

Vector-Valued Monte Carlo Integration Using Ratio Control Variates

HAOLIN LU, University of California San Diego, USA and Max Planck Institute for Informatics, Germany

DELIO VICINI, Google, Switzerland

WESLEY CHANG, University of California San Diego, USA

TZU-MAO LI, University of California San Diego, USA

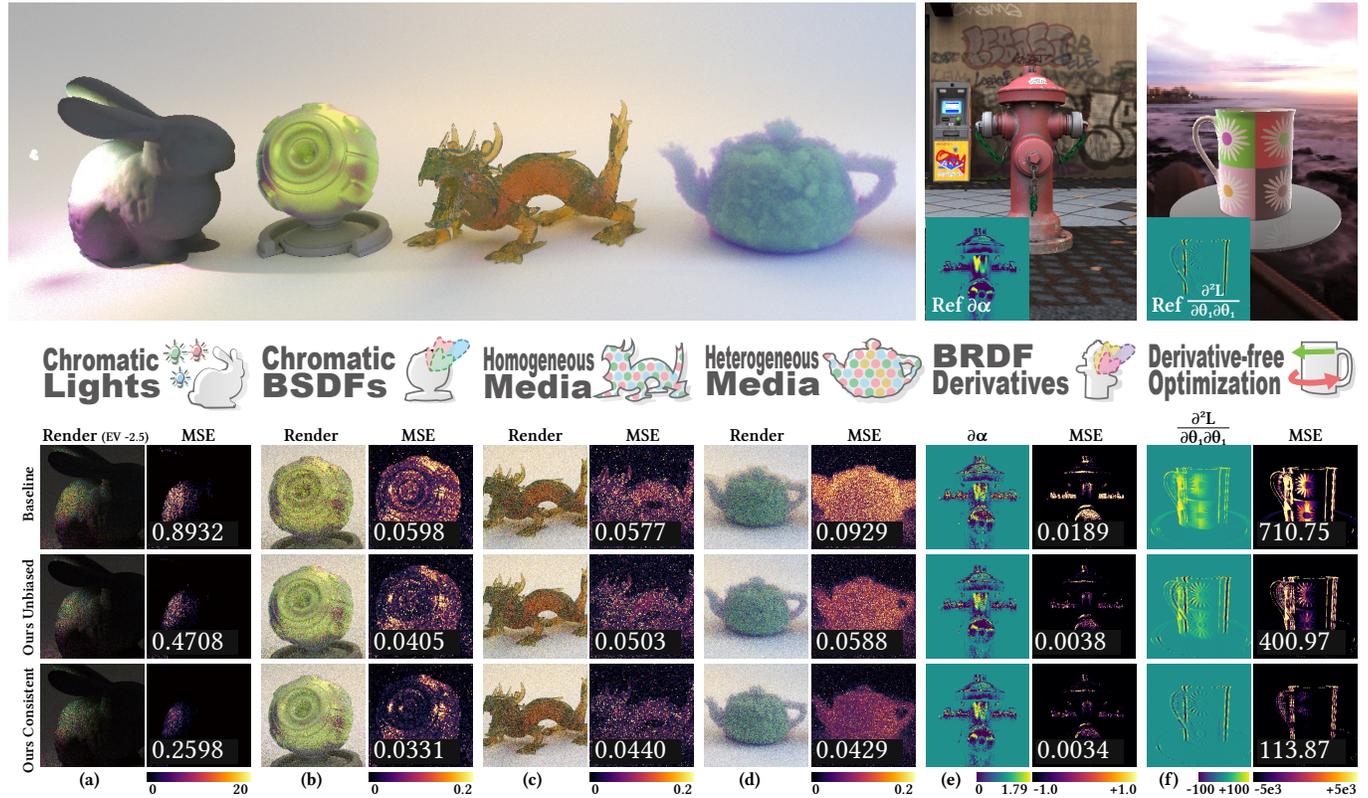


Fig. 1. Vector-valued integration is ubiquitous in forward and inverse rendering. In forward rendering, the integrated radiance often differs between color channels or wavelength due to chromatic (a) lights (BUNNY illuminated by three spherical area lights with RGB radiance values of [3000, 5, 6], [3, 3000, 4], and [4.8, 5.8, 3000]). For better visualization, we underexpose inset images), (b) BSDFs, (c) homogeneous, and (d) heterogeneous media. A common practice to tackle this challenge is sampling the luminance or a random mixture of each wavelength, leading to obvious color noise. In inverse rendering, we estimate (e) the vector-valued scene parameter gradient, or (f) the Hessian matrix (through a derivative-free method) for higher-order optimization. We propose a biased but consistent estimator and its unbiased variants to mitigate the challenge of vector-valued integration. Our methods can significantly reduce the variance with often negligible overhead, and can be easily integrated into existing forward and inverse rendering frameworks.

Variance reduction techniques are widely used for reducing the noise of Monte Carlo integration. However, these techniques are typically designed with the assumption that the integrand is scalar-valued. Recognizing that rendering and inverse rendering broadly involve vector-valued integrands, we identify the limitations of classical variance reduction methods in this context. To address this, we introduce ratio control variates, an estimator

that leverages a ratio-based approach instead of the conventional difference-based control variates. Our analysis and experiments demonstrate that ratio control variables can significantly reduce the mean squared error of vector-valued integration compared to existing methods and are broadly applicable to various rendering and inverse rendering tasks.

CCS Concepts: • Computing methodologies → Rendering.

Additional Key Words and Phrases: ratio control variates, importance sampling, differentiable rendering, inverse rendering

1 INTRODUCTION

Variance reduction methods in Monte Carlo integration are often designed for *scalar*-valued integrals. However, rendering and inverse

Authors' addresses: Haolin Lu, University of California San Diego, USA and Max Planck Institute for Informatics, Germany; Delio Vicini, vicini@google.com, Google, Switzerland; Wesley Chang, University of California San Diego, USA; Tzu-Mao Li, University of California San Diego, USA.



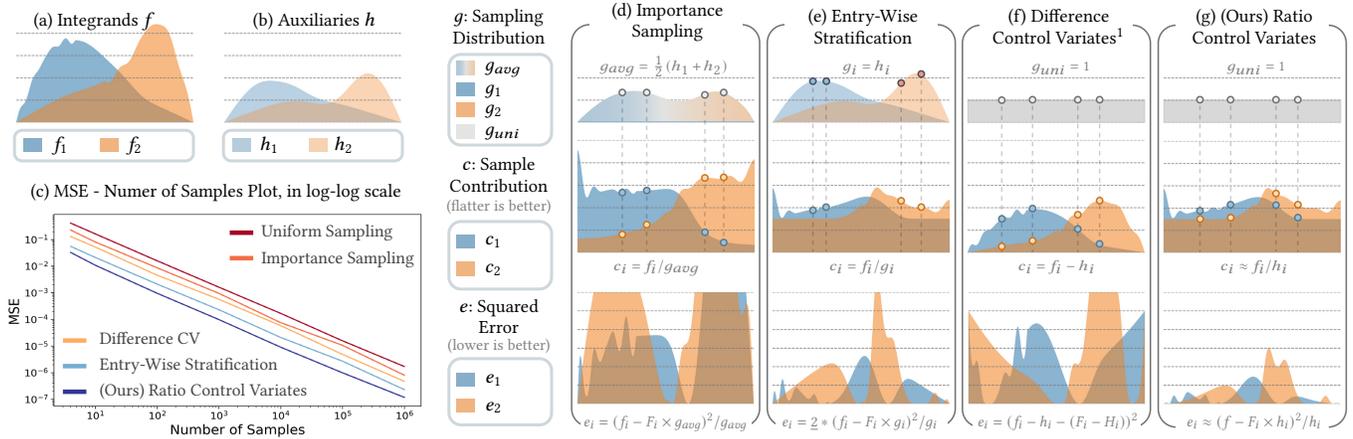


Fig. 2. *1D Toy Example*. Given (a) two integrands $[f_1, f_2]$, we assume prior knowledge in the form of (b) two auxiliary functions h_1, h_2 , which are roughly proportional to their respective integrands, normalized, and can be sampled. We compare four variance reduction techniques, and (c) plot their respective numerical error over the number of Monte Carlo samples. (d) *Vector-valued importance sampling* samples from a single distribution $g_{avg}(x)$ that combines both auxiliary functions. All samples are then used to evaluate both f_1 and f_2 , but the inability of g_{avg} to accurately represent either integrand leads to high variance. (e) *Entry-wise stratification* samples separately from h_1 and h_2 , with samples from each auxiliary used only for the corresponding integrand. This improves the accuracy of sampling distribution and reduces variance, but halves the sample count per integrand (e.g., 2 instead of 4 here). (f) *Difference control variates* is inefficient in this example due to the scaling relation between f and h , where $f \approx 2h$. This makes the difference $f - h$ far from constant, thereby reducing efficiency. (g) Our *ratio control variates* avoids the trade-off between specialization and sample count seen in importance sampling. It also effectively handles scaling effects, unlike difference control variates. In the third row, we demonstrate the efficiency of the estimators by visualizing the *mean squared error* (MSE) as a function $e(x)$. The final MSE can be computed as $MSE = \frac{1}{N} \int (e_1(x) + e_2(x)) dx$, so lower $e(x)$ indicates better performance.

rendering problems require solving *vector*-valued integrals since we need to estimate the radiance across multiple color channels or wavelengths, and for inverse rendering we need to estimate gradients for multiple parameters. In this work, we analyze different Monte Carlo estimators in the context of vector-valued integration in rendering, and adapt ratio control variates for a wide range of vector-valued integration problems in forward and inverse rendering.

Consider the example in Fig. 2, where we aim to estimate two integrals sharing the same domain (e.g., the two integrands represent two of the color channels). In rendering, we are usually also given two corresponding *auxiliary functions* that are roughly proportional to our integrands (e.g., the material importance sampling probability density functions of the two color channels). Standard importance sampling cannot find a single sampling distribution from the auxiliary functions that matches both integrands well. However, using a different distribution for each integrand, and treating the two integrals separately, means that a light path can only contribute to one channel, effectively doubling the computational cost.

Standard *difference* control variates can be easily extended for vector-valued problems by using a different auxiliary function for each integrand. However, difference control variates are only effective when the difference between the integrand and auxiliary functions is nearly constant. In actual rendering applications, this is usually not true, as the auxiliary function usually only models part of the integrand. For example, material sampling lobes would ignore the incoming radiance. In contrast, importance sampling remains effective when the *ratio* of the integrand and the auxiliary function is close to a constant, making it more robust.

Our key idea is to use a ratio estimator instead of the conventional difference estimator in control variates. By estimating the ratio of the integrands and the auxiliary function, our method is invariant to the scale difference between the integrand and the auxiliary function and shares the same robustness as importance sampling. Additionally, as our method is a control variates technique, it naturally extends to vector-valued problems. While a basic ratio control variates estimator is consistent but biased, we show that it can be efficiently debiased with minimal overhead. Finally, ratio control variates are a direct generalization of previously used *weighted importance sampling* methods, as we will discuss later.

We show that our ratio control variates estimator have a wide variety of forward and inverse rendering applications, by using existing importance sampling constructions, and only requires minimal modification of the renderer. These applications include sampling chromatic materials and lights and combining them using multiple importance sampling, many-light sampling, global illumination, multiple scattering in both homogeneous and heterogeneous media, differentiable path tracing, derivative-free optimization for inverse rendering, and even Monte Carlo partial differential equation solvers. Our method significantly reduces the error compared to standard techniques with minimal computational and memory overhead.

Our technical contributions are:

- We formulate forward and inverse rendering as vector-valued Monte Carlo integration.

¹For simplicity, we use a uniform sampling distribution for both the difference and ratio control variates. In practice, they can be combined with importance sampling.

- We analyze standard variance reduction techniques including importance sampling and difference control variates, and demonstrate their issues both theoretically and empirically.
- We introduce ratio control variates, its biased and unbiased variants for vector-valued integration, analyze its error, and show that it is *scale* invariant, as opposed to standard difference control variates that is *shift* invariant.
- We show how to apply ratio control variates to a wide range of rendering scenarios, and propose extensions to work with multiple importance sampling, global illumination, and volumetric scattering.
- We propose a *marginal* ratio control variates to handle differentiable path tracing and reuse existing importance sampling distributions.

An open-source implementation of our method is available under <https://github.com/suikasibyl/vvmc>.

2 RELATED WORK

Existing variance reduction techniques for Monte Carlo integration [Owen 2013; Veach 1998] largely focus on scalar-valued estimators. We review related work on variance reduction for path tracing, inverse rendering and gradient descent, as well as ratio estimators.

Variance reduction in rendering. Practical Monte Carlo path tracing [Kajiya 1986; Pharr et al. 2023] relies on the use of analytical importance sampling distributions [Arvo 1995; Dupuy and Benyoub 2023; Eto and Tokuyoshi 2023; Heitz 2018; Lin and Yuksel 2020; Miller et al. 2019; Novák et al. 2018; Ureña et al. 2013]. Recently, methods that progressively refine sampling distributions [Bitterli et al. 2020; Lin et al. 2022; Müller et al. 2017; Müller et al. 2019; Xu et al. 2024] have led to further improvements on challenging scenes.

Control variates is another popular technique [Crespo et al. 2021; Fan et al. 2006; Hua et al. 2023; Lafortune and Willems 1995; Müller et al. 2020; Salaün et al. 2022; Szirmay-Kalos et al. 2001; Szécsi et al. 2004]. Prior work in rendering focuses on *difference* control variates, which estimate the difference between the integrand and an auxiliary function. We instead investigate the use of *ratio* control variates. There are many other variance reduction techniques in rendering [Doignies et al. 2024; Yan et al. 2024] that are orthogonal to our method.

