An unbiased ray-marching transmittance estimator

MARKUS KETTUNEN, NVIDIA
EUGENE D'EON, NVIDIA
JACOPO PANTALEONI, NVIDIA
JAN NOVÁK, NVIDIA

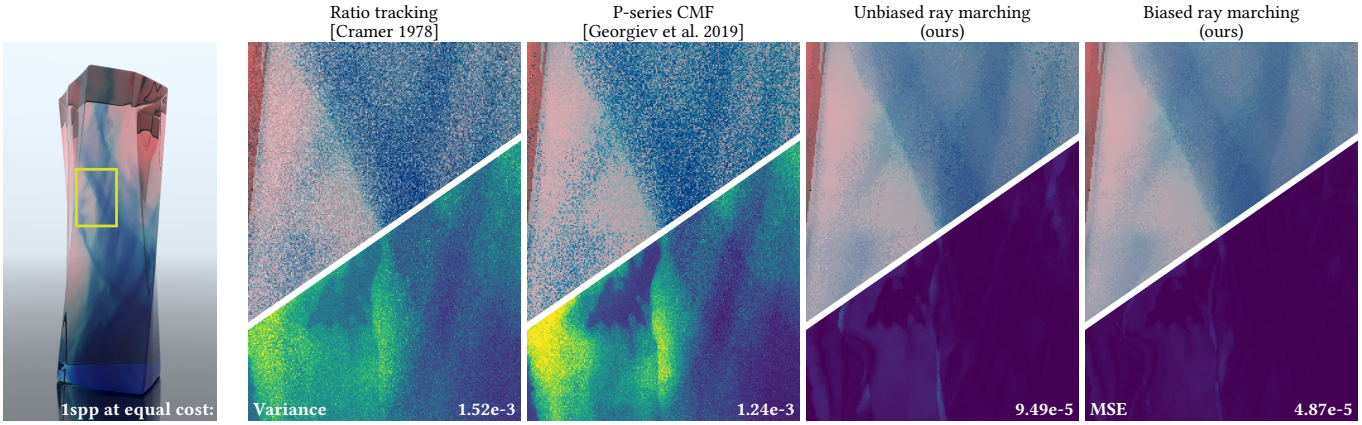

Fig. 1. We propose a new unbiased Monte Carlo estimator for volumetric transmittance based on a power series expansion. The zeroth-order term in our estimator corresponds to a variant of ray marching. The higher-order terms ensure a bias-free estimate and are evaluated infrequently. The result can have multiple orders of magnitude less variance than previous work with a similar number of density evaluations.

We present an in-depth analysis of the sources of variance in state-of-the-art unbiased volumetric transmittance estimators, and propose several new methods for improving their efficiency. These combine to produce a single estimator that is universally optimal relative to prior work, with up to several orders of magnitude lower variance at the same cost, and has zero variance for any ray with non-varying extinction. We first reduce the variance of truncated power-series estimators using a novel efficient application of U-statistics. We then greatly reduce the average expansion order of the power series and redistribute density evaluations to filter the optical depth estimates with an equidistant sampling comb. Combined with the use of an online control variate built from a sampled mean density estimate, the resulting estimator effectively performs ray marching most of the time while using rarely-sampled higher-order terms to correct the bias.

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

Additional Key Words and Phrases: transmittance, Poisson estimator, U-statistics, comb filter, power series

ACM Reference Format:
Markus Kettunen, Eugene d'Eon, Jacopo Pantaleoni, and Jan Novák. 2021. An unbiased ray-marching transmittance estimator. *ACM Trans. Graph.* 40, 4, Article 137 (August 2021), 20 pages. https://doi.org/10.1145/3450626.3459937

Authors' addresses: Markus Kettunen, NVIDIA, mkettunen@nvidia.com; Eugene d'Eon, NVIDIA, edeon@nvidia.com; Jacopo Pantaleoni, NVIDIA, jpantaleoni@nvidia.com; Jan Novák, NVIDIA, jnovak@nvidia.com.

1 INTRODUCTION

The visibility between two points in a scene is a fundamental quantity in light transport simulation. In a vacuum, it takes on a binary value. In a participating medium, however, scalar radiative transfer [Chandrasekhar 1960] is used to statistically account for the presence of scattering and absorbing particles. The number of particles intersecting a given ray is a random variable and visibility becomes a fractional quantity: the probability of traversing uncollided from a to b,

$$T(a,b) = \exp\left(-\int_a^b \mu(x)\,\mathrm{d}x\right), \tag{1}$$

where $\mu(x)$ is a known deterministic non-negative function (the *extinction coefficient* at position x). The probability $T(a,b)$ is sometimes called *transmittance*, and efficiently computing this value is essential for rendering scenes with haze, fog, and clouds.

The integral in Equation 1 is rarely known in closed form. Exceptions include piecewise-homogeneous volumes, and simple atmospheric models [Novák et al. 2018]. The general-purpose approach, therefore, is to use Monte Carlo to estimate transmittance by point-sampling $\mu(x)$ at a number of locations x along the ray. Various estimators have been proposed for this purpose, but no one estimator is optimal in all cases, and their efficiency depends on several parameters that are difficult to determine automatically.

In this paper, we present new methods for unbiased estimation of Equation 1. After reviewing previous work in Section 2, we present a new parametric variance analysis in Section 3 that reveals several

key factors that limit the performance of existing estimators. This inspires a number of novel variance reduction methods, which we detail in Section 4. Combining all of these methods together, we propose a new estimator in Section 4.7 that seems universally more efficient than prior work, and can in many cases yield transmittance estimates with orders of magnitude less variance at the same cost.

Our proposed estimator is based on a low-order Taylor series expansion of the exponential function near a relatively accurate estimate of the real optical depth obtained by multiple density lookups. The use of a low-order expansion frees up sampling budget for more accurate evaluation of both the expansion point and the Taylor series terms, which further allows lowering the evaluation order and improving the samples. This self-reinforcing loop leads to an unbiased low-variance estimator that most of the time only evaluates the already quite accurate zeroth-order term. The proposed evaluation of this term can be interpreted as the classical random-offset ray marching solution [Pauly et al. 2000], whereas the remaining terms can be interpreted as probabilistically-sampled correction terms that make it unbiased, so we refer to our technique as *unbiased ray marching*.

2 BACKGROUND AND RELATED WORK

In this section we review the main approaches for transmittance estimation in particle and light transport literature and also identify new connections to work outside of transport theory. For a review of related distance-sampling techniques, we refer the reader to the survey by Novák et al. [2018].

For brevity, we will sometimes abuse the term "density" to mean the extinction coefficient $\mu(x)$ (which is the product of the number density of particles at x with the total cross section), and the interval will sometimes be omitted (e.g., T refers to $T(a, b)$).

2.1 Ray Marching and Bias

Transmittance is the exponential of the negative optical depth,

$$T(a,b) = \exp\left(-\tau(a,b)\right) = \exp\left(-\int_a^b \mu(x)\,\mathrm{d}x\right). \tag{2}$$

The *optical depth* τ can be easily approximated using *ray marching* (uniformly-spaced samples along the interval) or by jittered and unbiased Monte Carlo approaches, but the exponential of these estimates will result in a biased estimator of $\exp(-\tau)$ [Raab et al. 2006]. The *jackknife* method and its generalizations [Miller 1974], and high-order quadrature rules [Muñoz 2014], can be used to reduce the bias in some cases, but the error may still not be acceptable for certain applications. The key challenge of transmittance estimation, then, is to form unbiased estimates of T, given only point samples of $\mu(x)$. This relates more broadly to estimating a functional $\exp(-\lambda)$ when λ is easily estimated in an unbiased fashion (see Jacob et al. [2015] for an extensive analysis of the challenges posed by general unbiased functional integration).

