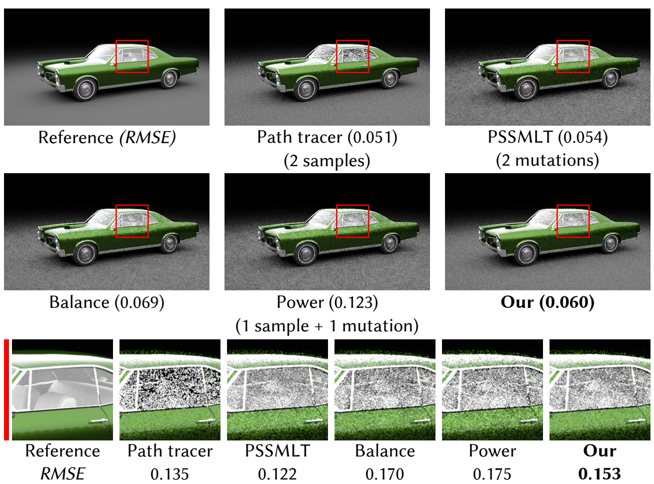

Fig. 10. Equal-sample comparison for the proof-of-concept Metropolis combination. The combination with our heuristic is the most robust: it retains the lower variance of both the Metropolis and the MC approach.

basing our estimates on a single sample per pixel. There could, however, be applications or particularly challenging scenes where errors in the v_t estimates cause issues.

There are several options to ameliorate such potential issues. The images storing the v_t factors can be filtered or denoised and potential outliers could be clamped. Another approach, naturally, is to increase the number of v_t estimation samples, or to accumulate better estimates progressively.

Narrow-support sampling pdfs. So far we have only considered the case where the pdf of each individual technique is non-zero over the entire integration domain. If this is not the case, weighting based on the variance no longer makes sense. In that case, the problem can be made recursive: A set of techniques, whose pdfs together cover the entire domain, can be combined via the regular balance heuristic, which is equivalent to sampling from the mixture of these pdfs [Veach 1997]. This mixture can then be combined with other sampling techniques using our variance-aware heuristic. An alternative is to simply set the variance factors to one for all narrow-support techniques.

7.2 Other applications

Efficiency. Our MIS weights help achieve robustness by preventing the noise from poorly performing techniques deteriorate the quality that could be achieved by using only the samples from the better techniques. However, for an algorithm to be truly efficient, better MIS weights are only the first step: Ideally, we also want to ensure that most samples are taken from the best-performing techniques. There have been attempts to improve the sample allocation based on the MIS-weighted contribution of previously taken samples [Šik et al. 2016; Grittmann et al. 2018; Hachisuka et al. 2014]. In combination with such approaches, our MIS weights could achieve a considerable increase in the efficiency of algorithms like vertex connection and merging (VCM) [Georgiev et al. 2012a].

VCM. Bidirectional path tracing, on top of which we apply our method, is a subset of the more powerful VCM algorithm [Georgiev et al. 2012a; Hachisuka et al. 2012]. Therefore, we can expect that the full VCM algorithm will benefit from similar improvements. Specifically, since photon mapping and light tracing for direct illumination are almost identical techniques, the direct-illumination results with VCM will be identical to those of BPT. It is possible, that our approach will improve on the VCM algorithm even further: Recent work [Jendersie and Grosch 2018; Jendersie 2019] has shown that the MIS weights for the vertex merging technique in VCM can perform poorly in some cases. Our method could also help there.

Metropolis light transport. Our observation that existing MIS heuristics neglect stratification has inspired us to experiment with a novel MIS combination: Combining a Metropolized path tracer with a regular Monte Carlo path tracer. Rendering methods based on the Metropolis algorithm are often criticized for their lack of image plane stratification [Cline et al. 2005; Šik and Křivánek 2018]. In a proof-of-concept experiment, we therefore tried combining a primary sample space Metropolis path tracer (PSSMLT) [Kelemen et al. 2002] with a stratified MC path tracer, using MIS.