Path tracing is inherently vector-valued, as it estimates the radiance for different color channels or wavelengths. Similarly, differentiable rendering computes a vector of parameter derivatives. While some previous works have used per-channel control variates [Hua et al. 2023; Müller et al. 2020], importance sampling methods either sample proportional to luminance or use a random mixture across wavelengths [Wilkie et al. 2014]. In general, both difference control variates and importance sampling struggle with vector-valued problems, as we will discuss further in Section 3.

Variance reduction in stochastic gradient descent. Stochastic gradient descent approximates parameter gradients using a vector-valued Monte Carlo sum estimator. Common variance reduction techniques such as importance sampling [Johnson and Guestrin 2018; Katharopoulos and Fleuret 2019; Salaün et al. 2023, 2024; Zhao and Zhang 2015] and control variates [Defazio et al. 2014; Johnson

Table 1. Summary of notations used throughout the paper.

f	A scalar-valued function.
\mathbf{f}	A vector-valued function, where $\mathbf{f}(x) = [f_1(x), \dots, f_M(x)]$.
F	The integral of scalar-valued function f , i.e. $F = \int f(x) \mathbf{d}\mathbf{x}$.
\mathbf{F}	The integral of vector-valued function \mathbf{f} , i.e. $\mathbf{F} = \int \mathbf{f}(x) \mathbf{d}\mathbf{x}$.
g	A probability density function for importance sampling.
h	An auxiliary function for various control variates.
H	The integral of h , i.e. $H = \int h(x) \mathbf{d}\mathbf{x}$.
$\hat{F}, \hat{\mathbf{F}}$	A Monte Carlo estimator of values F or \mathbf{F} .
\bar{f}, \bar{h}	The average of samples, e.g. $\bar{f} = \frac{1}{N} \sum_{i=1}^N f(X_i)$.
N	Random sample count.
M	Number of entries in a vector-valued function \mathbf{f} .
K	Number of auxiliary functions used for control variates.

and Zhang 2013; Schmidt et al. 2017] have also been used. Most of these techniques can only leverage limited prior information. In contrast, our work exploits rendering-specific prior knowledge such as BSDF sampling distributions.

Variance reduction in differentiable rendering. Physically-based differentiable rendering [Azinović et al. 2019; Gkioulekas et al. 2013; Li et al. 2018; Nimier-David et al. 2020; Vicini et al. 2021; Zhang et al. 2020] enables solving inverse light transport problems using gradient descent, and is inherently vector-valued as it estimates the gradient. Variance reduction techniques are used to reduce gradient variance [Belhe et al. 2024; Chang et al. 2023, 2024; Nicolet et al. 2021, 2023; Su and Gkioulekas 2024; Zeltner et al. 2021; Zhang et al. 2021a, 2020, 2021b]. While importance sampling [Belhe et al. 2024] can selectively reduce variance of individual partial derivatives, it is unable to simultaneously address all components of the gradient vector. We show that ratio control variates can achieve substantial variance reduction for the entire gradient vector.

Ratio estimator. Ratio estimators [Cochran 1977, Chapter 6] are a statistical tool for estimating the ratio of the means of two random variables, often in the context of finite populations. A well-known example is weighted importance sampling [Powell and Swann 1966; Spanier 1979; Spanier and Maize 1994], which is a biased variance-reduction technique for Monte Carlo integration. This estimator can be interpreted as a ratio control variate, which enables us to derive debiased versions, multi-auxiliary variants, and combinations with MIS techniques. Owen [2013, Chapter 8] briefly mentions ratio control variates, but without in-depth discussion or establishing the connection to weighted importance sampling.

In rendering, ratio estimators have been used for various purposes [Balázs et al. 2003; Bekaert et al. 2000; Fraboni et al. 2022; Stachowiak and Uludag 2015], e.g., to reduce variance of pixel filter evaluation [Pharr et al. 2023, Section 5.4.3] or direct illumination rendering [Heitz et al. 2018]. We advance both theory and applications of ratio estimators, including weighted importance sampling, in rendering and inverse rendering, particularly in the context of vector-valued integration. While Misso et al. [2022] discussed general debiasing strategies for biased rendering algorithms, we debias our ratio control variates using the Hartley-Ross estimator [1954], a specialized technique for ratio estimators that has not previously been used for rendering.

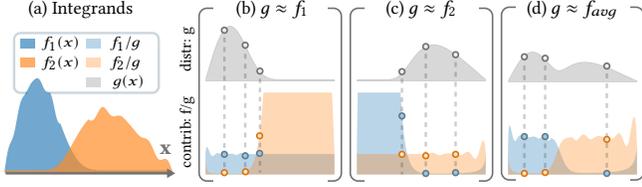


Fig. 3. *Limitations of using a shared sampling distribution.* (a) For integrands $[f_1, f_2]$, we show three choices of g and their corresponding *sample contributions* f_j/g , where a more uniform f_j/g indicates lower variance. (b) Choosing g close to f_1 makes f_1/g uniform but not f_2/g . (c) The opposite happens if g is closer to f_2 . (d) Choosing $g \approx f_{avg} = \frac{1}{2}(f_1 + f_2)$ keeps both f_1/g and f_2/g moderately uniform, but less so than using their respective better distributions in (b) and (c).

3 VECTOR-VALUED MONTE CARLO INTEGRATION

3.1 Problem formulation

The aforementioned applications motivate us to formulate a vector-valued Monte Carlo integration framework. We summarize the notation used throughout our derivations in Table 1.

Computing a vector-valued integral is equivalent to integrating M functions, $f_j : \Omega \rightarrow \mathbb{R}$, separately:

$$F_j = \int_{\Omega} f_j(x) dx, \quad \text{where } j = 1, \dots, M.$$

While it is possible to solve such a vector-valued problem using M separate, scalar-valued Monte Carlo estimators, the vector-valued formulation is meaningful if one of the following assumptions holds:

- (1) Given a sample X_i , evaluating one integrand function value $f_j(X_i)$ is nearly as costly as evaluating all M integrands.
- (2) Generating a sample X_i is significantly more expensive than evaluating the integrand functions $f_j(X_i)$.

If at least one of these assumptions holds, we would like to use each sample X_i for multiple integrals. As a result, the sampling distribution no longer necessarily corresponds to one specific integrand.

The applications that motivate our discussion all follow these assumptions: 1) In physically-based rendering, sampling a light path is costly and it is therefore advantageous to have each path contribute to all color channels. 2) In differentiable rendering, reverse-mode automatic differentiation [Griewank and Walther 2008] reuses intermediate computations, making gradient computation only marginally more expensive than evaluating a single partial derivative. Therefore, the following derivations assume a negligible cost of evaluating individual integrand functions and only reason about the total sample count N . We will keep using Fig. 2 as a motivating example to drive our discussion.

3.2 Importance sampling

Vector-valued importance sampling. We begin by analyzing the use of importance sampling for vector-valued integration problems. A straightforward solution is to draw all samples from a single importance sampling distribution $g(x)$, and use each sample X_i to evaluate all integrands f_j , as shown in Fig. 2(d):

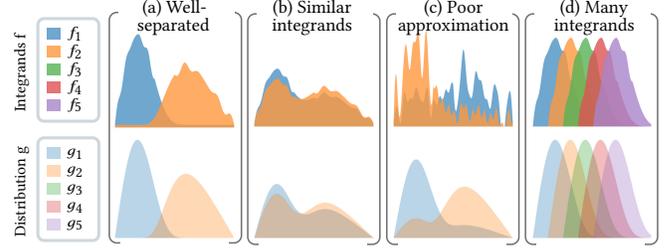


Fig. 4. *Limitations of entry-wise stratification.* Entry-wise stratification works well when (a) different integrands are well-separated. However, it becomes suboptimal if (b) the integrands are very similar, (c) the distributions poorly approximate the integrands, or (d) the integrand vector has too many entries, drastically decreasing the sample count N_j per estimator.

$$\hat{F}^{IS} = \left\{ \hat{f}_j^{IS} = \frac{1}{N} \sum_{i=1}^N \frac{f_j(X_i)}{g(X_i)} \right\}, \quad \text{with } j = 1, \dots, M. \quad (1)$$

The issue with this estimator is that $g(x)$ has to be suitable for all integrands f_j simultaneously. If the integrands have different shapes, no single g can be proportional to all of them, prohibiting zero variance estimation. Fig. 3 illustrates this limitation on an example problem. We derive the theoretically optimal sampling distribution and analyze its variance in the supplemental material.

Entry-wise stratification. Alternatively, for each integrand f_j , we can draw N_j samples from a specific distribution g_j and use these samples exclusively to evaluate f_j , as shown in Fig. 2 (e):

$$\hat{F}^{EWS} = \left\{ \hat{f}_j^{EWS} = \frac{1}{N_j} \sum_{i=1}^{N_j} \frac{f_j(X_{i,j})}{g_j(X_{i,j})} \right\}, \quad \text{with } j = 1, \dots, M, \quad (2)$$

where $N_1 + \dots + N_M = N$ to maintain an equal-cost comparison. We assume that all sampling techniques are of similar computational complexity.

Using a dedicated sampling distribution for each integrand can significantly reduce variance, but at the cost of a lower sample count per integrand. Each sample now contributes only to its corresponding integrand. Fig. 4 illustrates the limitations of this approach.

A dilemma of importance sampling. Using a single distribution allows sample reuse across all integrands, but g cannot precisely match all f . Entry-wise stratification circumvents this issue at the cost of a decreased sample count per integrand. We show in the supplementary material that neither strategy consistently outperforms the other, and both are often suboptimal.

3.3 Difference control variates

Scalar-valued difference control variates. We now turn to discussing standard control variates, which we will in the following refer to as *difference control variates (DCV)*. The scalar-valued difference control variates estimator is typically formulated as:

$$\hat{f}^{DCV} = \frac{1}{N} \sum_{i=1}^N f(X_i) - c \cdot (h(X_i) - H), \quad (3)$$

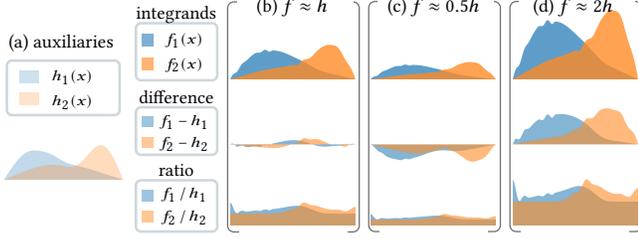


Fig. 5. *Impact of scaling factor on difference and ratio.* (a) We assume a set of fixed and normalized auxiliaries h_1 and h_2 . (b) When $f \approx h$, difference control variates has a uniform sample contribution, $f - h$, where more uniform is better. However, with a scaling factor, like (c) $f \approx 0.5h$ or (d) $f \approx 2h$, the difference becomes non-uniform (2nd row). In contrast, the sample contribution of importance sampling, i.e. the ratio f/h (3rd row), always remains uniform.

where h is an auxiliary function with a known integral H , and c is a constant to be carefully chosen. Analytical BRDF and light sampling distributions are often used as h , since they are normalized and thus $H = 1$, and c is regressed in a data-driven manner [Fan et al. 2006]. Data-driven methods may also leverage additional samples and data structures to adapt auxiliary functions to a specific scene, similar to path guiding. In this paper, we do not consider such methods and always set $c = 1$, similar to prior work [Heitz et al. 2018].

Unlike importance sampling, control variates naturally decouple sample reuse and auxiliary specialization. The corresponding vector-valued formulation is straightforward:

$$\hat{F}^{DCV} = \left\{ \hat{f}_j^{DCV} = \frac{1}{N} \sum_{i=1}^N f_j(X_i) - h_j(X_i) + H_j \right. \quad (4)$$

Shift and scale invariance. The mean square error (MSE) of a scalar-valued difference control variates with $c = 1$ is:

$$MSE[\hat{F}^{DCV}] = \frac{1}{N} \int (f(x) - h(x) - (F - H))^2 dx, \quad (5)$$

where for notational simplicity, and without loss of generality, we assume that the integration domain is the $[0, 1]$ interval. For a fixed h , the estimator maintains zero variance for any $f(x) = h(x) + a$, where a is an arbitrary constant. We call this property shift invariance. In contrast, for importance sampling:

$$MSE[\hat{F}^{IS}] = \frac{1}{N} \int \frac{(f(x) - F \cdot g(x))^2}{g(x)} dx. \quad (6)$$

The estimator maintains zero variance for any $f(x) = a \cdot g(x)$, since g is a normalized probability density and thus $F = \int a \cdot g(x) dx = a$. We call this property of importance sampling scale invariance.

Empirical evidence shows that scale invariance is much more important than shift invariance in forward and inverse rendering applications, since terms the contribution of a given light path is a product. Unfortunately, difference control variates are sensitive to scaling, as illustrated in Fig. 5. As a result, directly applying difference control variates with $c = 1$ may be suboptimal [Heitz et al. 2018], and often performs worse than the baseline (e.g., Fig. 10).

4 RATIO CONTROL VARIATES

We now introduce *ratio control variates* (RCV), which can significantly reduce variance for vector-valued integration problems. We first introduce a biased RCV estimator (Section 4.1) and later analyze its bias and introduce an unbiased version (Section 4.2).

4.1 Basic estimator

Ratio control variates approximate the integral F as $\widehat{F}/H \cdot H$, where \widehat{F}/H is an estimator of the ratio F/H and H is the known integral of an auxiliary function h . In contrast, the previously discussed difference control variates estimate F as $F \approx \widehat{F} - \widehat{H} + H$.

A straightforward ratio control variate estimator uses the *ratio of means* \widehat{f}/\widehat{h} to estimate F/H . Combining this with importance sampling, we get:

$$\hat{F}^{RCV} = H \cdot \frac{\sum_{i=1}^N f(X_i)/g(X_i)}{\sum_{i=1}^N h(X_i)/g(X_i)}. \quad (7)$$

The MSE of this estimator can be approximated as:

$$MSE[\hat{F}^{RCV}] \approx \frac{1}{N} \int \frac{(f(x) - (F/H) \cdot h(x))^2}{g(x)} dx + O\left(\frac{1}{N^2}\right). \quad (8)$$

For any a , if $f(x) = a \cdot h(x)$, then $F = a \cdot H$ and the numerator becomes zero. Therefore, ratio control variates satisfy the same scale invariance as importance sampling. We provide a full proof of Eq. (8) in the supplementary material.