2.2 Poisson Point Processes

Many unbiased methods have been devised to estimate exponentiated integrals like Equation 1, and these methods are closely related to the zero-order estimation problem for point processes. A *point process* $N(\ell)$ is a stochastic counting process of the number of events (such as particles) occurring in some time (or along a ray of length) ℓ. For a *Poisson* point process (PPP), the events are independent and $N(\ell) \sim \mathrm{Po}(\lambda_\ell)$ is Poisson-distributed with *rate* λ_ℓ [Cox and Lewis 1966]. This rate is the integral of the *intensity function* $\lambda(x)$ of the process over the interval,

$$\lambda_\ell = \int_0^\ell \lambda(x)\,\mathrm{d}x, \tag{3}$$

and allows the mean density of points to vary over the domain. It is well known that a PPP is exactly the process governing the scattering and absorption events in classical radiative transfer [Cox and Lewis 1966; Mikhailov 1992], due to the assumption of independent scattering centers. The correspondence between the two is established by equating the rate of the point process $\lambda(x)$ to the extinction coefficient of the medium $\mu(x)$ as the particle moves across the interval when starting from a, $\lambda(x) = \mu(a + x)$. Transmittance is then the probability of finding no points/particles along the interval

$$T(a,b) = \Pr\left[N(\ell) = 0\right], \quad \ell = b - a. \tag{4}$$

Since the mean of a PPP is the rate $\lambda_\ell = \mathbb{E}[N(\ell)] = \tau(a,b)$, the exponential free paths of classical radiative transfer follow from the zero-order probability of the Poisson distribution (the probability mass function of a Poisson distribution with rate τ is $e^{-\tau}\tau^k/k!$, which is an exponential for $k = 0$).

2.3 Tracking Estimators

The best-known unbiased transmittance estimators are called *tracking estimators* due to the fact that they track a particle moving from a to b by sampling a PPP to determine an ordered sequence of collisions with the medium. For a constant-density medium, the exponentially-distributed free-path lengths between collisions are easily sampled [Novák et al. 2018]. For a nonhomogeneous medium, the PPP can be sampled using the method of *delta-tracking* [Bertini 1963; Butcher and Messel 1958; Coleman 1968; Galtier et al. 2013; Mikhailov 1970; Skullerud 1968; Woodcock et al. 1965; Zerby et al. 1961]. Using a *majorant* $\bar{\mu}(x) \geq \mu(x)$, a denser process is sampled whose rate / optical depth is easily computable:

$$\bar{\tau}(a,b) = \int_a^b \bar{\mu}(x)\,\mathrm{d}x. \tag{5}$$

A rejection process is then used to thin the denser process down to the desired result whereby each sampled point x_i is kept with probability $\mu(x_i)/\bar{\mu}(x_i)$. This rejection embodies the fictitious/null collision concept of the transport literature. We note that this is equivalent to a method in the point process literature known as *thinning* [Pasupathy 2011] (identification of this correspondence appears to be new). The earliest use of either method would appear to be attributed to Ulam and von Neumann shortly after the war, according to Carter et al. [1972].

While the majorant $\bar{\mu}$ is often a constant, efficiency of delta-tracking is improved with a majorant that more tightly bounds the target density. Piecewise-linear [Klein and Roberts 1984] or piecewise-polynomial [Szirmay-Kalos et al. 2011] majorants can be efficiently sampled. For a general survey of methods for sampling nonhomogeneous PPPs, see Pasupathy [2011].

Somewhat remarkably, without knowing τ, delta-tracking can be used to sample the number of real collisions in the estimation interval, i.e., the Poisson-distributed variable $N(\ell) \sim \mathrm{Po}(\tau)$, by not stopping at the first real collision. When $\bar{v}$ is the mean of n independent samples of $N(\ell)$, the minimum-variance unbiased estimator for transmittance (given only $\bar{v}$) is [Johnson 1951]

$$\widehat{T}_{\mathrm{J}} = \left(1 - \frac{1}{n}\right)^{n\bar{v}}. \tag{6}$$

The single-sample ($n = 1$) form of this estimator produces (assuming $0^0 = 1$) the delta-tracking[1] transmittance estimator [Cramer 1978; Novák et al. 2018], which returns a binary estimate depending upon whether or not $N = 0$. The case with $n > 1$ provides an interesting generalization of this estimator and, to the best of our knowledge, it has not been applied to light transport. While this estimator is optimal (given only $\bar{v}$), in practice it can be improved upon by using the sampled densities $\mu(x_i)$ directly. Another related estimator has been proposed by Raab et al. [2006], obtained by averaging together n partially stratified delta-tracking estimates.

2.3.1 Ratio tracking. Weighted tracking on a line [Cramer 1978] (also known as *ratio tracking* in graphics [Novák et al. 2014]) applies an expected-value optimization to the $n = 1$ delta-tracking estimator to form a product of ratios of densities (*null density* $\mu_{\mathrm{n}}(x) = \bar{\mu}(x) - \mu(x)$ to total density $\bar{\mu}(x)$). This is closely related to a distance-sampling scheme known as weighted delta-tracking (see e.g. Galtier et al. [2013] or Legrady et al. [2017]). Like delta-tracking, a majorant PPP samples $N \sim \mathrm{Po}(\bar{\tau})$ points x_i in the interval (a, b). Instead of returning 0 as soon as a real particle is sampled, the ratio-tracking estimator imparts a fractional opacity to each sampled particle based on its probability of being fictitious,

$$\widehat{T}_{\mathrm{rt}} = \prod_{i=1}^{N} \left(1 - \frac{\mu(x_i)}{\bar{\mu}(x_i)}\right) = \prod_{i=1}^{N} \frac{\mu_{\mathrm{n}}(x_i)}{\bar{\mu}(x_i)}. \tag{7}$$

Ratio tracking outperforms delta-tracking in most cases. However, delta-tracking can use early termination after the first real particle is sampled and avoid many unnecessary density evaluations. Therefore it can be beneficial to switch to delta-tracking after the running product in Equation 7 goes below some threshold [Novák et al. 2014].

2.4 Control Variates

A common theme in transmittance estimation is the utilization of auxiliary density functions (null-collision density, control density, etc.). While these auxiliary functions can serve different purposes, for example to facilitate sampling of collisions and/or to reduce variance, they (or their combination) can be interpreted as a control variate (CV) [Georgiev et al. 2019; Novák et al. 2014]. Given an analytically integrable control variate $\mu_{\mathrm{c}}(x)$ with $\tau_{\mathrm{c}} = \int_a^b \mu_{\mathrm{c}}(x)\,\mathrm{d}x$, the optical depth integral can be rewritten as[2]

$$\tau(a, b) = \tau_{\mathrm{c}}(a, b) + \int_a^b \mu(x) - \mu_{\mathrm{c}}(x)\,\mathrm{d}x. \tag{8}$$

We will refer to $\mu_{\mathrm{c}}(x)$ and $\mu_{\mathrm{r}}(x) = \mu(x) - \mu_{\mathrm{c}}(x)$ as the *control* and *residual* extinction coefficients, and to their respective integrals τ_{c} and τ_{r} as the control and residual optical depth. The *majorant residual* coefficient $\bar{\mu}_{\mathrm{r}}(x) = \bar{\mu}(x) - \mu_{\mathrm{c}}(x)$ and optical depth $\bar{\tau}_{\mathrm{r}} = \bar{\tau} - \tau_{\mathrm{c}}$ follow. The transmittance is then

$$T(a, b) = T_{\mathrm{c}}(a, b)\, T_{\mathrm{r}}(a, b) = \exp\left(-\tau_{\mathrm{c}}\right) \exp\left(-\int_a^b \mu_{\mathrm{r}}(x)\,\mathrm{d}x\right). \tag{9}$$

This transformation can dramatically reduce variance, particularly when the control closely matches the true density.

2.4.1 Residual Ratio Tracking / Poisson estimator. The first application of control variates in transmittance estimation was with ratio tracking to produce the *residual ratio tracking* (RRT) estimator [Novák et al. 2014]. Applying the control variate in Equation 9, the estimator reads

$$\widehat{T}_{\mathrm{rrt}} = \exp\left(-\tau_{\mathrm{c}}\right) \prod_{i=1}^{N} \left(1 - \frac{\mu_{\mathrm{r}}(x_i)}{\bar{\mu}_{\mathrm{r}}(x_i)}\right), \tag{10}$$

where $N \sim \mathrm{Po}(\bar{\tau}_{\mathrm{r}})$ points x_i are generated by sampling a PPP with intensity $\lambda(x) = \bar{\mu}(x) - \mu_{\mathrm{c}}(x) \geq 0$ in the interval (a, b). In simple words, N is the number of jumps landing inside the interval before the bound b is reached. While different articles may propose different approaches for setting the intensity $\lambda(x)$ of the PPP, this estimator is conceptually equivalent to the one known as the *Poisson estimator* [Beskos et al. 2006; Chen and Huang 2012; Fearnhead et al. 2008; Jacob et al. 2015; Papaspiliopoulos 2011], first presented by Wagner [1987].

Concurrently, Jonsson et al. [2020] have also connected the ratio tracking and Poisson estimator literature and proposed several new variants of RRT that use online estimation of a constant control. We also propose online control estimation, but include additional variance-reduction techniques such as comb-filtering. We then apply this idea to a power-series formulation, which improves performance and naturally includes ray marching as a biased member of the general formalism.