An equal-time comparison is shown in Fig. 10. The MC path tracer, thanks to the stratification, can easily handle the direct illumination in that scene, but struggles with the indirectly lit car interior. PSSMLT easily resolves that more challenging interior, but due to the lack of stratification exhibits much stronger noise in the direct illumination (e.g., on the car exterior and the floor). While the balance- and power-heuristic combinations retain the better performance on the car interior, their unawareness of image plane stratification results in higher levels of noise in the simpler direct illumination. Our approach retains the better performance of both techniques, achieving a more efficient combination.

The optimal weights [Kondapaneni et al. 2019] are not applicable to such a combination. With Metropolis methods, the exact sampling densities are unknown and require approximations for use in MIS [Šik et al. 2016; Kelemen et al. 2002]. Therefore, computing the optimal MIS weights is impossible – a heuristic like ours is needed for such a combination to perform well.

8 CONCLUSION

We propose a novel weighting heuristic for multiple importance sampling that incorporates variance estimates. We show that existing MIS heuristics neglect the impact of stratification, correlation, and other effects on the variance – a shortcoming that can be addressed by incorporating variance estimates. We apply our theory to bidirectional path tracing, defensive sampling, and a proof-of-concept combination involving Metropolis sampling. Throughout all our tests, even coarse estimates of the variance have been sufficient to achieve significant improvements over the balance heuristic in some cases, while – most importantly – never performing worse.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their valuable feedback. The test scenes are slightly modified versions of those in the scene repositories of PBRT [Pharr et al. 2016] and Benedikt Bitterli [Bitterli 2016]. This work was supported by the Czech Science Foundation Grant 19-07626S and Charles University Grant SVV-2017-260452.

REFERENCES

Benedikt Bitterli. 2016. Rendering resources. https://benedikt-bitterli.me/resources/.

David Cline, Justin Talbot, and Parris Egbert. 2005. Energy redistribution path tracing. In *ACM Trans. Graph. (SIGGRAPH '05)*, Vol. 24. ACM, 1186–1195.

Iliyan Georgiev, Jaroslav Křivánek, Tomáš Davidovič, and Philipp Slusallek. 2012a. Light Transport Simulation with Vertex Connection and Merging. *ACM Trans. Graph. (SIGGRAPH Asia '12)* 31, 6, Article 192 (Nov. 2012), 10 pages.

Iliyan Georgiev, Jaroslav Křivánek, Stefan Popov, and Philipp Slusallek. 2012b. Importance Caching for Complex Illumination. *Comput. Graph. Forum (EG '12)* 31 (2012), 701–710.

Pascal Grittmann, Arsène Pérard-Gayot, Philipp Slusallek, and Jaroslav Křivánek. 2018. Efficient Caustic Rendering with Lightweight Photon Mapping. In *Comput. Graph. Forum (EGSR '18)*, Vol. 37. 133–142.

Toshiya Hachisuka, Anton S Kaplanyan, and Carsten Dachsbacher. 2014. Multiplexed metropolis light transport. *ACM Trans. Graph.* 33, 4 (2014), 100.

Toshiya Hachisuka, Jacopo Pantaleoni, and Henrik Wann Jensen. 2012. A Path Space Extension for Robust Light Transport Simulation. *ACM Trans. Graph. (SIGGRAPH Asia '12)* 31, 6, Article 191 (Nov. 2012), 10 pages.

J.M. Hammersley and D.C. Handscomb. 1968. Monte Carlo Methods (Methuen, London).

Hera Y He and Art B Owen. 2014. Optimal mixture weights in multiple importance sampling. *arXiv preprint arXiv:1411.3954* (2014).

Sebastian Herholz, Oskar Elek, Jiří Vorba, Hendrik Lensch, and Jaroslav Křivánek. 2016. Product Importance Sampling for Light Transport Path Guiding. *Comput. Graph. Forum (EGSR '16)* 35, 4 (2016), 67–77.

Tim Hesterberg. 1995. Weighted average importance sampling and defensive mixture distributions. *Technometrics* 37, 2 (1995), 185–194.

Johannes Jendersie. 2019. Variance Reduction via Footprint Estimation in the Presence of Path Reuse. In *Ray Tracing Gems* (1 ed.), Eric Haines and Tomas Akenine-Möller (Eds.). Vol. 1. Apress, 557–569.