Similar to difference control variates, specialized auxiliaries h_j can easily be applied to each integrand f_j for vector-valued integrals:

$$\hat{F}^{RCV} = \left\{ \hat{F}_j^{RCV} = H_j \cdot \frac{\sum_{i=1}^N f_j(X_i)/g(X_i)}{\sum_{i=1}^N h_j(X_i)/g(X_i)} \right. \quad (9)$$

Eq. (8) suggests that the ideal h_j for efficient ratio control variates should be proportional to the integrand f_j and have a known integral H_j . PDFs of importance sampling distributions meet these criteria well, as they are crafted to be proportional to the integrand and always integrate to 1. Therefore, our approach can easily incorporate the lessons learned from importance sampling:

- (1) We can directly use the PDFs of importance sampling distributions, such as BSDF sampling, as auxiliaries. This allows our RCV to be widely applied, with minimal modification to the renderer.
- (2) We can leverage techniques that improve importance sampling efficiency, such as multiple importance sampling and positivization, to make RCV more efficient as well.

More detailed discussions are provided in Section 5, where we introduce our extensions of RCV estimators to integrate them in forward and inverse renderers.

Relation to weighted importance sampling. If h is a normalized PDF, the ratio-of-means formulation in Eq. (7) is identical to weighted importance sampling. However, our RCV formulation has several advantages: it enables debiasing and using auxiliary functions that are not PDFs, or which require estimating H . Moreover, we believe that the general application to vector-valued integration problems is novel and not explicitly discussed in prior work on weighted importance sampling.

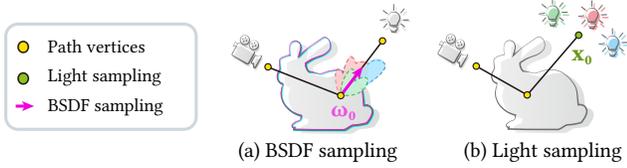


Fig. 6. *Vector-valued integrands in surface light transport.* (a) The BSRF may exhibit different responses across color channels. (b) Multiple lights are likely to have varying colors. Conventional luminance-based sampling methods cannot leverage chromatic information.

4.2 Bias analysis and debiasing

Bias analysis. The basic ratio control variates estimator is biased, but the bias decreases in $O(N^{-1})$ [Cochran 1977]:

$$\text{Bias}[\hat{F}^{RCV}] \approx \frac{F}{H} \left[\frac{\text{Var}[\hat{H}]}{H} - \frac{\text{Cov}[\hat{H}, \hat{F}]}{F} \right] + O\left(\frac{1}{N^2}\right). \quad (10)$$

Moreover, Hartley and Ross [1954] proved that the absolute bias ratio (ABR) is bounded by the coefficient of variation of $h(x)$:

$$\text{ABR}[\hat{F}^{RCV}] = \frac{|\text{Bias}[\hat{F}^{RCV}]|}{\sigma[\hat{F}^{RCV}]} \leq \frac{\sigma[\hat{H}]}{H}, \quad (11)$$

where $\sigma[X]$ is the standard deviation of a random variable X . This implies that the bias depends on how well-suited the sampling distribution g is to integrate h . The closer g is to being proportional to h , the smaller the bias. Hence, we in practice need a “safe” sampling distribution g to avoid extreme bias, e.g., a mixture of all $h_j(x)$.

Unbiased ratio estimator. The expectation of the ratio of means estimator \bar{f}/\bar{h} is intractable. Therefore, its bias can only be reduced, but not completely eliminated [Beale 1956; Quenouille 1956; Tin 1965]. As an alternative, Hartley and Ross [1954] proposed the *mean of ratios* estimator $\sum \frac{f(x_i)}{h(x_i)}$, which has a tractable expectation. Debiasing this estimator results in the following unbiased ratio control variates estimator:

$$\hat{F}^{HR} = \left\{ \frac{H_j}{N} \sum_{i=1}^N \frac{f_j(x_i)}{h_j(x_i)} + \frac{N}{N-1} \left[\frac{1}{N} \sum_{i=1}^N \frac{f_j(x_i)}{g(x_i)} - \frac{1}{N^2} \sum_{i=1}^N \frac{f_j(x_i)}{h_j(x_i)} \sum_{i=1}^N \frac{h_j(x_i)}{g(x_i)} \right] \right\}. \quad (12)$$

It has the same approximate MSE as the biased ratio estimator in Eq. (8) and can be used as a drop-in replacement. A minor limitation is that the sample count N must be greater than one to avoid division by zero. In Appendix A, we further discuss how to avoid numerical instabilities. In the supplementary material, we provide expanded discussions on vector-valued integration and additional variants of the RCV estimator.

5 APPLICATIONS

In practice, all known importance sampling distributions, including those for BSDFs, light sources, media, derivatives, and others, can be used as auxiliary functions for our ratio control variates, since their PDFs have a known integral of 1. This allows us to build on previous importance sampling research, improving vector-valued integral problems in surface light transport (Section 5.1), volumetric light transport (Section 5.2), and differentiable rendering (Section 5.3). In their respective subsections, we discuss our extensions of ratio

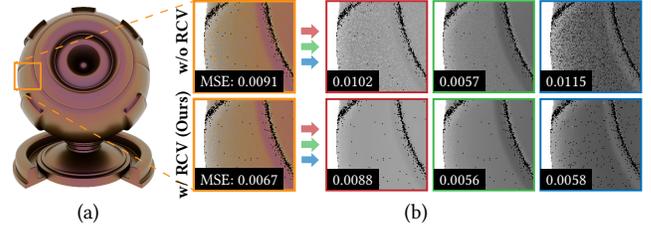


Fig. 7. *Wavelength-dependent BRDF.* (a) The IBIZA SUNSET material [Dupuy and Jakob 2018] has a wavelength-dependent BRDF. We render the material under direct illumination from a constant environment map, using only BRDF importance sampling and 1 sample per pixel (SPP). (b) We separately visualize each color channel both without (top row) and with our ratio control variates (bottom row).

control variates to address challenges when applying to multiple importance sampling, global illumination, and differentiable path tracing.

Implementation. All methods are implemented in a custom renderer using the Vulkan API with hardware-accelerated ray tracing. The results were rendered on a laptop with an NVIDIA GeForce RTX 3070 GPU.

5.1 Surface light transport

We apply our method to reduce variance on scenes that feature chromatic BSDFs and light sources. For this, we leverage existing BSDF and light sampling techniques as ratio control variates (Fig. 6).

Chromatic BSDF sampling. Simple analytical BSDF models usually have the same shape across color channels, as chromatic diversity is only introduced through a multiplicative constant (reflectance or albedo). However, some models like goniochromatic BSDFs [Belcour and Barla 2017] and measured BRDFs [Dupuy and Jakob 2018; Matusik et al. 2003] exhibit wavelength-dependent reflectance, thus the shape of each color channel may vary significantly.

As shown in Fig. 7, luminance-based importance sampling disproportionately focuses on the green channel², which leads to significant noise in the red and blue channels, even under constant lighting. Our RCV effectively reduce this noise, and requires only 0.03% more storage³ for RGL dataset and negligible runtime overhead.

Chromatic many-light sampling. Different lights in a scene often have different colors, yet the standard many-light sampling method typically also considers only luminance. We extend the light BVH sampling method [Conty Estevez and Kulla 2018] to support ratio control variates. It increases per-node storage from 28 bytes [Pharr et al. 2023] to 32 bytes⁴, enabling the computation of the PDF for each color channel.

As shown in Fig. 9, luminance-based sampling can cause severe noise even in scenes with extremely simple BRDFs, visibility and

²Luminance-based importance sampling is the suggested solution for the RGL dataset [Dupuy and Jakob 2018]; luminance = $0.2126 \times R + 0.7152 \times G + 0.0722 \times B$.

³Since we need to compute the PDF for each color channel, we must store the normalization constants for each data slice.

⁴We use two float16 to store the power of the U and V channels alongside the luminance (Y channel), so that we can compute PDF for each color channel.

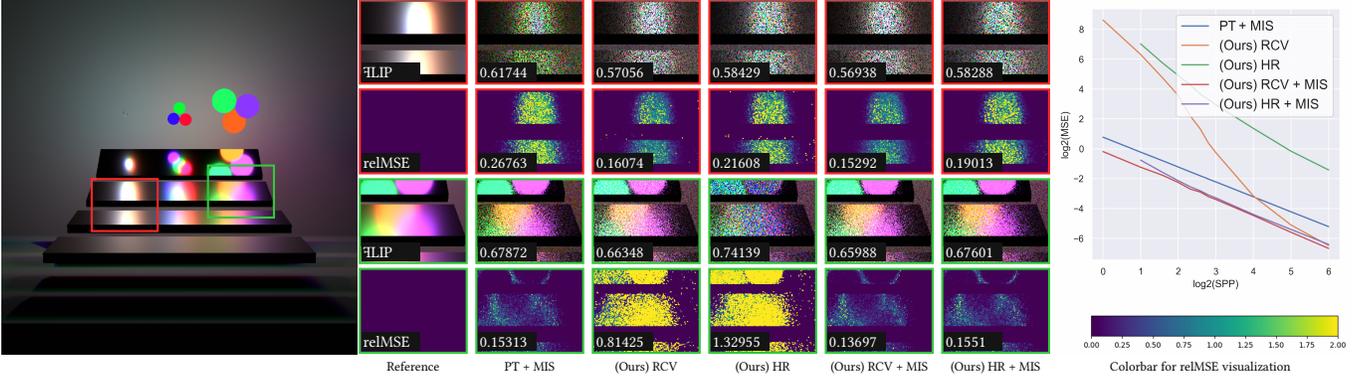


Fig. 8. *Multiple Importance Sampling (MIS)*: Our estimators are compatible with MIS. We use nine chromatic sphere lights to illuminate the VEACHMIS scene. We compare the visual differences (using $\bar{\Delta}$ LIP) and relative Mean Squared Error (relMSE) of standard path tracing (PT), ratio control variates, and the Hartley-Ross estimator applied to the entire integrands (RCV, HR), as well as methods applied only to the next event estimation sub-integrands (RCV-MIS, HR-MIS). All methods were rendered with 3 SPP and direct illumination only. Correctly combining MIS with the ratio estimator significantly reduces the error.

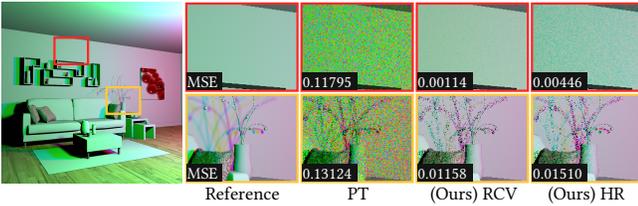


Fig. 9. *Three lights*. The BEDROOM scene is directly illuminated by three spherical area lights with RGB radiance values of [300, 4, 5], [4, 300, 5], and [4, 5, 300]. Our ratio control variates and Hartley-Ross estimators significantly reduce the noise, under the equal-sample-and-time comparison (2 SPP).

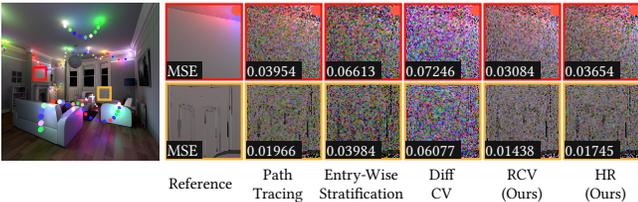


Fig. 10. *Many lights*. We apply our technique in combination with a light BVH to render the LIVINGROOM scene illuminated by over 100 chromatic sphere lights. In an equal-sample-and-time comparison (3 SPP), our estimators outperform the baseline and other variance reduction methods for vector-valued integration, including element-wise stratification and difference control variates.

lighting. Our ratio control variates and Hartley-Ross estimators significantly reduce the noise at negligible overhead.

In Fig. 10, we validate our approaches in a more challenging scene with over 100 lights, and further compare our methods with other vector-valued variance reduction techniques. Element-wise stratification has only 1/3 of the samples per channel compared to the other methods. This results in higher variance due to factors not captured by the sampling distributions (e.g., visibility). Difference control variates on the other hand suffer from significant noise due to the lack of scale invariance. As a result, both methods perform even worse than the baseline.

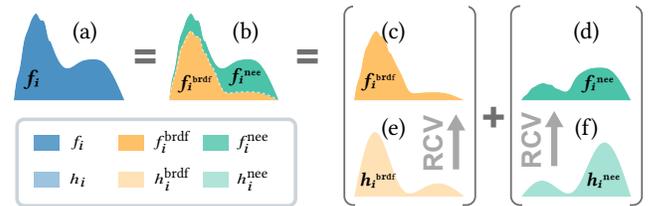


Fig. 11. *Combining MIS and RCV*. In surface light transport, we often have two sets of auxiliary functions, based on (e) BRDF and (f) NEE (next event estimation) sampling distributions. However, neither can independently match (a) the complete integrand accurately. (b) MIS decomposes the original integrand into a sum of sub-integrands (c) and (d), by multiplying with MIS weights. Each sub-integrand then aligns more closely with the corresponding auxiliaries, and can be independently estimated using RCV. While the example shows a single integrand, this approach extends seamlessly to vector-valued problems by applying the decomposition per entry.