2.5 Power-Series Formulation

Another family of unbiased transmittance estimators follows from a power series (Taylor) expansion of the exponential in Equation 1. This approach has been suggested as early as [Cameron 1954] and has been used in estimation problems involving transformed observations [Neyman and Scott 1960]. One such form, which is used to estimate the exponential of the Hamiltonian in particle physics Markov-chain Monte Carlo simulations, due to Bhanot and Kennedy [Bhanot and Kennedy 1985; Lin et al. 2000; Wagner 1987, 1988], can be applied directly to transmittance estimation. Related applications of this idea to transmittance estimation were independently presented by several authors [El Hafi et al. 2021; Georgiev et al. 2019;

[1] Also known as the track-length transmittance estimator [Georgiev et al. 2019]

[2] In standard literature the CV and its integral are typically weighted by a coefficient that controls the strength of applying the CV. Since we design our CVs heuristically with the goal of maximizing positive correlations, we simply absorb the scaling factor into the CV for brevity.

Jonsson et al. 2020; Longo 2002]. Similar work has been proposed by Lyne et al. [2015] in the context of Bayesian inference.

Georgiev et al. [2019] first introduced this formulation to computer graphics, showing how it can in fact be seen as a very general framework for expressing and analyzing all transmittance estimators. Following their derivation [Georgiev et al. 2019, Equations (15) and (16)], transmittance (1) can be expressed as:

$$T(a,b) = \sum_{k=0}^{\infty} \frac{(-\tau)^k}{k!} = \sum_{k=0}^{\infty} \frac{1}{k!} \prod_{i=1}^{k} \left(-\int_a^b \mu(x)\, \mathrm{d}x \right) = 1 - \frac{\tau}{1!} + \frac{\tau^2}{2!} - \frac{\tau^3}{3!} + \cdots . \tag{11}$$

Monte Carlo estimation is then applied to each integer power of the optical depth τ^k in the expansion. This is typically[3] achieved by using k numerical estimates $\{X_1, \ldots, X_k\}$ of the negative optical depth. As long as $\{X_1, \ldots, X_k\}$ are *independent* and *unbiased*, i.e.,

$$\mathbb{E}[X_i] = -\tau(a,b) = -\int_a^b \mu(x)\, \mathrm{d}x\,, \tag{12}$$

it follows that their product provides an unbiased estimate of the k-th power of $-\tau$ (we drop the index since all X_i have the same expectation):

$$\mathbb{E}\left[\prod_{i=1}^{k} X_i \right] = \prod_{i=1}^{k} \mathbb{E}\left[X_i\right] = \mathbb{E}\left[X\right]^k = (-\tau)^k\,. \tag{13}$$

This observation allows formulating the transmittance function as the series of products of unbiased, independent estimates of (negative) optical depth:

$$T(a,b) = e^{\mathbb{E}[X]} = \sum_{k=0}^{\infty} \frac{1}{k!}\, \mathbb{E}\left[\prod_{i=1}^{k} X_i \right]. \tag{14}$$

Various estimators then follow from estimating random finite portions of this expansion with the appropriate weight corrections (explained below).

The above derivation highlights the importance of using independent and unbiased estimates of τ within a single term τ^k of the power series. Correlations across the terms of the sum, however, are perfectly acceptable. In fact, high-order terms are typically computed from the low-order ones using the *recursive formulation* [Bhanot and Kennedy 1985; Georgiev et al. 2019]

$$T(a,b) = 1 - \frac{\tau}{1}\left(1 - \frac{\tau}{2}\left(1 - \frac{\tau}{3}\left(\ldots\right.\right.\right., \tag{15}$$

which has been used in Equation 14.

A control variate $\mu_{\mathrm{c}}(x)$ is often applied to Equation 14 to yield

$$\begin{aligned} T(a,b) = T_{\mathrm{c}}(a,b)\, T_{\mathrm{r}}(a,b) &= e^{-\tau_{\mathrm{c}}} e^{-\tau_{\mathrm{r}}} \\ &= e^{-\tau_{\mathrm{c}}} \sum_{k=0}^{\infty} \frac{1}{k!} \prod_{i=1}^{k} \left(-\int_a^b \mu_{\mathrm{r}}(x)\, \mathrm{d}x \right) \\ &= e^{-\tau_{\mathrm{c}}} \sum_{k=0}^{\infty} \frac{1}{k!}\, \mathbb{E}\left[\prod_{i=1}^{k} Y_i \right], \end{aligned} \tag{16}$$

[3] See the appendix of Glasser [1962] for an interesting alternative.

where Y_i are unbiased, independent estimates of (negative) residual optical depth (we use X and Y to distinguish estimates of optical depth and residual optical depth, respectively).

Galtier et al. [2013], El Hafi et al. [2021] and Georgiev et al. [2019] proposed to set the control variate to a strict majorant, $\mu_{\mathrm{c}}(x) = \bar{\mu}(x)$, to avoid sign oscillations in power series estimates of $T_{\mathrm{r}}(a,b)$. Based on our analysis from Section 3, we will revisit this decision and propose a new way of setting the control variate in Section 4. Notice how $-\tau_{\mathrm{c}}$ effectively acts as a pivot for the Taylor series expansion. This interpretation is central to our investigations and we will refer to $-\tau_{\mathrm{c}}$ as the *pivot* in the rest of the text.

2.5.1 *Numerical evaluation.*

In practice, the evaluation of the infinite power series needs to be limited to sampling a finite number of terms; Georgiev et al. [2019] have shown that virtually all unbiased transmittance estimators can be ultimately related to sampling this power series expansion. In this respect, existing unbiased estimators can be classified into two broad categories.

Single-term estimation. Georgiev et al. [2019] showed that the delta-tracking and ratio-tracking estimators can be described in the power series formulation by noting that these estimators estimate a single term in Equation 11 at a time; when N points are sampled by the $\bar{\mu}$-driven PPP, these estimators estimate $(-\tau)^N/N!$. The general form of the single-term power series estimator is called the *generalized Poisson estimator* [Fearnhead et al. 2008],

$$\widehat{T}_{\mathrm{single}} = e^{-\tau_{\mathrm{c}}} \frac{1}{N!\, \hat{p}(N)} \prod_{i=1}^{N} Y_i\,, \tag{17}$$

where $\hat{p}(N)$ is the probability mass function of N. Using delta-tracking results in a Poisson distribution, $N \sim \mathrm{Po}(\bar{\tau}_{\mathrm{r}})$, but other distributions can be used [Fearnhead et al. 2008; Jonsson et al. 2020]. For the standard delta-tracking estimator, the random variable Y is replaced by a μ_{c}-weighted Bernoulli random variable and the power series derivation of this estimator is a special case of a more general derivation [Glasser 1962, appendix].

Truncated-series estimators. The recursive power series relation in Equation 15 directly produces a *truncated-series* estimator that estimates all terms in the Taylor expansion up to and including τ^N (see Figure 2 for an illustration). If N is a discrete random variable and $P_k = \Pr[N \geq k]$, the truncated estimator for Equation 16 is

$$\widehat{T}_{\mathrm{trunc}} = e^{-\tau_{\mathrm{c}}} \sum_{k=0}^{N} \frac{1}{k!\, P_k} \prod_{i=1}^{k} Y_i\,. \tag{18}$$

Instead of selecting N from a Poisson process, Russian roulette is commonly employed and P_k becomes the product of the continuation probabilities.

2.5.2 *Bhanot & Kennedy roulette.*

A useful scheme where the expansion is *always* evaluated up to order K and then terminated using term-wise roulette decisions follows from the equivalence (which we generalize here) [Bhanot and Kennedy 1985]

$$e^x = \sum_{k=0}^{K} \frac{x^k}{k!} + \frac{c}{K+1}\left(\frac{x^{K+1}}{cK!} + \frac{c}{K+2}\left(\frac{x^{K+2}}{c^2K!} + \ldots\right.\right. . \tag{19}$$

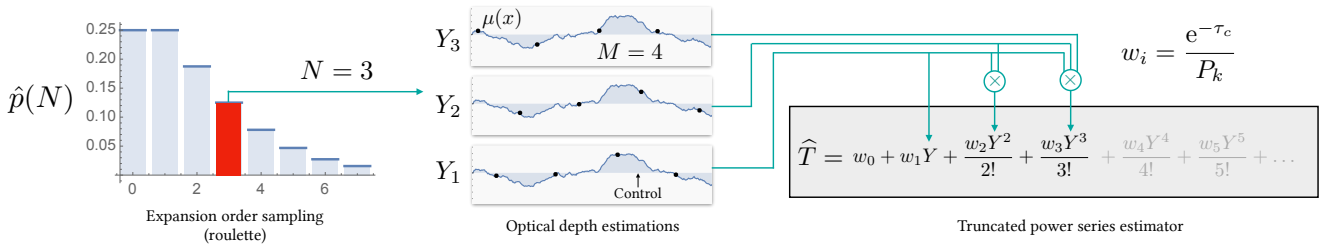

Fig. 2. Transmittance estimation via truncated power series (here illustrated for sampled order $N = 3$) using a constant control variate and multiple ($M = 4$) density evaluations (black dots) per each estimate Y_i of the negative residual optical depth.