Johannes Jendersie and Thorsten Grosch. 2018. An Improved Multiple Importance Sampling Heuristic for Density Estimates in Light Transport Simulations. In *Proc. of Eurographics Symposium on Rendering EI&I Track (EGSR)*. Eurographics Association, 65–72.

James T. Kajiya. 1986. The Rendering Equation. In *Proceedings of SIGGRAPH '86*. ACM, New York, NY, USA, 143–150.

Csaba Kelemen, László Szirmay-Kalos, György Antal, and Ferenc Csonka. 2002. A simple and robust mutation strategy for the metropolis light transport algorithm. In *Comput. Graph. Forum (EG '02)*, Vol. 21. Wiley Online Library, 531–540.

A. Keller, L. Fascione, M. Fajardo, I. Georgiev, P. Christensen, J. Hanika, C. Eisenacher, and G. Nichols. 2015. The Path Tracing Revolution in the Movie Industry. In *ACM SIGGRAPH 2015 Courses (SIGGRAPH '15)*. ACM, New York, NY, USA, Article 24, 24:1–24:7 pages.

David Kirk and James Arvo. 1991. Unbiased sampling techniques for image synthesis. In *ACM Trans. Graph. (SIGGRAPH '91)*, Vol. 25. ACM, 153–156.

Ivo Kondapaneni, Petr Vévoda, Pascal Grittmann, Tomaš Skřivan, Philipp Slusallek, and Jaroslav Křivánek. 2019. Optimal Multiple Importance Sampling. *ACM Trans. Graph. (SIGGRAPH '19)* 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 Trans. Graph. (SIGGRAPH '14)* 33, 4 (Aug. 2014), 1–13.

Eric P Lafortune and Yves D Willems. 1993. Bi-directional Path Tracing. (1993).

Art Owen and Yi Zhou. 2000. Safe and Effective Importance Sampling. *J. Amer. Statist. Assoc.* 95, 449 (2000), 135–143.

Matt Pharr, Wenzel Jakob, and Greg Humphreys. 2016. *Physically Based Rendering: From Theory to Implementation* (3rd ed.). Morgan Kaufmann.

Stefan Popov, Ravi Ramamoorthi, Fredo Durand, and George Drettakis. 2015. Probabilistic Connections for Bidirectional Path Tracing. *Comput. Graph. Forum (EGSR '15)* 34, 4 (2015), 075–086.

Fabrice Rousselle, Wojciech Jarosz, and Jan Novák. 2016. Image-space Control Variates for Rendering. *ACM Trans. Graph. (SIGGRAPH Asia '16)* 35, 6, Article 169 (Nov. 2016), 12 pages.

Mateu Sbert and Vlastimil Havran. 2017. Adaptive multiple importance sampling for general functions. *The Visual Computer* 33, 6-8 (2017), 845–855.

Mateu Sbert, Vlastimil Havran, and László Szirmay-Kalos. 2016. Variance Analysis of Multi-sample and One-sample Multiple Importance Sampling. In *Comput. Graph. Forum (Pacific Graphics '16)*, Vol. 35. Wiley Online Library, 451–460.

Mateu Sbert, Vlastimil Havran, László Szirmay-Kalos, and Víctor Elvira. 2018. Multiple importance sampling characterization by weighted mean invariance. *The Visual Computer* 34, 6-8 (2018), 843–852.

Martin Šik and Jaroslav Křivánek. 2018. Survey of Markov Chain Monte Carlo Methods in Light Transport Simulation. *IEEE transactions on visualization and computer graphics* (2018).

Martin Šik, Hisanari Otsu, Toshiya Hachisuka, and Jaroslav Křivánek. 2016. Robust light transport simulation via metropolised bidirectional estimators. *ACM Trans. Graph.* 35, 6 (2016).

Eric Veach. 1997. *Robust Monte Carlo Methods for Light Transport Simulation.* Ph.D. Dissertation. Stanford University.

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 *Proceedings of SIGGRAPH '95*. ACM, 419–428.

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. (SIGGRAPH '14)* 33, 4 (2014), 101.