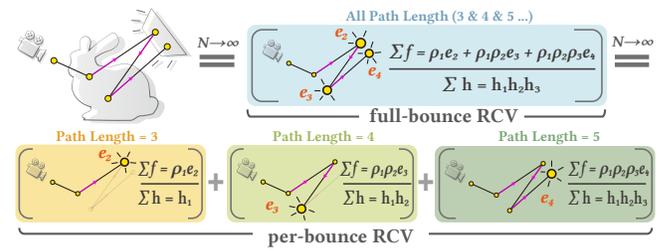


Fig. 12. *Full-bounce and per-bounce RCV*. After constructing paths via iterative BxDF importance sampling, ratio control variates can be applied either per-bounce or full-bounce. In full-bounce RCV, contributions from all path lengths are considered together. In per-bounce RCV, the integrands are decomposed into the sum of fixed-length path contributions, and RCV is applied separately to each. At a path vertex x_i , we denote BxDF throughput as ρ_i , the corresponding auxiliary function as h_i , and the emission as e_i .

Multiple importance sampling. In practice, BxDF and light sampling distributions each only partially describe the actual integrands.

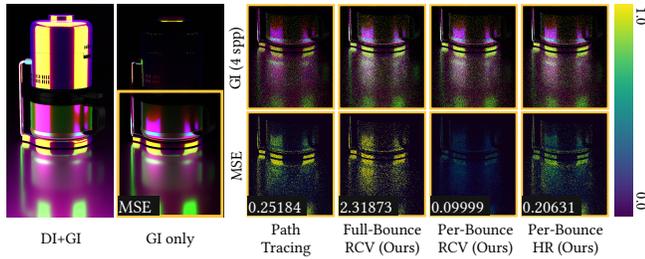


Fig. 13. *Full-bounce versus per-bounce RCV.* The COFFEE MAKER scene is rendered using GGX BRDFs with wavelength-dependent roughness and a maximum path length of 5. The per-bounce RCV and HR estimators improve over both path tracing and the full-bounce RCV.

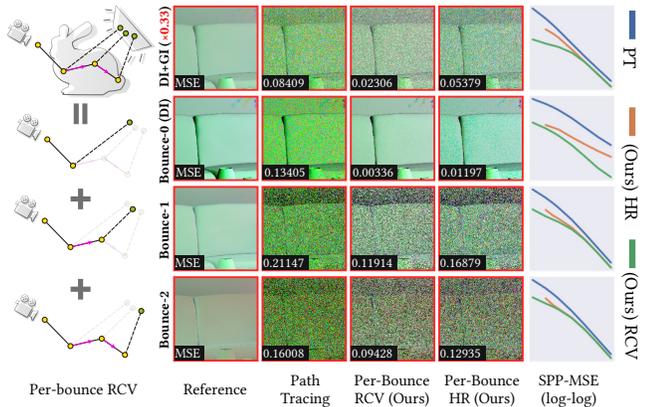


Fig. 14. *Per-bounce RCV with NEE.* Our per-bounce RCV can also be applied to next event estimation. We separately show contributions of paths of different lengths, demonstrating the benefits of our methods for multi-bounce indirect lighting. We adjusted the exposure of the images for optimal display.

Our estimators degrade if the used auxiliary function poorly matches the actual integrand. Fortunately, both ratio control variates and the Hartley-Ross estimator can leverage multiple importance sampling (MIS) [Veach 1998] to combine several auxiliaries. Fig. 11 illustrates this on a 1D example.

Multi-sample MIS decomposes the integrand into two parts by multiplying with MIS weights w_1 and w_2 . Each part can be evaluated using a corresponding auxiliary function that is more likely to match the MIS-weighted integrand:

$$\hat{F}_{MIS}^{Ratio} = \frac{\sum_{i=1}^N w_1(X_i) f(X_i) H_1}{\sum_{i=1}^N h_1(X_i)} + \frac{\sum_{i=1}^N w_2(X_i) f(X_i) H_2}{\sum_{i=1}^N h_2(X_i)},$$

Fig. 8 shows the results of using our chromatic Light BVH with and without MIS using balance heuristic weights. The specular regions are dominated by the BRDF, which is not captured by the next event estimation auxiliaries, resulting in high noise with our methods. However, by applying MIS, these regions are assigned to be handled by BRDF importance sampling, leading to significant improvements.

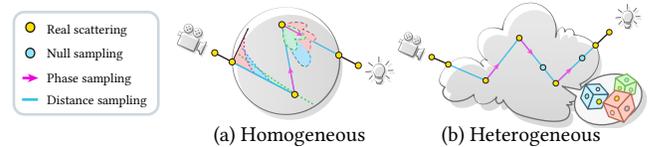


Fig. 15. *Vector-valued integrations in volumetric light transport.* (a) A homogeneous medium with wavelength-dependent absorption and scattering properties results in wavelength-dependent distance sampling, while the phase function may also vary across color channels. (b) Rendering heterogeneous media using delta tracking additionally requires sampling, potentially wavelength-dependent, discrete decisions at each path vertex.

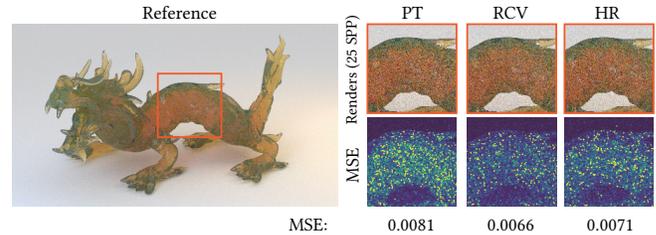


Fig. 16. *Homogeneous volume rendering.* Our approach can reduce the mean squared error (MSE) of homogeneous media. DRAGON uses a dielectric BSDF to enclose a homogeneous but chromatic medium, with property values of $\sigma_a = [0.05, 0.11, 0.74]$, $\sigma_s = [0.064, 0.45, 1.23]$, and $g = [0.3, 0.0, -0.2]$.

Path-space ratio control variates. For multi-bounce path samples, we can use the PDF of the entire path as the auxiliary. Since the path PDF always integrates to 1, all analysis remain valid. On the other hand, a naive recursive application of RCV would lead to an inconsistent estimator, as discussed in the supplemental material.

Full-bounce and per-bounce ratio control variates. A straightforward way to handle global illumination with path-space ratio control variates is to use the total radiance of the path as the integrand f , and the entire path PDF as the auxiliary h , referred to as the *full-bounce RCV*. However, as shown in Fig. 12, radiance can be contributed by each vertex along the path, not just the last one. Contributions from earlier vertices are independent of how subsequent vertices are sampled. As a result, the total path PDF does not correlate well with these terms and is thus not an effective auxiliary.

Alternatively, we can consider the sub-integral of contributions from a specific path length, and use the corresponding prefix path PDF as the auxiliary function for better correlation. This approach is referred to as the *per-bounce method*. Fig. 13 shows how per-bounce RCV and HR provide more reliable variance reduction compared to full-bounce variants.

The per-bounce formulation can easily be extended to next event estimation, as depicted by Fig. 14. It is also trivial to use MIS on each bounce as discussed in Fig. 11. Therefore, our methods are fully compatible with standard path construction techniques.

5.2 Volumetric light transport

Volumetric path tracing constructs light paths incrementally by sampling both phase function and free-flight distance. Rendering heterogeneous media using delta tracking [Miller et al. 2019; Woodcock et al. 1965] additionally requires discrete sampling of the type

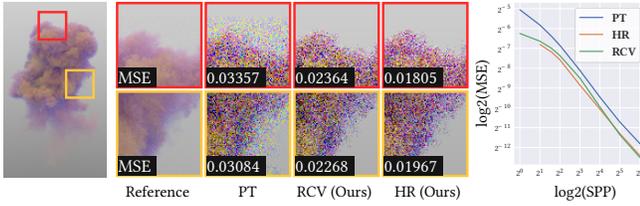


Fig. 17. *Heterogeneous volume rendering*. Our approaches can effectively handle heterogeneous media with wavelength-dependent properties. RCV and HR significantly reduce chromatic noise along the boundaries of heterogeneous media, even at low sample counts (8 SPP in the insets).

Table 2. Comparison of performance without and with our method

Scene	SPP	Baseline (ms)	Ours (ms)
IBIZA SUNSET (Fig. 7)	32	57.9	58.7
LIVING ROOM (Fig. 10)	32	97.4	97.6
GROUND EXPLOSION (Fig. 17)	4	221.1	221.2

of event at each interaction (i.e., absorption, scattering, or null scattering). All these factors can be wavelength-dependent, as shown in Fig. 15. The standard solution is hero-wavelength sampling [Wilkie et al. 2014], which samples a random mixture across color channels.

Using path-space ratio control variates, applying ratio control variates to volumetric light transport becomes straightforward. As shown in Fig. 16, our RCV significantly reduces the mean square error in homogeneous media without noticeable additional computational or memory costs.

Our methods also benefit heterogeneous media, as demonstrated in Fig. 17. In this example, the simple RCV suffers from visible bias at low sample counts. However, as shown in Fig. 18, this bias rapidly diminishes within approximately 16 SPP, thanks to the $1/N$ convergence rate.

Runtime overhead. In forward rendering, the main overhead of our method comes from evaluating h_i for each color channel. While this cost is negligible for analytical distributions, even tree-based methods like light BVH incur minimal overhead. This is because our method does not require additional tree traversal – it only reads a few more bytes at the leaf nodes. Quantitative results across test scenes confirm this efficiency (Table 2).

Memory overhead. For analytical distributions, there is no memory overhead, as the PDF can be computed directly. For tree-based distributions, the original data structures typically support evaluating the PDF only for luminance-based sampling, not for individual color channels. As a result, we precompute normalization constants for each channel, which introduces a small memory cost, as discussed in Section 5.1.

5.3 Differentiable rendering

We discuss two differentiable rendering applications in which ratio control variates can reduce the variance. One for when we have access to a differentiable path tracer (Section 5.3.1). One for when we do not have access to a differentiable renderer and have to rely

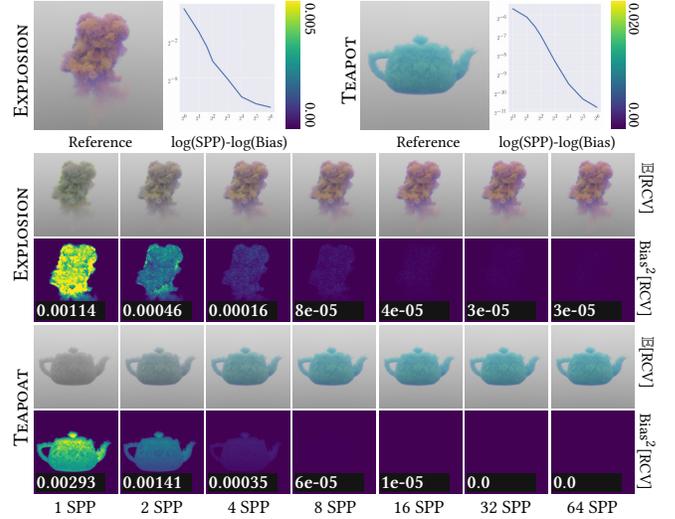


Fig. 18. *Bias convergence*. While the bias of RCV may be significant at very low sample counts, it decreases rapidly thanks to the $1/N$ convergence rate. We visualize the per-pixel squared bias, and provide its average value in the inset text.

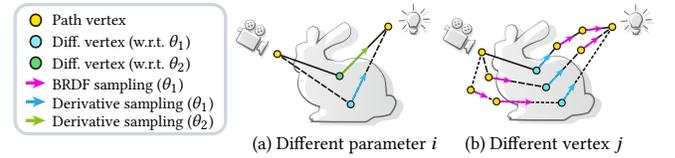


Fig. 19. *Vector-valued integration in differentiable path tracing*. Differentiable path tracing is a vector-valued problem due to gradient contributions of (a) different parameters θ_i and (b) different path vertices along a path. Ideally, different sampling distributions would be used for each integral.

on stochastic derivative-free optimization [Deliot et al. 2024; Fischer and Ritschel 2023; Wang et al. 2024].

5.3.1 Optimization via differentiable path tracing. In this section, we formulate gradient estimation in differentiable rendering as a vector-valued integration in the path space, with the vector-valued integrand having components both from different parameters and from different path vertices (Fig. 19). Applying ratio control variates to differentiable path tracing, however, is not trivial since the auxiliary functions we used are designed for the entire path space, and not for a specific parameter. Below, we show how our *marginal* ratio control variates can be used to extend importance sampling distributions to a stochastic auxiliary function to significantly reduce the variance. Finally, we discuss our *positivized ratio estimator* to handle the case when the auxiliary integral H is 0. We focus on the *interior term* in this discussion [Zhao et al. 2020], and leave discontinuity handling [Bangaru et al. 2020; Yan et al. 2022; Zhang et al. 2023] to future work.