Here c is a roulette control parameter restricted to $0 < c < K + 1$ and $c/(K + k)$ is the probability of expanding from order $K + k - 1$ to order $K + k$. Bhanot and Kennedy proposed using a continuous *expansion parameter* $c > 0$, setting $K = \lfloor c \rfloor$, which we will refer to as the *Bhanot & Kennedy* (BK) estimator. The BK roulette, specifically the parameter c, provides a very explicit control over the cost of the estimator.

Independently, Georgiev et al. [2019] proposed the *p-series CMF* estimator that sets $c = \bar{\tau}$ to the majorant and selects K such that 99% of the majorant cumulative mass function (CMF) is accumulated (assuming $\bar{\tau}$ is a safe and reasonable guess for the true optical depth when selecting K).

2.6 Additional Related Work

Delta-tracking and ratio tracking each have variations known as the *next-flight* estimators [Cramer 1978; Novák et al. 2018] that fall somewhat in between the tracking and truncated series forms. Also, Georgiev et al. [2019] introduced a number of additional estimators, including the p-series cumulative estimator that employs a different roulette strategy than described above, but concluded the p-series CMF was best overall. We refer the reader to these works for further details.

In this work we use multiple correlated density evaluations per estimate of optical depth, which was mentioned by Georgiev et al. [2019] but, to the best of our knowledge, has not been applied before.

3 EFFICIENCY ANALYSIS

In this section, we investigate the efficiency of single-term and truncated power series estimators. We measure the sensitivity of each estimator's variance to various factors, which leads to key insights that inform the design of new estimators.

3.1 Efficiency and Cost

Following prior work we define the *efficiency* of an estimator to be the reciprocal of the variance-cost product,

$$\mathrm{Eff}[\widehat{T}] = \frac{1}{\mathrm{Var}[\widehat{T}]\mathrm{Cost}[\widehat{T}]} \tag{20}$$

where $\mathrm{Cost}[\widehat{T}]$ is the mean number of density $\mu(x)$ evaluations. For both single-term and truncated power series estimators, this will depend on N: the number of unbiased estimates of negative residual optical depth (Y_i) needed to estimate a subset of the power series in Equation 16. By abandoning the physical picture of tracking estimators, the power series formulation permits a new parameter M that we call the *tuple size*, which is the number of density evaluations per estimate Y_i (see Figure 2). The estimator for Y is then

$$\widehat{Y} = -\frac{1}{M}\sum_{i=1}^{M}\frac{\mu_r(x_i)}{p(x_i)} = -\frac{1}{M}\sum_{i=1}^{M}\frac{\mu(x_i) - \mu_c(x_i)}{p(x_i)} \tag{21}$$

where $p(x)$ is the density for sampling $x \in (a, b)$ and the total cost is $\mathrm{Cost}[\widehat{T}] = \mathbb{E}[N] \cdot M$.

The efficiency of a given estimator will depend on a number of parameters that can be adjusted: the control variates, the tuple size M, and (in the case of BK roulette) the power series expansion parameter c. Ideally, an automatic procedure would optimally configure these parameters given only limited knowledge of the density statistics in the scene. To discern more about how this could be achieved, we need a detailed picture of how these parameters influence variance. While it is known that the efficiency of residual ratio tracking improves with increasing majorant [Georgiev et al. 2019] and basic heuristics for setting control variates have been discussed [Jonsson et al. 2020; Novák et al. 2014], to the best of our knowledge, little to no detailed investigation of tuple size M or expansion parameter c has been presented for any truncated estimator.

3.2 Roulette Variance in the Uniform Medium

Ultimately, variance in a transmittance estimator will arise due to two factors, which we will call *Y-variance* and *roulette variance*, and we will show that they are in fact weakly coupled. Transmittance estimators are random functions $f_{\widehat{T}}(Y_1, \dots, Y_N)$ of N random variables Y_i. By *Y-variance*, we mean the variance in the optical depth estimates Y_i themselves, which leads to variance in $\widehat{T}$ upon insertion into $f_{\widehat{T}}$. To better understand the influence of Y-variance, we can turn it off by considering a uniform medium and uniform sampling $p(x_i) = 1/\ell$. The only variance that remains is then due to N being a random variable, causing $f_{\widehat{T}}$ to evaluate different portions of the power series. This variance arises due to the roulette scheme of a truncated estimator (or PPP sampling for a single-term estimator), and we call it *roulette variance*.

In Figure 3 we compare the roulette variance of single-term and truncated power series estimators. We use residual ratio tracking, with known variance (Equation 56), as the single-term estimator and the Bhanot & Kennedy estimator to represent truncated power series; we derive the variance of the BK estimator in Appendix C.2.1.

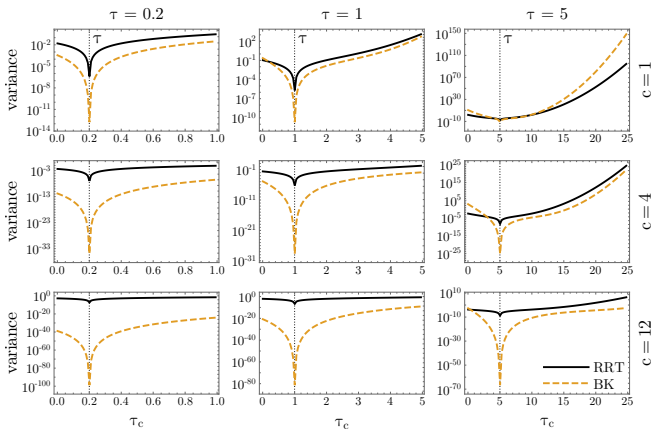

Fig. 3. Variance of cost-matched single-term and truncated estimator, represented by RRT (*solid*) and BK (*dashed*) estimators for uniform media as a function of negative pivot τ_c in three configurations of optical depth τ and expansion parameter c.

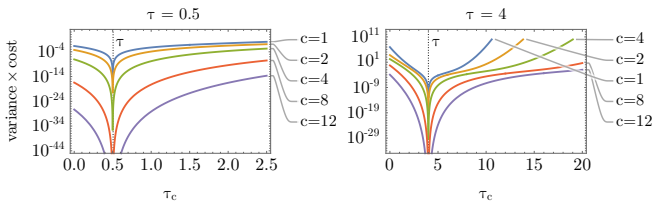

Fig. 4. Inverse efficiency of the BK estimator in a uniform medium as a function of negative pivot τ_c in two configurations of optical depth τ and five configurations of expansion parameter c.

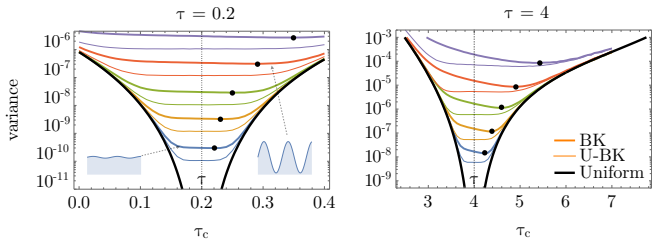

Fig. 5. Variance of the BK (thick) and U-BK (thin) estimators ($K = c = 2$) for two different optical thicknesses. The uniform medium (black) is compared to five different levels of density fluctuations (colored). The pivot with the lowest variance for the BK estimator is indicated by black dots and shifts to the right with increasing density fluctuations, making the optimal pivot difficult to predict. Application of U-statistics always reduces the variance and also widens the range of pivots where the minimum variance is achieved.