Path space integration for gradient. Our goal is to derive the vector-valued integrand for which we can use our ratio control variates to reduce the variance. In inverse rendering, the derivative of the loss function \mathcal{L} with respect to one specific parameter θ_i (where

$i = 1, \dots, M)$ is:

$$\partial_{\theta_i} \mathcal{L} = \partial_1 \mathcal{L} \cdot \partial_{\theta_i} \mathbf{I} = \sum_{k=1}^K \partial_{I_k} \mathcal{L} \cdot \partial_{\theta_i} \mathbf{I}_k, \quad (13)$$

where \mathbf{I} is the rendered image, \mathbf{I}_k is one pixel of the image, and K is the number of pixels. The derivative $\partial_{\theta_i} \mathbf{I}_k$ can further be formulated as a path space integration [Nimier-David et al. 2020]. We consider the contributions from paths with different fixed path lengths N separately, where $N = 2, \dots, N^{\max}$:

$$\partial_{\theta_i} \mathbf{I}_k = \sum_{N=2}^{N^{\max}} \partial_{\theta_i} \mathbf{I}_k^N = \underbrace{\sum_{N=2}^{N^{\max}} \int_{\mathcal{X}_k^N} \partial_{\theta_i} \left[\prod_{j=0}^{N-1} f_j(\bar{\mathbf{x}}) \right] d\bar{\mathbf{x}}}_{\text{paths with a fixed length } N} \quad \text{all paths contributes to pixel } \mathbf{I}_k \quad (14)$$

where \mathcal{X}_k^N is the path space consisting of all paths $\bar{\mathbf{x}}$ contributing to pixel \mathbf{I}_k and have fixed path length N , and $f_j(\bar{\mathbf{x}})$ represents the forward radiance contribution from each path vertex \mathbf{x}_i :

$$f_i(\bar{\mathbf{x}}) = \begin{cases} W_e(\mathbf{x}_0, \omega_{\mathbf{x}_1, \mathbf{x}_0}) G(\mathbf{x}_0, \mathbf{x}_1), & \text{if } i = 0, \\ G(\mathbf{x}_i, \mathbf{x}_{i+1}) \rho_s(\mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{x}_{i+1}), & \text{if } 0 < i < N, \\ L_e(\mathbf{x}_N, \omega_{\mathbf{x}_N, \mathbf{x}_{N-1}}), & \text{if } i = N, \end{cases} \quad (15)$$

where W_e is the sensor importance, G is the geometry term, ρ_s is the BSDF, and L_e is the light emission. To further simplify notation, we introduce $f_{\theta_i}^*(\bar{\mathbf{x}}, j)$ to be the derivative contribution from path vertex \mathbf{x}_j to parameter θ_i ,

$$f_{\theta_i}^*(\bar{\mathbf{x}}, j) = f_0(\bar{\mathbf{x}}) f_1(\bar{\mathbf{x}}) \cdots f_N(\bar{\mathbf{x}}) \cdot \partial_{\theta_i} f_j(\bar{\mathbf{x}}) / f_j(\bar{\mathbf{x}}). \quad (16)$$

Therefore, we can further decompose the contribution from paths of fixed length N into contributions from each vertex along the path:

$$\partial_{\theta_i} \mathbf{I}_k^N = \int_{\mathcal{X}} \underbrace{[(\partial_{\theta_i} f_0) f_1 \cdots f_N]}_{f_{\theta_i}^*(\bar{\mathbf{x}}, 0)} + \cdots + \underbrace{[f_0 \cdots f_{N-1} (\partial_{\theta_i} f_N)]}_{f_{\theta_i}^*(\bar{\mathbf{x}}, N)} d\bar{\mathbf{x}} \quad (17)$$

$$= \int_{\mathcal{X}} [f_{\theta_i}^*(\bar{\mathbf{x}}, 0) + f_{\theta_i}^*(\bar{\mathbf{x}}, 1) + \cdots + f_{\theta_i}^*(\bar{\mathbf{x}}, N)] d\bar{\mathbf{x}}. \quad (18)$$

The derivative can then be computed as a sum over pixels, path lengths, and vertex indices:

$$\partial_{\theta_i} \mathcal{L} = \sum_{k=1}^K \sum_{N=2}^{N^{\max}} \sum_{j=0}^{N-1} \int_{\mathcal{X}_k^N} [\partial_{I_k} \mathcal{L} \cdot f_{\theta_i}^*(\bar{\mathbf{x}}, j)] d\bar{\mathbf{x}} \quad (19)$$

The integral above is vector-valued. The integrands $\partial_{I_k} \mathcal{L} \cdot f_{\theta_i}^*(\bar{\mathbf{x}}, j)$ vary for different pixel k , parameter θ_i , path vertex j , and path length N . In practice, however, storing a vector with K pixels, M parameters, and for a total of N^{\max} different path lengths each having N different vertices, is impractical. We therefore marginalize over all pixels and only store a vector for each parameter θ_i , each path vertex j , and each path length N [Chang et al. 2023]:

$$\partial_{\theta_i}^{N,j} \mathcal{L} = \sum_{k=1}^K \int_{\mathcal{X}_k^N} [\partial_{I_k} \mathcal{L} \cdot f_{\theta_i}^*(\bar{\mathbf{x}}, j)] d\bar{\mathbf{x}}. \quad (20)$$

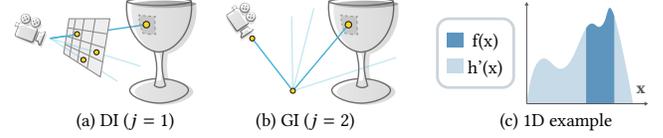


Fig. 20. *Inefficiency of naive auxiliary function.* Only a subset of sampled light paths will hit a specific texel and contribute to its derivative. (a) For direct illumination, paths from most pixels can never contribute to a given target texel and (b) a similar scenario arises for global illumination. (c) We illustrate the intuition in 1D: while the auxiliary function aligns well with the integrand where it is non-zero, the integrand’s domain is much narrower, as most samples fail to touch the target texel being optimized. This makes h' a bad auxiliary function.

Marginal ratio control variates. We now want to design an auxiliary function h for the vector-valued integrand in Eq. (20). For BSDF derivatives, when pixel k , parameter i , path length N , and differential vertex index j are all fixed, importance sampling of the integrands $f_{\theta_i}^*(\bar{\mathbf{x}}, j)$ is known: by using BSDF derivative sampling [Belhe et al. 2024] at \mathbf{x}_j and using BSDF sampling for other vertices. We denote the PDF of such incremental path construction schema as:

$$h'_{(\theta_i, j)}(\bar{\mathbf{x}}) = p_{(\theta_i, j)}(\mathbf{x}_0) \cdot p_{(\theta_i, j)}(\mathbf{x}_1) \cdots p_{(\theta_i, j)}(\mathbf{x}_N), \quad (21)$$

where the local probability density $p_{(\theta_i, j)}(\mathbf{x}_0)$ at each path vertex \mathbf{x}_j is:

$$p_{(\theta_i, j)}(\mathbf{x}_m) \propto \begin{cases} W_e(\mathbf{x}_0, \omega_{\mathbf{x}_1, \mathbf{x}_0}) G(\mathbf{x}_0, \mathbf{x}_1), & \text{if } m = 0, \\ G(\mathbf{x}_m, \mathbf{x}_{m+1}) \rho_s(\mathbf{x}_{m-1}, \mathbf{x}_m, \mathbf{x}_{m+1}), & \text{if } 0 < m < N, m \neq j \\ G(\mathbf{x}_i, \mathbf{x}_{m+1}) \partial_{\theta_i} \rho_s(\mathbf{x}_{m-1}, \mathbf{x}_m, \mathbf{x}_{m+1}), & \text{if } 0 < m < N, m = j \\ 1, & \text{if } m = N, \end{cases}$$

which accounts for every term except the emission in Eq. (16).

However, directly using h' (Eq. (21)) as auxiliary is very inefficient. Fig. 20 reveals the reason for that: not all path generated by h' will contribute to the integrand $f_{\theta_i}^*(\bar{\mathbf{x}}, j)$. For example, in texture optimization, a path that does not hit the target texel at \mathbf{x}_j will never contribute to its derivative. Such paths can frequently be generated using incremental construction with BSDF and BSDF derivative sampling. This fact makes $h'(x)$ a very poor auxiliary function to describe actual integrand $f_{\theta_i}^*(\bar{\mathbf{x}}, j)$.

To alleviate this, we aim to construct a more accurate auxiliary:

$$h_{\theta_i, j}(\bar{\mathbf{x}}) = p(\mathbf{x}_0) \cdot p(\mathbf{x}_1) \cdots p(\mathbf{x}_N) \cdot \mathbb{I}(\mathbf{x}_j, \theta_i), \quad (22)$$

where $\mathbb{I}(\bar{\mathbf{x}}, \theta_i)$ is an indicator function that returns 1 if vertex \mathbf{x}_j actually touches θ_i , and 0 otherwise. $h_{(\theta_i, j)}$ is far more precise than $h'_{(\theta_i, j)}$, but we do not know its integral $H_{(\theta_i, j)}$.

To use ratio control variates, we need to estimate H using another estimator \hat{H} , making the new estimator to be $\hat{H} \cdot \hat{f} / \hat{h}$. Specifically, if we use the sample average \hat{h} to estimate H , it cancels out the denominator, leaving only \hat{f} . To make it meaningful, we need an estimator \hat{H} more accurate than \hat{h} .

As illustrated in Fig. 21, our key idea to estimate the integration \hat{H} is to numerically integrate on the prefix of the path before vertex \mathbf{x}_j , and analytically marginalize the suffix after the vertex. The

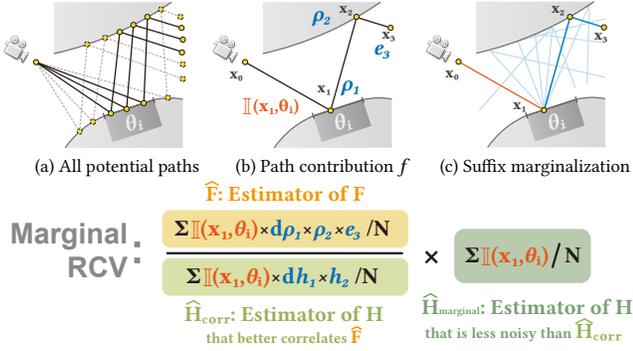


Fig. 21. *Marginal ratio control variates*. (a) We optimize the parameter θ_i for a single texel, focusing on paths of length $N = 3$ with contributions solely from the path vertex x_1 . Only paths that hit the texel at x_1 contribute non-zero values to the gradient. (b) The path contribution consists of the BRDF derivative $d\rho_1$, the indicator $\mathbb{I}(x_1, \theta_i)$, the BRDF ρ_2 , and the emission e_3 . There are corresponding auxiliaries dh_1 for derivative $d\rho_1$ and h_2 for BRDF ρ_2 , and the path-space auxiliary $dh_1 \cdot \mathbb{I}(x_1, \theta_i) \cdot h_2$ correlates the path contribution f well. (c) By partitioning the path at vertex x_j , we observe the auxiliary function of *suffix path* has no indicator function and thus always integrates to 1 since we use PDFs as auxiliaries. For estimating H , the suffix only introduces more noise, so we analytically marginalize the suffix and numerically integrate only the *prefix path*. This approach yields a less noisy estimator for H . Dividing the better-correlated estimator of H and multiplying the marginalized estimator of H forms our *marginal RCV*.

marginalization can be written as:

$$H_{(\theta_i, j)} = \int p(\mathbf{x}_0) \cdot p(\mathbf{x}_1) \cdots p(\mathbf{x}_N) \cdot \mathbb{I}(\mathbf{x}_j, \theta_i) d\bar{\mathbf{x}} \quad (23)$$

$$= \int p(\mathbf{x}_0) \cdots p(\mathbf{x}_{j-1}) \cdot \mathbb{I}(\mathbf{x}_j, \theta_i) d\mathbf{x}_0 \cdots \mathbf{x}_{j-1} \quad (24)$$

$$\cdot \int p(\mathbf{x}_j) \cdots p(\mathbf{x}_N) d\mathbf{x}_j \cdots \mathbf{x}_N \quad (25)$$

$$= \int p(\mathbf{x}_0) \cdot p(\mathbf{x}_1) \cdots p(\mathbf{x}_{j-1}) \cdot \mathbb{I}(\mathbf{x}_j, \theta_i) d\bar{\mathbf{x}}. \quad (26)$$

For a path $\bar{\mathbf{x}}$, whether its j -th vertex \mathbf{x}_j touches θ_i is only relevant to the prefix path $\mathbf{x}_0 \mathbf{x}_1 \cdots \mathbf{x}_{j-1} \mathbf{x}_j$ and irrelevant to the suffix path $\mathbf{x}_j \cdots \mathbf{x}_N$. Therefore, we can extract the integration of the suffix path out, as in Eq. (25), whose integral is known to be 1.

The remaining prefix integral Eq. (26) is a better estimator of H than \bar{h} (Eq. (23)). This variance reduction method by marginalizing suffix is known as conditioning [Owen 2013] or Rao-Blackwellization [Blackwell 1947]. Therefore, we use a Monte Carlo estimator of Eq. (26) as \hat{H} and still leave Eq. (23) in denominator as it correlated the numerator better.

Positized ratio estimator. For ratio control variates, $\bar{f}/\bar{h} \cdot H$, if the auxiliary integral H is 0, applying ratio control variates results in a division by zero. For example, Belhe et al. [2024] showed that the derivatives of all microfacet normal distribution functions (NDF) integrate to zero, so we cannot directly use the NDF derivative as a valid auxiliary. As shown in Fig. 22, to address this, we decompose

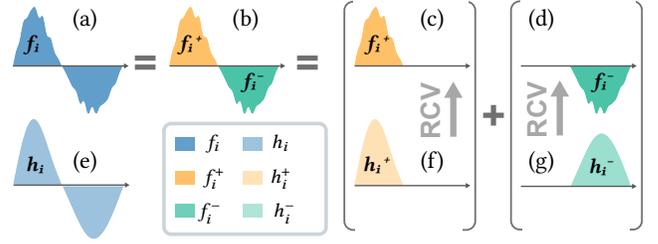


Fig. 22. *Positized RCV*. In cases where (e) the auxiliary function h has a known integral $H = 0$, RCV becomes meaningless. To address this, we can decompose h into (f) positive and (g) negative components, so that each part has a non-zero integral. We then apply the same decomposition to (a) integrands as well and apply our RCV separately. While the example shows a single integrand, this approach extends seamlessly to vector-valued problems by applying the decomposition per entry.

the integrands into positive and negative parts [Owen and Zhou 2000], f^+ and f^- , and estimate each part separately with ratio control variates, using positized sampling distributions, h^+ and h^- , as auxiliaries:

$$\hat{F}_{\text{Pos}}^{\text{Ratio}} = \frac{\sum_{i=1}^N f^+(X_i)}{\sum_{i=1}^N h^+(X_i)} \cdot H^+ + \frac{\sum_{i=1}^N f^-(X_i)}{\sum_{i=1}^N h^-(X_i)} \cdot H^-$$

Despite $H^+ + H^- = 0$, since $H^+ \neq 0$ and $H^- \neq 0$, the ratio control variates will work as usual for each term. We can easily extend this to any *single-signed decomposition* [Belhe et al. 2024].

Texture optimization with direct illumination. `rveectMarginal` and positized variants provide the theoretical underpinning for using RCV to differentiable rendering. Now, we consider optimizing a mixture BRDF with three parameters ($M = 3$):

$$f = \mathbf{w} \cdot \text{Oren-Nayar}(\sigma) + (1 - \mathbf{w}) \cdot \text{Isotropic-GGX}(\alpha), \quad (27)$$

where w is the mixture weight, σ is Oren-Nayar roughness and α is isotropic GGX roughness. Belhe et al. [2024] described methods for importance sampling the derivative of each parameter individually. Applying positization to $\partial\alpha$ and mixture decomposition to ∂w and $\partial\sigma$, we obtain six integrands and six corresponding auxiliaries.

In Fig. 23, we compare path replay backpropagation (PRB) [Vicini et al. 2021], entry-wise BRDF derivative importance sampling [Belhe et al. 2024] and PRB with our ratio control variates, under direct illumination. The results demonstrate robust improvements achieved by our method across various settings. Under direct illumination, whether a path will touch a texel is determined solely by camera rays, as shown in Fig. 20 (a). In this case, the variance in \hat{H} is entirely due to the random sampling of the pixel footprint.

Texture optimization with global illumination. To further highlight the marginal RCV's ability to handle stochastic \hat{H} estimation and multiple bounces, we compare our methods with PRB under 1-bounce global illumination, incorporating indirect contributions. As illustrated in Fig. 20 (b), the estimation of Eq. (26) becomes stochastic in this scenario. Fig. 24 shows that our methods can improve upon the PRB baseline across various settings.

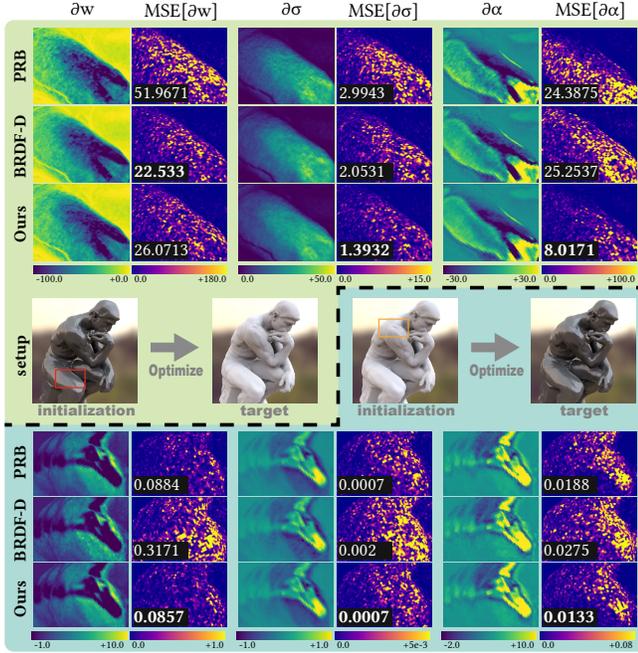


Fig. 23. *Texture optimization under direct illumination.* We optimize a 512^2 texture that stores all three parameters of the mixture material, under equal sample comparison (80 primal and 12 gradient SPP). When the initial state is glossy, the gradient variance is dominated by the BRDF derivative, so our approach significantly improves upon PRB. When the initial state is diffuse, the gradient variance is dominated by lighting and visibility, and our approach becomes much better than entry-wise BRDF derivative sampling, which effectively have only $1/3$ samples per parameter. In most cases, our approach outperforms both baselines.

Performance. In forward rendering, applying ratio control variates to per-pixel integrands incurs only a few additional arithmetic operations that can efficiently be carried out in registers. However, in differentiable rendering, auxiliary information needs to be accumulated per parameter and using atomic additions. In Fig. 24, the use of RCV incurs an overhead of around 29% for each gradient sample, compared to the PRB baseline. However, the overall gradient computation time remains almost identical, since we take significantly more primal samples (80 SPP) than gradient samples (1 SPP). This is a common strategy to reduce the influence of primal rendering noise [Azinović et al. 2019]. We believe more advanced optimization of atomic accumulation [Bangaru et al. 2023, Section 4.3] or a parameter-oriented formulation [Chang et al. 2023; Nimier-David et al. 2021] could reduce the overhead of RCV.

Table 3. Comparison of performance on Fig. 24

Setting	Forward	Inverse (PRB)	Inverse (Ours)
$\alpha = 5e^{-3}$	56 ms / 80 SPP	2.6 ms / 1 SPP	3.4 ms / 1 SPP
$\alpha = 5e^{-2}$	96 ms / 80 SPP	4.1 ms / 1 SPP	6.1 ms / 1 SPP

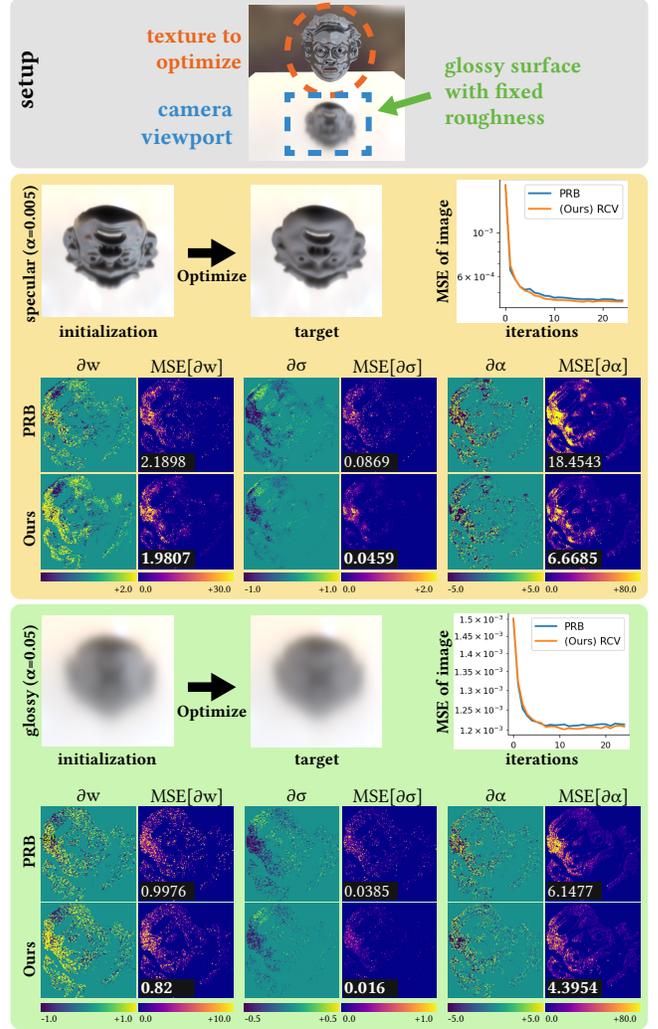


Fig. 24. *Texture optimization under indirect illumination.* We optimize a 512^2 texture that stores all three parameters of the mixture material, under equal sample comparison (80 primal and 1 gradient SPP). All derivatives are contributed from the indirect bounces over the glossy panel. Increasing the specular roughness of the panel increases the noise of our estimation of H (Eq. (26)). However, thanks to marginalization, our method can achieve improvements across various specularities of the panel. While the improvement on the final loss value is not significant on this scene, we believe that variance reduction will benefit future use of inverse rendering.

Memory overhead. Applying our method in inverse path tracing requires an additional buffer to accumulate auxiliary values, especially when positization is needed. In that case, each parameter stores five floating-point values: f^+ , f^- , h^+ , h^- , h^{marginal} .

5.3.2 Stochastic derivative-free optimization. When a differentiable renderer is not accessible, we may still have a renderer capable of generating an image $\mathbf{I}(\theta)$ for arbitrary parameters θ to compute a loss $\mathcal{L}(\mathbf{I}(\theta))$. In this case, we can stochastically estimate the derivatives using ideas from derivative-free optimization [Rechenberg

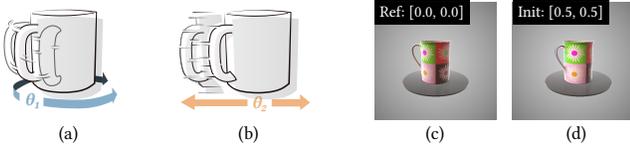


Fig. 25. *Optimization of a rigid transformation.* We optimize both (a) the rotation along the y-axis (θ_1) and (b) the translation along the x-axis (θ_2) of a mug. (c) The reference image. (d) The initial state.

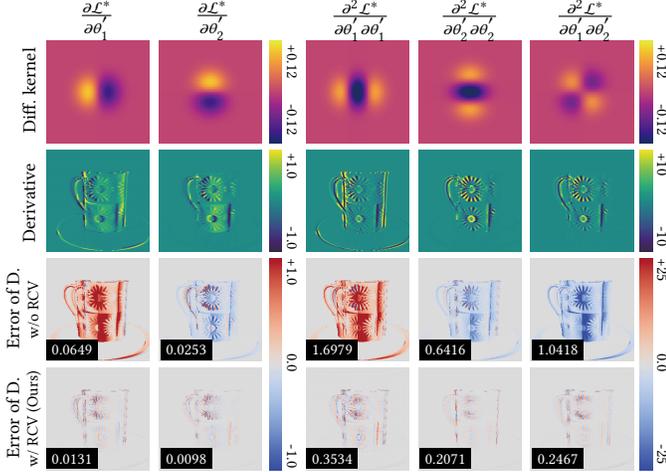


Fig. 26. *RCV for first- and second-order derivatives estimation.* With two parameters to optimize, the rendering loss has to be convolved with five different kernels to estimate the first- and second-order derivatives. Our RCV significantly reduces the mean absolute error of these derivatives (as reported in the inset text), compared to importance sampling using the average of distributions corresponding to these kernels.

and Eigen 1973; Spall 1992; Staines and Barber 2012; Wierstra et al. 2014], by smoothing the loss through perturbing parameters:

$$\min_{\theta} \mathcal{L}(\mathbf{I}(\theta)) \leq \min_{\theta'} E_{\theta \sim p(\theta|\theta')} [\mathcal{L}(\mathbf{I}(\theta))] = \min_{\theta'} \mathcal{L}^*(\theta'), \quad (28)$$

where $p(\theta|\theta')$ is usually the density a multivariate Gaussian distribution whose mean is parameterized by θ' , and the covariance is usually set to $\sigma^2 \cdot I$ for some standard deviation σ .

Previous work [Deliot et al. 2024; Fischer and Ritschel 2023; Le Lidec et al. 2021; Wang et al. 2024] has applied the idea to estimate the derivatives of rendering. Both first- and second-order derivatives of the Gaussian mean θ' can be estimated unbiasedly without ever having access to the rendering derivative:

$$\nabla_{\theta'} \mathcal{L}^* = \int_{\Theta} \nabla_{\theta'} p(\theta|\theta') \mathcal{L}(\mathbf{I}(\theta)) d\theta, \quad (29)$$

$$\nabla_{\theta'}^2 \mathcal{L}^* = \int_{\Theta} \nabla_{\theta'}^2 p(\theta|\theta') \mathcal{L}(\mathbf{I}(\theta)) d\theta, \quad (30)$$

where Θ is the domain of the parameters and $\nabla_{\theta'}^2$ is the operator that computes the Hessian matrix with respect to θ' . These derivatives can then be used in a gradient-based optimizer (e.g., Adam) or even a second-order optimizer (e.g., Newton’s method).

The integrals above are vector-valued, and thus the variance can be reduced using our ratio control variates. We apply our ratio

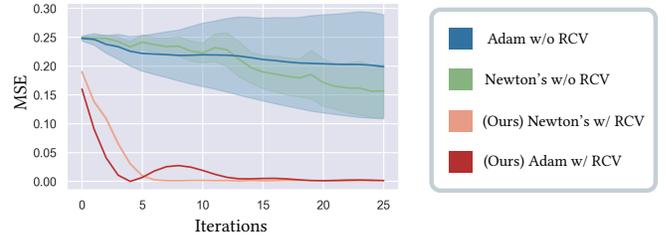


Fig. 27. We plot the mean squared error (MSE) of the optimized parameters $[\theta'_1, \theta'_2]$ over optimization iterations, with 16 images rendered per iteration to estimate gradients and Hessians. The shaded region indicates the variance of the MSE across 50 optimization runs, while the solid line is the mean. We compare both first-order optimization (using Adam) and second-order optimization (using Newton’s method), demonstrating that RCV significantly enhances the efficiency and robustness of the optimization process.

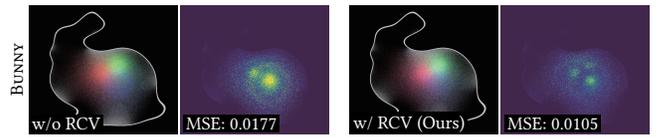


Fig. 28. *RCV on Walk on Sphere.* We apply RCV to sample chromatic source terms in the Walk on Sphere method, using 20 walks per pixel, achieving significantly reduced MSE.

control variates on Wang et al. [2024]’s positized distributions. We apply our estimators to optimize the rigid pose of an object, see Fig. 25 for the problem setup. Fig. 26 and Fig. 27 show that our ratio control variates significantly reduce variance.^{5 6}

5.4 Discussion and future work

Applications beyond rendering. Our techniques may also benefit applications beyond rendering. For example, Monte Carlo Walk on Spheres (WoS) [Sawhney and Crane 2020] shares many similarities with rendering. As illustrated in Fig. 28, we show that RCV could be beneficial to sample chromatic source terms, and it should be promising to further extend to vector-valued Green’s functions and chromatic spatially-varying coefficients [Sawhney et al. 2022].