In each plot, the RRT rate is cost-matched to the BK estimator by adjusting the majorant $\bar{\mu}$ such that $\bar{\tau}_r = \bar{\tau} - \tau_c = \ell(\bar{\mu} - \mu_c) = \mathbb{E}[N_{BK}]$, where ℓ is the length of the estimation interval (see also Equation 53). The most efficient estimator is the one with the lowest variance. We observe two important trends as the pivot $-\tau_c$ is varied; first, we can achieve arbitrarily low variance by moving the negative pivot close to the true optical depth of the medium ($\tau_c = \tau$). Second, when the pivot is near this optimal value, the truncated estimator (BK) is universally better than the single-term estimator (RRT).

This analysis hints at the possibility of finding a single transmittance estimator that performs best in all cases, but also highlights the need for an accurate pivot estimator in order to achieve this. It is known that ratio tracking can outperform truncated estimators in some cases [Georgiev et al. 2019] and we see this again here for the uniform medium: when the pivot is far from its optimal value, the truncated estimator sees a significant explosion of variance, while the single-term estimator sees far less. This issue is lessened by estimating more of the series via the expansion parameter c, but at a cost of more density evaluations. Increasing c also widens the performance gap between the two estimators (Figure 3, bottom row) and extends the range of pivots where the truncated estimator is better than the single-term.

Figure 4 further demonstrates how the expansion parameter c influences the efficiency of power series estimators by plotting inverse efficiency. We see that increasing c widens the range of pivots where high efficiency can be obtained. It also shows a universal trend shared with the single-term estimator: increasing $\mathbb{E}[N]$ monotonically improves the overall efficiency, regardless of the pivot.

3.3 Varying Media

While rays passing through uniform density are common in practice (especially empty space with $T = 1$), the analysis above is not representative of the full picture. We now introduce increasing amounts of Y-variance to observe how and when the total variance changes.

In Figure 5 we compare the variance of the truncated estimator as different amounts of fluctuation in the density $\mu(x)$ are introduced while preserving the mean of $\mu(x)$. Each thick colored line corresponds to a different amount of fluctuation in $\mu(x)$ (thereby introducing Y-variance). The uniform medium (pure roulette variance) is shown in black for reference. This comparison is comprehensive in that, like with ratio tracking, it follows from the power series formulation that the variance of the BK estimator is purely a function of the mean and variance of the Y estimates (together with the pivot value)—the exact profile of the density fluctuations is irrelevant (this is because the BK roulette is independent of Y_i).

We find that the variance of the BK estimator is dominated by either Y-variance or roulette variance: *they are weakly coupled*. Far from the optimal pivot (where the black curves merge with the rest) the variance is essentially the same as that of a medium with constant density, so increasing M will have little impact. Conversely, no matter how good the pivot, variance in the samples Y_i limits the minimum-achievable variance. Further, as Y-variance decreases (by increasing M, say), the pivot needs to be closer to the optimal value to avoid roulette variance limiting the gains (to stay inside the black curves in Figure 5), suggesting that the sample budget of any online pivot estimation should be positively correlated to M.

3.4 Summary

From the analysis in this section we take away several key insights:

- Regardless of the optical depth or variation of density along a ray, the pivot is a critical parameter for achieving optimal efficiency with either single-term or truncated estimators.
- Near the optimal pivot value, truncated estimators outperform single-term when Y-variance is low.
- The lowest achievable variance of the estimator is ultimately limited by variance in the optical depth estimates.

4 HIGH-EFFICIENCY POWER SERIES ESTIMATORS

In this section, we propose new truncated power series estimators inspired by the previous analysis. These estimators build on previous work through the introduction of several novel methods, largely under the theme of Y-variance reduction and pivot estimation. We describe each of these methods separately, with both experimental and theoretical motivation for each, before detailing their combination. For notational simplicity, and unless stated otherwise, we will discuss estimation of transmittance *without* the application of the control variate. Extending the proposed improvements to residual transmittance T_r is trivial, necessitating mere substitution of the corresponding terms.

4.1 Symmetrization via U-Statistics

In order to estimate transmittance with a power series estimator that evaluates all terms up to order N (Equation 18), we need to obtain N estimates of the negative optical thickness, X_1 to X_N, and evaluate the following sum:

$$\widehat{T}_\mathrm{trunc} = 1 + \frac{X_1}{1!\ P_1} + \frac{X_1 X_2}{2!\ P_2} + \cdots + \frac{X_1 \cdots X_N}{N!\ P_N}, \tag{22}$$

where P_k is the probability of evaluating *at least* k orders. This specific estimator follows from the recursive formulation of the power series [Bhanot and Kennedy 1985; Georgiev et al. 2019] but is not the only unbiased estimator with the correct expectation. We show how to reduce the variance of this estimator with no additional density evaluations.

The key insight in reducing the variance of Equation 22 is noting that the first estimate X_1 appears in all of the terms, but the last estimate X_N is used only once, and so increasing N has little impact on the variance of the linear term, and so on. Our goal is to ensure that all estimates are in a symmetric position with respect to impacting the sum, and that we utilize the estimates maximally for each term in the estimator. We can achieve this for the first-order term in Equation 22 by replacing X_1 by the mean of all estimates:

$$m_1 \coloneqq \frac{X_1 + \cdots + X_N}{N} \tag{23}$$

Analogously, we replace the X_1X_2 product in the second-order term by the mean of all two-term products X_iX_j:

$$m_2 \coloneqq \frac{X_1X_2 + \cdots + X_1X_N + X_2X_3 + \cdots + X_{N-1}X_N}{\binom{N}{2}}. \tag{24}$$

In order to generalize this idea to the k-th order, we sum the products of all possible k-wide combinations—the k-th *elementary symmetric sum*:

$$s_k \coloneqq \sum_{1 \le i_1 < \cdots < i_k \le N} X_{i_1} \cdots X_{i_k}, \tag{25}$$

where $s_0 \coloneqq 1$, and divide it by the number of k-wide combinations; this yields a general formula for computing the k-th symmetric mean:

$$m_k \coloneqq \frac{s_k}{\binom{N}{k}}. \tag{26}$$

This variance-reduction procedure is well-known in probability theory: Equation 22 is a statistic $f(X_1, \cdots, X_N)$ of N independent and identically-distributed random variables. It is known [Halmos 1946] that the unique and minimum-variance estimator of such a statistic is the symmetric function $f^{[N]}(X_1, \cdots, X_N)$ that is invariant to the order of the X_i inputs. The generalized means m_k are known as U-statistics [Lee 1990].

Utilizing U-statistics as the numerators in Equation 22 yields a novel *U-statistics power series estimator* of transmittance:

$$\widehat{T}_\mathrm{U} = 1 + \frac{m_1}{1!\ P_1} + \frac{m_2}{2!\ P_2} + \cdots + \frac{m_N}{N!\ P_N}. \tag{27}$$

This approach can lower the variance of any estimator that estimates more than one term of the power series at the same time (tracking estimators are already fully symmetric). When combined with the generalized Bhanot & Kennedy roulette scheme (19) we refer to this estimator as the *U-BK estimator*.

In addition to reducing variance, U-statistics make it easier to find the optimal pivot. This can be seen in the variance comparisons in Figure 5 (varying Y-variance with c fixed) and Figure 6 (varying c with Y-variance fixed). For BK, the negative pivot τ_c that achieves minimum variance shifts to the right of τ as Y-variance or c increases, making this a difficult parameter to automatically determine. In addition to universally lowering the variance, we see that U-statistics flattens the variance profile in regions not dominated by roulette variance (see Figure 5). Importantly, regardless of Y-variance or c, the true optical depth τ is a (near) optimal setting for τ_c in all cases, making the goal of pivot estimation simple: to estimate the negative optical depth.

The main caveat of naively evaluating a U-statistics estimator is the exponential computational complexity: the total number of combinations required for evaluating the series up to order N is

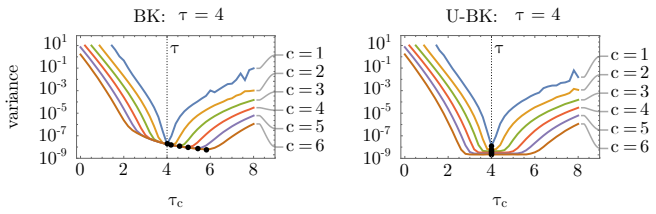

Fig. 6. Variance of BK (left) and U-BK (right) estimators for a ray with optical depth $\tau = 4$ and small Y-variance. The pivot that minimizes variance for each value of expansion parameter c is indicated (approximately) by the black dots. In addition to reducing variance, U-statistics flattens the variance profile with respect to pivot. This helps to mitigate any increase in variance due to errors in online pivot estimation and provides a simple common goal for pivot estimation—the negative optical depth of the ray.

$\binom{N}{0} + \binom{N}{1} + \cdots + \binom{N}{N} = 2^N$. A simple approximation of the optimal estimator could be built in $O(N^2)$ time by averaging N rotations of the X_i estimates (taking the average of $\{f(X_1, X_2, \cdots, X_N), f(X_2, X_3, \cdots, X_1), \ldots\}$). However, efficient full symmetrization is possible by using the Girard-Newton formulas [Mead 1992], independently found by Albert Girard and Isaac Newton in the 17th century, which relate numbers x_1 to x_N to their elementary symmetric sums s_k. By precomputing the power sums $Q_k = \sum_{i=1}^{N} x_i^k$, we have

$$s_k = \frac{1}{k} \sum_{i=1}^{k} (-1)^{i-1} s_{k-i} Q_i, \tag{28}$$

a simple and efficient recurrence relation for the elementary symmetric sums.

Although the Girard-Newton formulas provide a convenient way for directly calculating the elementary symmetric means, we found them to suffer from numerical precision problems. Algorithm 1 provides pseudo-code for a novel incremental algorithm to compute the elementary symmetric means that we designed to address these robustness issues while potentially allowing to add new samples on-the-fly. An explanation of the algorithm is provided in Appendix B. We recommend using this algorithm in practical implementations.

Our algorithm and the Girard-Newton formulas both run in time $O(NZ)$, where N is the number of samples and Z is the number of orders evaluated. Normally we evaluate all orders ($Z = N$) but if the sample count N is high enough, the highest orders might not contribute and we might want to make Z smaller than N to save time. These algorithms reduce the time of evaluating the elementary symmetric means from $O(2^N)$ to $O(N^2)$, or $O(NZ)$, and make the combination estimator practical.

ALGORITHM 1: ElementaryMeans

Input : Samples $x_1, \ldots, x_N$; Evaluation order Z
Output: Elementary symmetric means $m_0, \ldots, m_Z$
$m_0 = 1$;
$m_k = 0$ (for $k = 1$ to Z) ;
for $n = 1$ *to* N **do**
  for $k = \min(n, Z)$ *to* 1 **do**
    $m_k = m_k + \frac{k}{n} (m_{k-1} x_n - m_k)$;
  end
end

4.2 Selecting a Pivot

We presented empirical evidence in Section 3 that the pivot plays a key role in minimizing the variance of transmittance estimators and that, with U-statistics, the negative optical depth is a universally good choice. We now provide additional theoretical motivation for this observation before discussing online pivot estimation.

Interpreting the negative control thickness $-\tau_c$ as the expansion point, or pivot, of the Taylor series of the exponential provides an insightful new way to analyze power series estimation. Consider the general problem of evaluating e^x, for a given $x \in \mathbb{R}$, using its

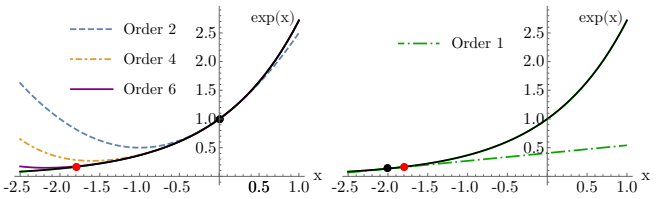

Fig. 7. Moving the pivot (black dot) of the Taylor series expansion closer to the point where we want to evaluate it (red) has a dramatic effect on its convergence. Under the power series formulation of transmittance estimation this means that a more accurate control variate permits more aggressive roulette on higher-order terms in the series, lowering the cost.

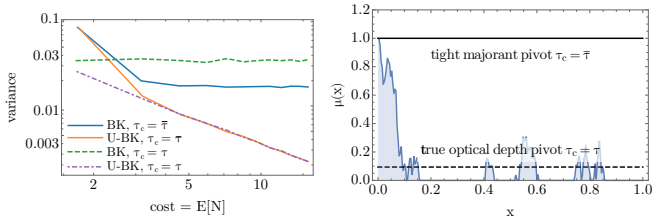

Fig. 8. The variance of a non-symmetric p-series (BK) estimator quickly reaches a plateau as the expansion order is raised (increasing c) while maintaining a fixed pivot (blue). This is because the additional samples are used only in the higher-order terms, which have an insignificant contribution in the Taylor expansion of T. Moving the pivot closer to the true optical depth (dashed green) worsens the efficiency because even fewer terms play a significant role in the expansion. Our symmetric U-BK estimators continually improve with larger c because additional samples improve all terms in the expansion.

series expansion centered at point p:

$$e^x = e^p \sum_{k=0}^{\infty} \frac{(x-p)^k}{k!} \approx e^p \sum_{k=0}^{N} \frac{(x-p)^k}{k!}. \tag{29}$$

Different values of the pivot p correspond to different polynomial fits: the closer p is to x, the faster the Taylor polynomial converges to the true value at x (see Figure 7 for demonstration). For transmittance estimation, this is another way of saying that the "mass" that each term in the series contributes to the final estimate of T shifts as the pivot changes (see [Georgiev et al. 2019, Figures 5 and 6]).

For a uniform medium, where all optical depth estimates are deterministic (zero variance), the optimal estimator is the one with pivot $p = -\tau$, where the Taylor expansion converges at the zeroth order: $e^x = e^{-\tau}(1 + \tau_r + \cdots)$ with $\tau_r = 0$. This may seem like a purely theoretical curiosity, because knowing τ immediately gives T, but it does ensure that any estimator using an estimate of $-\tau$ for p will have zero variance for rays with uniform density, such as through empty portions of the volume.

Whether or not $p = -\tau$ is a good choice in general, though, is more complicated. Without a symmetric estimator, once the optical depth estimates are random, a pivot derived from a majorant density to suppress alternating signs at consecutive orders tends to result in lower variance [El Hafi et al. 2021; Galtier et al. 2013; Georgiev et al. 2019]. This stems from a complicated interaction between the

sampling probabilities for each term in the series and their expected contributions (masses). We note that (lack of) symmetry can shed more light on why this happens.

In Figure 5 and Figure 6 we compare the variance of the truncated BK estimator as the pivot changes. As the expansion parameter c or the level of Y-variance change, the optimal pivot (black dots) moves. Note how in Figure 6 (left), at each of the optimal pivot locations, an increase in c has no effect on the variance, which plateaus despite the use of more samples. We show this also in Figure 8 with the medium and pivot held fixed as more samples are used by the estimator. The BK estimator is unable to utilize the extra samples and becomes wasteful, no matter whether it uses the tight majorant (blue line) or the true optical depth τ (orange line) as the pivot. In the latter case, the plateau is reached instantly.

The explanation for this effect is, to some extent, the lack of estimator symmetry, resulting in a bad sampling of the low order terms. Because the k-th sample Y_k impacts only the k-th and higher-order terms in the BK series, i.e., the $k-1$ preceding terms are not influenced by Y_k, the benefits of adding extra samples (and orders) diminish. Regardless of the pivot, the convergence will plateau as the expected contribution of the higher orders eventually tends to zero and additional samples stop improving the result. With a more accurate pivot, which results in even higher relative expected contribution of the low order terms, the problem is exacerbated, making the convergence plateau even sooner.

The situation changes once we utilize U-statistics because each additional sample improves *all of the terms* in the Taylor expansion. Computing the pivot using the majorant density no longer improves performance, while the approximate density mean always yields better variance, especially in the low sample-count setting (Figure 8). As we noted above, in addition to lowering the variance relative to the non-symmetric estimator, U-statistics creates a range of pivots containing $p = -\tau$ where the variance is near-optimal, and so we will refer to $p = -\tau$ as the *optimal pivot*. Appendix A contains a more thorough analysis of the effects of using the approximate mean pivot.

4.2.1 Variance of higher-order terms. Assuming we have obtained a good pivot, a natural next question is to understand how the Y-variance relates to the required order of the Taylor expansion. With a good pivot, the variance of the higher-order terms decreases exponentially with the order. Let us assume a relatively accurate pivot, $p \approx \mathbb{E}[X]$, such that our shifted samples $Y_i := X_i - p$ have approximately zero expectation. For simplicity, let us also assume that our estimates for the terms $\mathbb{E}[Y]^k$ are given by simple products $Y_1 \cdots Y_k$. The variance of the product is

$$\mathrm{Var}[Y_1 \cdots Y_k] \approx \mathbb{E}[Y^2]^k \approx \mathrm{Var}[Y]^k, \tag{30}$$

which relies on the assumption that $\mathbb{E}[Y]^2 \approx 0$. Since shifting a random variable does not change its variance, if we decrease the variance of our samples X_i to a factor s, the variance of our estimate for $\mathbb{E}[X-p]^k$ will fall geometrically to the factor of s^k, quickly making the higher-order terms insignificant. For instance, decreasing the variance of our samples by 50% would result in a decrease of the variance of the 10th order term to around 1/1000 of its original value.