Accuracy of auxiliary functions. Well-correlated auxiliary functions are important for RCV to be helpful. In the supplementary material, we present a method to estimate the confidence of auxiliary functions, allowing interpolation between vanilla Monte Carlo and RCV based on confidence levels. However, this approach requires estimating additional constants through regression, making it less effective at low sample counts. It is also interesting to investigate how to combine RCV with RIS-based sampling methods, which have no explicit PDF formula, such as ReSTIR [Bitterli et al. 2020].

Other quantitative error metrics. Ratio control variates has been proven to be efficient in reducing MSE. We show that in practice it may achieve improvements under FLIP, relative MSE (Fig. 8),

⁵To handle non-convex loss landscape, we modify the Hessian to ensure it is positive-definite, and clamp the gradient step length to prevent excessively large values.

⁶While our approach targets variance reduction, it is orthogonal to and does not diminish the contributions of Wang et al. [2024], which enable higher-order differentiability.

and L_1 error (Fig. 26), but these are less guaranteed. Since MSE is not always the best metric for rendering and gradients, developing vector-valued variance reduction techniques tailored to other metrics would be an interesting direction for future work.

6 CONCLUSION

We formalize vector-valued integration as a common pattern in forward and inverse rendering. We found that ratio control variates are a powerful and underexplored tool to reduce variance for such problems. Our methods can directly leverage existing importance sampling distributions to construct both consistent and unbiased estimators. Not only are ratio control variates applicable to a wide range of rendering problems, but their implementation is most often very simple and has negligible computational and memory overheads. We hope that our methods can benefit other applications of vector-valued Monte Carlo integration in computer graphics and scientific computing.

ACKNOWLEDGMENTS

This work was funded Google and Adobe's gifts, NSF IIS grants 2105806 and 2238839 and an NSERC Fellowship. We thank the anonymous reviewers for their constructive feedback, and Yash Belhe and Vishesh Gupta for proofreading. Scene assets are courtesy of Sketchfab users yellokab, n-malMBERG, yugengen, Riggers; Shadertoy user Marcel Fernandez; and Benedikt Bitterli.

REFERENCES

- James Arvo. 1995. Stratified sampling of spherical triangles. In *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '95)*. Association for Computing Machinery, New York, NY, USA, 437–438. <https://doi.org/10.1145/218380.218500>
- Dejan Azinović, Tzu-Mao Li, Anton Kaplanyan, and Matthias Nießner. 2019. Inverse Path Tracing for Joint Material and Lighting Estimation. In *Computer Vision and Pattern Recognition*. 2447–2456.
- Benedek Balázs, László Szirmay-Kalos, and Antal György. 2003. Weighted importance sampling in shooting algorithms. In *Proceedings of the 19th Spring Conference on Computer Graphics (Budmerice, Slovakia) (SCCG '03)*. Association for Computing Machinery, New York, NY, USA, 177–184. <https://doi.org/10.1145/984952.984981>
- Sai Praveen Bangaru, Tzu-Mao Li, and Frédo Durand. 2020. Unbiased warped-area sampling for differentiable rendering. *ACM Trans. Graph.* 39, 6, Article 245 (Nov. 2020), 18 pages. <https://doi.org/10.1145/3414685.3417833>
- Sai Praveen Bangaru, Lifan Wu, Tzu-Mao Li, Jacob Munkberg, Gilbert Bernstein, Jonathan Ragan-Kelley, Frédo Durand, Aaron Lefohn, and Yong He. 2023. SLANG.D: Fast, Modular and Differentiable Shader Programming. *ACM Trans. Graph.* 42, 6, Article 264 (Dec. 2023), 28 pages. <https://doi.org/10.1145/3618353>
- Evelyn Martin Lansdowne Beale. 1956. Some use of computers in operational research. 31 (1956), 27–28.
- Philippe Bekaert, Mateu Sbert, and Yves D. Willems. 2000. Weighted Importance Sampling Techniques for Monte Carlo Radiosity. In *Proceedings of the Eurographics Workshop on Rendering Techniques 2000*. Springer-Verlag, Berlin, Heidelberg, 35–46.
- Laurent Belcour and Pascal Barla. 2017. A practical extension to microfacet theory for the modeling of varying iridescence. *ACM Trans. Graph.* 36, 4, Article 65 (July 2017), 14 pages. <https://doi.org/10.1145/3072959.3073620>
- Yash Belhe, Bing Xu, Sai Praveen Bangaru, Ravi Ramamoorthi, and Tzu-Mao Li. 2024. Importance Sampling BRDF Derivatives. *ACM Trans. Graph.* 43, 3, Article 25 (apr 2024), 21 pages. <https://doi.org/10.1145/3648611>
- Benedikt Bitterli, Chris Wyman, Matt Pharr, Peter Shirley, Aaron Lefohn, and Wojciech Jarosz. 2020. Spatiotemporal reservoir resampling for real-time ray tracing with dynamic direct lighting. *ACM Trans. Graph.* 39, 4, Article 148 (Aug. 2020), 17 pages. <https://doi.org/10.1145/3386569.3392481>
- David Blackwell. 1947. Conditional Expectation and Unbiased Sequential Estimation. *The Annals of Mathematical Statistics* 18, 1 (1947), 105–110. <http://www.jstor.org/stable/2236107>
- Wesley Chang, Venkataram Sivaram, Derek Nowrouzezahrai, Toshiya Hachisuka, Ravi Ramamoorthi, and Tzu-Mao Li. 2023. Parameter-space ReSTIR for Differentiable and Inverse Rendering. In *ACM SIGGRAPH 2023 Conference Proceedings* (Los Angeles, CA, USA) (*SIGGRAPH '23*). Association for Computing Machinery, New York, NY, USA, Article 18, 10 pages. <https://doi.org/10.1145/3588432.3591512>
- Wesley Chang, Xuanda Yang, Yash Belhe, Ravi Ramamoorthi, and Tzu-Mao Li. 2024. Spatiotemporal Bilateral Gradient Filtering for Inverse Rendering. In *SIGGRAPH Asia 2024 Conference Papers* (Tokyo, Japan) (*SA '24*). Association for Computing Machinery, New York, NY, USA, Article 70, 11 pages. <https://doi.org/10.1145/3680528.3687606>
- William G Cochran. 1977. Sampling techniques. *John Wiley & Sons* (1977).
- Alejandro Conty Estevez and Christopher Kulla. 2018. Importance Sampling of Many Lights with Adaptive Tree Splitting. *Proc. ACM Comput. Graph. Interact. Tech.* 1, 2, Article 25 (Aug. 2018), 17 pages. <https://doi.org/10.1145/3233305>
- Miguel Crespo, Adrian Jarabo, and Adolfo Muñoz. 2021. Primary-space Adaptive Control Variates Using Piecewise-polynomial Approximations. *ACM Trans. Graph.* 40, 3, Article 25 (July 2021), 15 pages. <https://doi.org/10.1145/3450627>
- Aaron Defazio, Francis Bach, and Simon Lacoste-Julien. 2014. SAGA: a fast incremental gradient method with support for non-strongly convex composite objectives. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 1* (Montreal, Canada) (*NIPS'14*). MIT Press, Cambridge, MA, USA, 1646–1654.
- Thomas Deliot, Eric Heitz, and Laurent Belcour. 2024. Transforming a Non-Differentiable Rasterizer into a Differentiable One with Stochastic Gradient Estimation. *Proc. ACM Comput. Graph. Interact. Tech.* 7, 1, Article 3 (May 2024), 16 pages. <https://doi.org/10.1145/3651298>
- Bastien Doignies, David Coeurjolly, Nicolas Bonneel, Julie Digne, Jean-Claude Iehl, and Victor Ostromoukhov. 2024. Differentiable Owen Scrambling. *ACM Trans. Graph.* 43, 6, Article 255 (Nov. 2024), 12 pages. <https://doi.org/10.1145/3687764>
- Jonathan Dupuy and Anis Benyoub. 2023. Sampling Visible GGX Normals with Spherical Caps. arXiv:2306.05044 [cs.GR] <https://arxiv.org/abs/2306.05044>
- Jonathan Dupuy and Wenzel Jakob. 2018. An Adaptive Parameterization for Efficient Material Acquisition and Rendering. *Transactions on Graphics (Proceedings of SIGGRAPH Asia)* 37, 6 (Nov. 2018), 274:1–274:18. <https://doi.org/10.1145/3271217.3275059>
- Kenta Eto and Yusuke Tokuyoshi. 2023. Bounded VNDF Sampling for Smith-GGX Reflections. In *SIGGRAPH Asia 2023 Technical Communications* (Sydney, NSW, Australia) (*SA '23*). Association for Computing Machinery, New York, NY, USA, Article 13, 4 pages. <https://doi.org/10.1145/3610543.3626163>
- Shaohua Fan, Stephen Chenney, Bo Hu, Kam-Wah Tsui, and Yu-Chi Lai. 2006. Optimizing Control Variate Estimators for Rendering. *Comput. Graph. Forum* 25 (09 2006), 351–357. <https://doi.org/10.1111/j.1467-8659.2006.00954.x>
- Michael Fischer and Tobias Ritschel. 2023. Plateau-Reduced Differentiable Path Tracing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4285–4294.
- Basile Fraboni, Antoine Webanck, Nicolas Bonneel, and Jean-Claude Iehl. 2022. Volumetric Multi-view Rendering. *Computer Graphics Forum (Eurographics 2022)* 41, 2 (2022).
- Ioannis Gkioulekas, Shuang Zhao, Kavita Bala, Todd Zickler, and Anat Levin. 2013. Inverse Volume Rendering with Material Dictionaries. *ACM Trans. Graph.* 32, 6 (nov 2013), 162:1–162:13.
- Andreas Griewank and Andrea Walther. 2008. *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation* (second ed.). Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.
- H. O. Hartley and A. Ross. 1954. Unbiased Ratio Estimators. *Nature* 174, 4423 (01 Aug 1954), 270–271. <https://doi.org/10.1038/174270a0>
- Eric Heitz. 2018. Sampling the GGX Distribution of Visible Normals. *Journal of Computer Graphics Techniques (JCGT)* 7, 4 (30 November 2018), 1–13. <http://jcg. org/published/0007/04/01/>
- Eric Heitz, Stephen Hill, and Morgan McGuire. 2018. Combining analytic direct illumination and stochastic shadows. In *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games* (Montreal, Quebec, Canada) (*IGD '18*). Association for Computing Machinery, New York, NY, USA, Article 2, 11 pages. <https://doi.org/10.1145/3190834.3190852>
- Qingqin Hua, Pascal Grittmann, and Philipp Slusallek. 2023. Revisiting controlled mixture sampling for rendering applications. *ACM Trans. Graph.* 42, 4, Article 64 (July 2023), 13 pages. <https://doi.org/10.1145/3592435>
- Rie Johnson and Tong Zhang. 2013. Accelerating stochastic gradient descent using predictive variance reduction. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 1* (Lake Tahoe, Nevada) (*NIPS'13*). Curran Associates Inc., Red Hook, NY, USA, 315–323.
- Tyler B Johnson and Carlos Guestrin. 2018. Training Deep Models Faster with Robust, Approximate Importance Sampling. In *Advances in Neural Information Processing Systems*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Eds.), Vol. 31. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2018/file/967990de5b3eac7b87d49a13c6834978-Paper.pdf
- James T. Kajiya. 1986. The Rendering Equation. *Comput. Graph. (Proc. SIGGRAPH)* 20, 4 (1986), 143–150.
- Angelos Katharopoulos and François Fleuret. 2019. Not All Samples Are Created Equal: Deep Learning with Importance Sampling. arXiv:1803.00942 [cs.LG] <https://arxiv.org/abs/1803.00942>