This means that even a small reduction in Y-variance makes the Taylor series converge with fewer terms. This, in turn, allows us to save computation by more aggressive Russian roulette. We can use this freed sampling budget for bringing the variance of the samples down even more – a potential self-amplifying feedback loop. However, all of this needs a good pivot that might not always be available.

4.2.2 Sampled pivots and additional symmetry. One way of obtaining accurate approximations of the optimal pivot is to subdivide the volume, precompute localized statistics, and query them along each ray. We take a lighter approach and propose to estimate the optimal pivot on-the-fly by taking an additional independent sample X_{N+1} of the integral $-\tau$.

A single sample might not seem enough to estimate $-\tau$. However, we observe that we can apply a procedure analogous to the rotations briefly mentioned in Section 4.1 to effectively increase the total number of samples to $N+1$: If we denote the entire set of samples $\{X_1, \cdots, X_{N+1}\}$ by X, we may consider all $N+1$ estimators resulting from taking each unbiased sample X_i as the pivot in turn, using the other samples $X \setminus \{X_i\}$ to build our symmetrized estimator from Section 4.1, and averaging the result. Formally, this unbiased estimator of order N reads

$$\widehat{T} = \frac{1}{N+1} \sum_{i=1}^{N+1} f_N\left(X_i, X \setminus \{X_i\}\right), \tag{31}$$

where the estimator f_N is given by

$$f_N(p, Y) = e^p \sum_{k=0}^{N} \frac{m_k(Y-p)}{k!\, P_k}. \tag{32}$$

Here, m_k is the k-th symmetric mean of the samples (excluding the pivot), with the samples X_i shifted to be relative to the pivot. The samples X_i are independent unbiased estimators of $-\tau$ and each sample acts as the pivot in turn within the average. This ensures that all samples in the expanded set have a symmetric contribution in the new estimator.

4.3 Combed Estimators

In this section, we focus on reducing *Y-variance*. Equation 30 suggests that the variance of higher-order terms k is proportional to the k-th power of the variance of the estimators X_i: hence, even a small reduction in variance of each individual X_i will transform into much larger reductions for the higher-order terms of the series expansion. We propose to use an unbiased, multi-sample estimator that strikes a better quality-cost trade-off than single-sample estimators $X_i = -\frac{\mu(x_i)}{p(x_i)}$, where $x_i \sim p$. The estimator is based on a randomized Cranley-Patterson (CP) rotation of equidistant points; other sampling patterns are briefly discussed in Section 6.

Without loss of generality, we assume the integration interval to be $[0, \ell)$. We use an M-tuple of equidistant points $\{u_j : \ell j M^{-1}\}_{j=1..M}$ that we randomly offset and wrap around the $[0, \ell)$ interval using the CP rotation. For each order i, we use a single random number $x_i \in [0, \ell)$ to obtain the rotated set $\{x_{ij} : (x_i + u_j) \bmod \ell\}_{j=1..M}$ and

estimate the optical depth as:

$$X_i = -\frac{1}{M}\sum_{j=1}^{M}\frac{\mu(x_{ij})}{p(x_{ij})}. \tag{33}$$

Notice that this estimator is equivalent to convolving the integrand with an M-point Dirac comb:

$$\mu^{\otimes}(s) = \frac{1}{M}\sum_{j=1}^{M}\mu\left((s+u_j) \bmod \ell\right). \tag{34}$$

Henceforth, we will refer to the resulting estimators as *combed estimators*.

Fast convergence rate of equidistant sampling. Using multiple density evaluations induces higher evaluation cost than single-sample optical depth estimators. It is thus important to consider whether the U-statistics estimator, which utilizes all estimates maximally, yields lower variance with few high-quality estimates or with many low-quality ones.

The reason why using combs dedicating M evaluations to each X estimate is advantageous stems from the very fast convergence rate of equidistant sampling: while random sampling with M evaluations yields error reduction of $O(1/\sqrt{M})$, integrating with equidistant combs features convergence rate of $O(1/M)$ (if the integrand has bounded slope). We also observed that with an equal density query budget, denser combs (i.e., more accurate optical depth estimates), together with fewer evaluated orders (to stay in the budget), is more efficient than trying to leverage the U-statistics estimator with more combinations of less accurate estimates.

Our proposed algorithms will take this idea to a logical maximum: we try to maximally benefit from the improved convergence rate by utilizing as dense sampling combs as possible (i.e., a large M), and compensate for the larger M by a very aggressive Russian roulette to keep the truncation order N low.

4.3.1 Combing as density reshuffling. As shown in Equation 34, an M-point equidistant sample of the density function corresponds to a single evaluation of the convolution of the density with an M-point Dirac comb. This convolution does not change the value of the density integral: it merely reshuffles its density into a form that is more suitable for Monte Carlo estimation (see Figure 9). This inspires a new general *invariance principle for transmittance estimation*: we can alter the density along the ray in any integral-preserving way and not change the result. With this principle, we can maintain the physical picture of a particle traversing the interval or use the Volterra integral formulation of transmittance [Georgiev et al. 2019] and still benefit from Y-variance reduction using a tuple size M. We can also design additional density-reshuffling transformations that further reduce Y-variance.

4.4 Endpoint Matching

The CP rotation utilized in the combed estimator may introduce an artificial discontinuity. This is easy to realize when noticing that rotating the set of samples around the integration interval is equivalent to rotating the integrand (while keeping the set of samples fixed). The original interval endpoints a and b coincide at a new location x_i where the rotated integrand μ^{cp} features a discontinuity–see Figure 10.

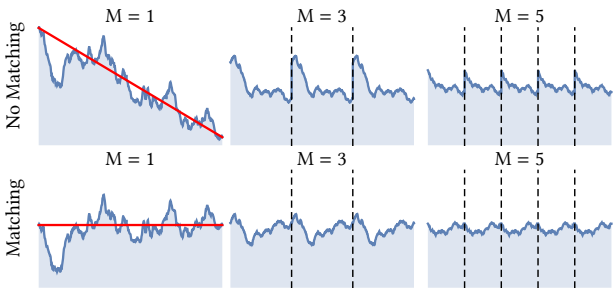

Fig. 9. Top row: M-tap Dirac comb filtering is used to reshuffle the original density $\mu(x)$ (left) to reduce Y variance. Bottom row: an affine transformation (red) that preserves optical depth is applied to the density to match the endpoints and remove the discontinuities in the combed densities with only two extra density evaluations.

In practice, if the original μ had a bounded maximum slope, which is often the case in practice, μ^{cp} no longer does. This breaks the assumption that guarantees the improved convergence rate of equidistant sampling.

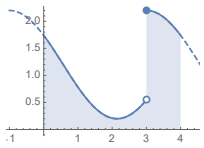

Fig. 10. Discontinuity.

We can remedy the discontinuity by another *density reshuffling* operation, namely subtracting a zero-mean affine control variate that interpolates the endpoints:

$$\mu^{\star}(s) := \mu(s) + \left(\frac{1}{2} - \frac{s}{\ell}\right)(\mu(\ell) - \mu(0)). \tag{35}$$

This modification eliminates the discontinuity caused by the random offset ($\mu^{\star}(0) = \mu^{\star}(\ell)$) while the integral remains unchanged, re-enabling the improved convergence rate from equidistant sampling. See the bottom row of Figure 9 for an illustration. Appendix D provides further formulas and simplifications.

4.5 Russian Roulette

Sampling the order of the series expansion N is typically performed incrementally by Russian roulette, and several methods have been discussed [Bhanot and Kennedy 1985; Booth 2007; Georgiev et al. 2019; Girolami et al. 2013; Moka et al. 2019; Papaspiliopoulos 2011]. Our work builds on the *Bhanot & Kennedy* roulette described in Section 2.5.2.

Using a sampled pivot that is close to the true optical thickness reduces the expected contributions of the first and higher orders of the Taylor expansion. Moreover, with the combed and endpoint-matched M-sample U-BK estimator, both the bias and variance of the zeroth order term are often very small, while most of the variance comes form the higher-order (correction) terms.

Using all the samples for the pivot would improve the estimator (which has superlinear convergence in M), but it would also make the method biased; however, we can still take advantage of this observation by allocating a larger portion of the samples to the zeroth order, and sampling the higher-order terms more infrequently.