- //arxiv.org/abs/1803.00942
- Eric P. Lafortune and Yves D. Willems. 1995. The Ambient Term as a Variance Reducing Technique for Monte Carlo Ray Tracing. In *Photorealistic Rendering Techniques*, Georgios Sakas, Stefan Müller, and Peter Shirley (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 168–176.
- Quentin Le Lidec, Ivan Laptev, Cordelia Schmid, and Justin Carpentier. 2021. Differentiable rendering with perturbed optimizers. *Advances in Neural Information Processing Systems* 34 (2021), 20398–20409.
- Tzu-Mao Li, Miika Aittala, Frédéric Durand, and Jaakko Lehtinen. 2018. Differentiable Monte Carlo Ray Tracing through Edge Sampling. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)* 37, 6 (2018), 222:1–222:11.
- Daqi Lin, Markus Kettunen, Benedikt Bitterli, Jacopo Pantaleoni, Cem Yuksel, and Chris Wyman. 2022. Generalized resampled importance sampling: foundations of ReSTIR. *ACM Trans. Graph.* 41, 4, Article 75 (July 2022), 23 pages. <https://doi.org/10.1145/3528223.3530158>
- Daqi Lin and Cem Yuksel. 2020. Real-Time Stochastic Lightcuts. *Proc. ACM Comput. Graph. Interact. Tech.* 3, 1, Article 5 (May 2020), 18 pages. <https://doi.org/10.1145/3384543>
- Wojciech Matusik, Hanspeter Pfister, Matt Brand, and Leonard McMillan. 2003. A Data-driven Reflectance Model. *ACM Trans. Graph. (Proc. SIGGRAPH)* 22, 3 (2003), 759–769.
- Bailey Miller, Iliyan Georgiev, and Wojciech Jarosz. 2019. A null-scattering path integral formulation of light transport. *ACM Transactions on Graphics (Proceedings of SIGGRAPH)* 38, 4 (July 2019). <https://doi.org/10/gfzorb>
- Zackary Misso, Benedikt Bitterli, Iliyan Georgiev, and Wojciech Jarosz. 2022. Unbiased and consistent rendering using biased estimators. *ACM Transactions on Graphics (Proceedings of SIGGRAPH)* 41, 4 (July 2022). <https://doi.org/10/gqjn66>
- Thomas Müller, Markus Gross, and Jan Novák. 2017. Practical Path Guiding for Efficient Light-Transport Simulation. *Computer Graphics Forum (Proceedings of EGSR)* 36, 4 (June 2017), 91–100. <https://doi.org/10.1111/cgf.13227>
- Thomas Müller, Brian McWilliams, Fabrice Rousselle, Markus Gross, and Jan Novák. 2019. Neural Importance Sampling. *ACM Trans. Graph.* 38, 5, Article 145 (Oct. 2019), 19 pages. <https://doi.org/10.1145/3341156>
- Thomas Müller, Fabrice Rousselle, Alexander Keller, and Jan Novák. 2020. Neural control variates. *ACM Trans. Graph.* 39, 6, Article 243 (nov 2020), 19 pages. <https://doi.org/10.1145/3414685.3417804>
- Baptiste Nicolet, Alec Jacobson, and Wenzel Jakob. 2021. Large steps in inverse rendering of geometry. *ACM Trans. Graph.* 40, 6, Article 248 (Dec. 2021), 13 pages. <https://doi.org/10.1145/3478513.3480501>
- Baptiste Nicolet, Fabrice Rousselle, Jan Novak, Alexander Keller, Wenzel Jakob, and Thomas Müller. 2023. Recursive Control Variates for Inverse Rendering. *ACM Trans. Graph.* 42, 4, Article 62 (July 2023), 13 pages. <https://doi.org/10.1145/3592139>
- Merlin Nimier-David, Zhao Dong, Wenzel Jakob, and Anton Kaplanyan. 2021. Material and Lighting Reconstruction for Complex Indoor Scenes with Texture-space Differentiable Rendering. In *Eurographics Symposium on Rendering - DL-only Track*, Adrien Bousseau and Morgan McGuire (Eds.). The Eurographics Association. <https://doi.org/10.2312/sr.20211292>
- Merlin Nimier-David, Sébastien Speierer, Benoit Ruiz, and Wenzel Jakob. 2020. Radiative Backpropagation: An Adjoint Method for Lightning-Fast Differentiable Rendering. *Transactions on Graphics (Proceedings of SIGGRAPH)* 39, 4 (July 2020). <https://doi.org/10.1145/3386569.3392406>
- Jan Novák, Iliyan Georgiev, Johannes Hanika, and Wojciech Jarosz. 2018. Monte Carlo methods for volumetric light transport simulation. *Computer Graphics Forum (Proceedings of Eurographics - State of the Art Reports)* 37, 2 (May 2018). <https://doi.org/10/gd2jqj>
- Art Owen and Yi Zhou. 2000. Safe and Effective Importance Sampling. *J. Amer. Statist. Assoc.* 95, 449 (2000), 135–143. <http://www.jstor.org/stable/2669533>
- Art B. Owen. 2013. *Monte Carlo theory, methods and examples*. <https://artowen.us.domains/mc/>.
- Matt Pharr, Wenzel Jakob, and Greg Humphreys. 2023. *Physically Based Rendering: From Theory to Implementation (4th ed.)* (4th ed.). The MIT Press, San Francisco, CA, USA. 1266 pages.
- M. J. D. Powell and J. Swann. 1966. Weighted Uniform Sampling — a Monte Carlo Technique for Reducing Variance. *IMA Journal of Applied Mathematics* 2, 3 (09 1966), 228–236. <https://doi.org/10.1093/imamat/2.3.228> arXiv:<https://academic.oup.com/imamat/article-pdf/2/3/228/2233907/2-3-228.pdf>
- M. H. Quenouille. 1956. Notes on Bias in Estimation. *Biometrika* 43, 3/4 (1956), 353–360. <http://www.jstor.org/stable/2332914>
- Ingo Rechenberg and Manfred Eigen. 1973. *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. Frommann-Holzboog Verlag.
- Corentin Salaün, Adrien Gruson, Binh-Son Hua, Toshiya Hachisuka, and Gurprit Singh. 2022. Regression-based Monte Carlo integration. *ACM Trans. Graph.* 41, 4, Article 79 (July 2022), 14 pages. <https://doi.org/10.1145/3528223.3530095>
- Corentin Salaün, Xingchang Huang, Iliyan Georgiev, Niloy J. Mitra, and Gurprit Singh. 2023. Efficient Gradient Estimation via Adaptive Sampling and Importance Sampling. arXiv:2311.14468 [cs.LG] <https://arxiv.org/abs/2311.14468>
- Corentin Salaün, Xingchang Huang, Iliyan Georgiev, Niloy J. Mitra, and Gurprit Singh. 2024. Multiple importance sampling for stochastic gradient estimation. arXiv:2407.15525 [cs.LG] <https://arxiv.org/abs/2407.15525>
- Rohan Sawhney and Keenan Crane. 2020. Monte Carlo Geometry Processing: A Grid-Free Approach to PDE-Based Methods on Volumetric Domains. *ACM Trans. Graph.* 39, 4 (2020).
- Rohan Sawhney, Dario Seyb, Wojciech Jarosz, and Keenan Crane. 2022. Grid-free Monte Carlo for PDEs with spatially varying coefficients. *ACM Trans. Graph.* 41, 4, Article 53 (July 2022), 17 pages. <https://doi.org/10.1145/3528223.3530134>
- Mark Schmidt, Nicolas Le Roux, and Francis Bach. 2017. Minimizing finite sums with the stochastic average gradient. *Math. Program.* 162, 1–2 (March 2017), 83–112. <https://doi.org/10.1007/s10107-016-1030-6>
- James C Spall. 1992. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE transactions on automatic control* 37, 3 (1992), 332–341.
- J. Spanier. 1979. A New Family of Estimators for Random Walk Problems†. *IMA Journal of Applied Mathematics* 23, 1 (01 1979), 1–31. <https://doi.org/10.1093/imamat/23.1.1> arXiv:<https://academic.oup.com/imamat/article-pdf/23/1/1/2356098/23-1-1.pdf>
- Jerome Spanier and Earl H. Maize. 1994. Quasi-Random Methods for Estimating Integrals Using Relatively Small Samples. *SIAM Rev.* 36, 1 (1994), 18–44. <http://www.jstor.org/stable/2132560>
- Tomasz Stachowiak and Yasin Uludag. 2015. Stochastic screen-space reflections. *ACM SIGGRAPH Courses: Advances in Real-time Rendering* 12 (2015).
- Joe Staines and David Barber. 2012. Variational optimization. *arXiv preprint arXiv:1212.4507* (2012).
- Tanli Su and Ioannis Gkioulekas. 2024. Path Sampling Methods for Differentiable Rendering. In *Eurographics Symposium on Rendering*, Eric Haines and Elena Garces (Eds.). The Eurographics Association. <https://doi.org/10.2312/sr.20241148>
- Laszlo Szirmay-Kalos, Ferenc Csonka, and György Antal. 2001. Global Illumination as a Combination of Continuous Random Walk and Finite-Element Based Iteration. *Computer Graphics Forum* (2001). <https://doi.org/10.1111/1467-8659.00521>
- László Szécsi, Mateu Sbert, and László Szirmay-Kalos. 2004. Combined Correlated and Importance Sampling in Direct Light Source Computation and Environment Mapping. *Computer Graphics Forum* 23, 3 (2004), 585–593. <https://doi.org/10.1111/j.1467-8659.2004.00790.x> arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8659.2004.00790.x>
- Myint Tin. 1965. Comparison of Some Ratio Estimators. *J. Amer. Statist. Assoc.* 60, 309 (1965), 294–307. <http://www.jstor.org/stable/2283154>
- Carlos Ureña, Marcos Fajardo, and Alan King. 2013. An area-preserving parametrization for spherical rectangles. In *Proceedings of the Eurographics Symposium on Rendering (Zaragoza, Spain) (EGSR '13)*. Eurographics Association, Goslar, DEU, 59–66. <https://doi.org/10.1111/cgf.12151>
- Eric Veach. 1998. *Robust Monte Carlo Methods for Light Transport Simulation*. Ph. D. Dissertation. Stanford University. Advisor(s) Guibas, Leonidas J.
- 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>
- Zican Wang, Michael Fischer, and Tobias Ritschel. 2024. Higher-order Differentiable Rendering. arXiv:2412.03489 [cs.GR] <https://arxiv.org/abs/2412.03489>
- Daan Wierstra, Tom Schaul, Tobias Glasmachers, Yi Sun, Jan Peters, and Jürgen Schmidhuber. 2014. Natural evolution strategies. *The Journal of Machine Learning Research* 15, 1 (2014), 949–980.
- A. Wilkie, S. Nawaz, M. Droske, A. Weidlich, and J. Hanika. 2014. Hero wavelength spectral sampling. In *Proceedings of the 25th Eurographics Symposium on Rendering (Lyon, France) (EGSR '14)*. Eurographics Association, Goslar, DEU, 123–131. <https://doi.org/10.1111/cgf.12419>
- E Woodcock, T Murphy, P Hemmings, and S Longworth. 1965. Techniques used in the GEM code for Monte Carlo neutronics calculations in reactors and other systems of complex geometry. In *Applications of Computing Methods to Reactor Problems*, Vol. 557.
- Kehan Xu, Sebastian Herholz, Marco Manzi, Marios Pappas, and Markus Gross. 2024. Volume Scattering Probability Guiding. *ACM Trans. Graph.* 43, 6, Article 184 (Nov. 2024), 17 pages. <https://doi.org/10.1145/3687982>
- Kai Yan, Christoph Lassner, Brian Budge, Zhao Dong, and Shuang Zhao. 2022. Efficient estimation of boundary integrals for path-space differentiable rendering. *ACM Trans. Graph.* 41, 4, Article 123 (July 2022), 13 pages. <https://doi.org/10.1145/3528223.3530080>
- Kai Yan, Vincent Pegoraro, Marc Droske, Jiří Vorba, and Shuang Zhao. 2024. Differentiating Variance for Variance-Aware Inverse Rendering. In *SIGGRAPH Asia 2024 Conference Papers (Tokyo, Japan) (SA '24)*. Association for Computing Machinery, New York, NY, USA, Article 117, 10 pages. <https://doi.org/10.1145/3680528.3687603>
- Tizian Zeltner, Sébastien Speierer, Iliyan Georgiev, and Wenzel Jakob. 2021. Monte Carlo Estimators for Differentiable Light Transport. *Transactions on Graphics (Proceedings of SIGGRAPH)* 40, 4 (Aug. 2021). <https://doi.org/10.1145/3450626.3459807>

Cheng Zhang, Zhao Dong, Michael Doggett, and Shuang Zhao. 2021a. Antithetic Sampling for Monte Carlo Differentiable Rendering. *ACM Trans. Graph.* 40, 4 (2021), 77:1–77:12.

Cheng Zhang, Bailey Miller, Kai Yan, Ioannis Gkioulekas, and Shuang Zhao. 2020. Path-Space Differentiable Rendering. *ACM Trans. Graph. (Proc. SIGGRAPH)* 39, 6 (2020), 143:1–143:19.

Cheng Zhang, Zihan Yu, and Shuang Zhao. 2021b. Path-Space Differentiable Rendering of Participating Media. *ACM Trans. Graph.* 40, 4 (2021), 76:1–76:15.

Ziyi Zhang, Nicolas Roussel, and Wenzel Jakob. 2023. Projective Sampling for Differentiable Rendering of Geometry. *Transactions on Graphics (Proceedings of SIGGRAPH Asia)* 42, 6 (Dec. 2023). <https://doi.org/10.1145/3618385>

Peilin Zhao and Tong Zhang. 2015. Stochastic Optimization with Importance Sampling. arXiv:1401.2753 [stat.ML] <https://arxiv.org/abs/1401.2753>

Shuang Zhao, Wenzel Jakob, and Tzu-Mao Li. 2020. Physics-Based Differentiable Rendering: A Comprehensive Introduction. In *SIGGRAPH Courses*. 14:1–14:30.

A DEFENSIVE AUXILIARY FUNCTIONS

In the basic ratio estimator Eq. (9), the denominator, $\sum h(X_i)/g(X_i)$, is a Monte Carlo estimator of H . At low sample counts, the estimator might become zero or very small, leading to numerical issues. Fig. 29 shows an example where this could happen.

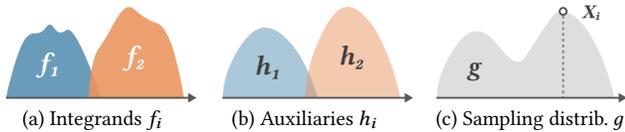


Fig. 29. (a) The overlap between the supports of integrands is sometimes small. (b) Auxiliaries typically share the same support as their corresponding integrands, if we directly use their importance sampling PDF as h . (c) To cover the union of the integrands’ supports, we often sample from mixtures of auxiliaries, resulting in samples like X_i . However, here $h_1(X_i)$ will become very small or even zero.

One practical solution is to heuristically discard the problematic zero denominators and fall back to the classical Monte Carlo estimator by considering only the numerator. While this approach can work for basic RCV, it is incompatible with the unbiased Hartley-Ross estimator. Moreover, the HR estimator involves a sum of ratios, $\sum_{i=1}^N \frac{f_j(X_i)}{h_j(X_i)}$. As a result, any single problematic sample can cause numerical issues, regardless of the sample count.



Fig. 30. (a) Original auxiliary function matches the integrands well but may have zero values over parts of the domain. (b) A uniform distribution ensures non-zero values across the domain but poorly matches f_i . (c) By linearly interpolating them, the newly composed auxiliary satisfies both requirements.

To address this issue, we must ensure that $h(X_i)$ is neither zero nor extremely close to it. Inspired by defensive sampling [Owen and Zhou 2000], we propose linearly blending h_i with a uniform distribution, as illustrated in Fig. 30. The resulting composite function $h'(x)$ is guaranteed to remain non-zero across the entire domain while preserving an analytical integral of 1. Using *defensive auxiliary functions* effectively preserves the unbiasedness of the HR estimator.