In order to do that, we terminate the series at the zeroth order with probability p_Z, and only sample the first and higher order terms with probability $1 - p_Z$ times the original BK roulette probabilities. This is equivalent to using the original BK roulette with the acceptance probability of the first-order term multiplied by $1 - p_Z$, leading to the following probabilities of evaluating at least the first K terms:

$$P_0 = 1 \tag{36}$$

$$P_1 = \cdots = P_K = 1 - p_Z\,, \tag{37}$$

and the following *conditional* probability for adding the subsequent terms:

$$P_{k|k-1} = \frac{c}{k}, \quad \text{where } k > K\,. \tag{38}$$

This amortizes the cost of the correction terms, allowing us to use larger tuple sizes M, which in turn improves the pivot and exponentially reduces the expected contribution and variance of the higher-order terms. Algorithm 2 lists the pseudo-code.

In practice, we found that truncating the series at the zeroth term in 90% of cases, i.e., $p_Z = 0.9$, provides a large increase in efficiency across all of our tests. Using the BK scheme with parameters $K = c = 2$ lowers the expected evaluation order from $e-1 \approx 1.71828$ to about 0.31945, decreasing our expected sample count from $e \approx 2.71828$ to 1.31945, and allowing us approximately twice larger tuple sizes M. The superlinear convergence obtained by increased tuple sizes more than offsets the variance increase caused by the higher weights of the correction terms.

Raising this probability to 99% yields even lower variance, but at the cost of occasional outliers (manifesting as "fireflies"). Moreover, evaluating only the zeroth-order term 90% of the time already provides 90% of the possible cost savings, effectively amortizing the cost of the higher orders, so higher values are unlikely to strike much better efficiency.

4.6 Tuple Size Deduction

With all the above improvements, we obtain an estimator that has superlinear convergence properties in the tuple size. While this allows very efficient reduction of transmittance noise by adding more samples, we need to strike a practical balance. We found that a sample count related to the one used in the p-series CMF estimator [Georgiev et al. 2019] works well in practice, and we describe its evaluation and use here.

As a first step, we employed a simple grid search to obtain a fit for the expected sample count used by the p-series CMF with 99% mass (given the control optical thickness $\bar{\tau}$):

$$\mathbb{E}[N_{\text{CMF}}] \approx \left\lceil \sqrt[3]{(0.015 + \bar{\tau})(0.65 + \bar{\tau})(60.3 + \bar{\tau})} \right\rceil. \tag{39}$$

This approximation is (by empirical analysis) asymptotically correct, has mean absolute error of 0.34 samples for $\bar{\tau} < 10$ and a maximum relative error of 9% for $\bar{\tau} \geq 10$. We found this more efficient than the approach used by Georgiev et al. [2019].

We then use Algorithm 3 to solve for M such that a generalized BK roulette with given K produces the same mean number of density evaluations as p-series CMF. This uses an exact formula, Equation 53, for the expected evaluation order $\mathbb{E}[N_{BK}]$. We need one sample per order plus one for the pivot, and hence, to achieve the same cost,

ALGORITHM 2: AggressiveBKRoulette

Input : $p_Z = 0.9$
Output: maximum order k and probabilities $P_0, \ldots, P_k$

$w_0 = 1$;
$P = 1 - p_Z$;
u = rand();
// Stop at the zeroth-order term with probability p_Z
if $P \leq u$ **then**
  return 0;
end
// BK with $K = c = 2$;
$K = c = 2$;
for $k = 1$ *to* K **do**
  $P_k = P$;
end
for $k = K + 1$ *to* ∞ **do**
  // Compute the continuation probabilities
  $c_k = \min(c/k, 1)$;
  // Update the probability of sampling at least order k
  $P = P \cdot c_k$;
  // Russian roulette termination
  if $P \leq u$ **then**
    return $k - 1$;
  end
  // Final probability for order k
  $P_k = P$;
end

ALGORITHM 3: BKExpectedEvalOrder

Input : p_Z, c, $K = \lfloor c \rfloor$ assumed
Output: expected evaluation order of our roulette
// Evaluate $\mathbb{E}[N_{BK}] = K + \left(K!\,/c^K\right)\left(e^c - \sum_{k=0}^{K} c^k/k!\right)$
$K = \lfloor c \rfloor$;
$t = 1$;
$sum = 1$;
for $k = 1$ *to* K **do**
  $t = t * c/k$;
  $sum = sum + t$;
end
$E_N = K + (\exp(c) - sum)/t$;
// Non-zero orders are evaluated with probability $1 - p_Z$
return $(1 - p_Z) \cdot E_N$;

ALGORITHM 4: DetermineTupleSize (for unbiased ray marching)

Input : control optical thickness $\bar{\tau}$
Output: tuple size M matching the p-series CMF cost
$N_{CMF} = \left\lceil \sqrt[3]{(0.015 + \bar{\tau})(0.65 + \bar{\tau})(60.3 + \bar{\tau})} \right\rceil$;
N_{BK} = BKExpectedEvalOrder(2) ; // ≈ 0.31945.
return $\max(1, \lfloor N_{CMF}/(N_{BK} + 1) + 0.5 \rfloor)$;

the expected sample count $\mathbb{E}[N_{BK}] + 1$ times the tuple size M must match $\mathbb{E}[N_{CMF}]$. In other words, our desired tuple size is given by

$$M = \frac{\mathbb{E}[N_{CMF}]}{\mathbb{E}[N_{BK}] + 1}. \tag{40}$$

Algorithm 4 provides pseudo-code for the final algorithm.

In practice, in order to not oversample high-density but low-variance volumes, we recommend using the difference between the majorant and minorant optical thicknesses as the control parameter $\bar{\tau}$, when a minorant is available.

4.7 Assembling the Estimators

The following paragraphs summarize the construction of our final estimators.

4.7.1 The unbiased ray-marching estimator. Our final unbiased estimator is summarized in Algorithm 5 and works as follows: We first determine the number of density evaluations, M, for estimating each sample of negative optical thickness X_i (as described in Section 4.6), and we determine the highest order of the power series, N, using our modified BK roulette (Section 4.5); these first two steps do not impact each other.

Then we compute $N + 1$ combed estimates $(X_1, \ldots, X_{N+1})$ using equidistant, CP-rotated evaluations (Section 4.3) and apply endpoint matching (Section 4.4).

Finally, we use each X_i as the pivot (Equation 31) and evaluate the Taylor series using the symmetrized estimator (Section 4.1). Specifically, we use our new elementary-means algorithm (Algorithm 1) with the remaining N estimates $X \setminus \{X_i\}$.

Notice that endpoint matching is beneficial only if the ends of the interval have different density values, and we hence treat it as an optional robustness feature. Moreover, at very low sample counts, e.g. $M < 8$, we disable endpoint matching, since we found that the additional overhead of its two additional evaluations $\mu(0)$ and $\mu(\ell)$ was not worth the resulting variance reduction.

Most of the time our estimator evaluates only the zeroth-order term, when the roulette samples evaluation order $N = 0$. In this case, the estimate is simply e^{X_1}, where X_1 is the single estimation of negative optical depth using equidistant evaluations of the density from a to b. This zeroth-order term corresponds to the biased *random-offset ray marching* method [Pauly et al. 2000], optionally coupled with our endpoint-matching control variate, and the step size user parameter set by our method. The higher-order terms correct the bias, so we call our estimator *unbiased ray marching*.

4.7.2 The biased ray-marching estimator. One of the interesting conclusions from Section 4.5 is that with all our optimizations in place and enough equidistant samples and a sampled pivot, we can make the Russian roulette most often truncate at the constant term and still obtain very little variance. This is possible because our sampled pivots become increasingly good estimates for the real integral with the addition of more equidistant samples, and hence even a zeroth-order Taylor polynomial often results in a very good and cost-effective estimate for the real integral.

This behavior is partially explained by the following observation: When the pivot X_1 is an unbiased estimate for the integral, the zeroth-order approximation e^{X_1} is actually in a sense accurate to the first order, essentially gaining an order of accuracy for free:

$$\begin{aligned}\mathbb{E}[e^{X_1} - e^{\mathbb{E}[X]}] &= e^{\mathbb{E}[X]}\,\mathbb{E}\left[e^{X_1 - \mathbb{E}[X]} - 1\right] \\ &= e^{\mathbb{E}[X]}\,\mathbb{E}\left[(X_1 - \mathbb{E}[X]) + \frac{(X_1 - \mathbb{E}[X])^2}{2!} + \cdots\right] \\ &= e^{\mathbb{E}[X]}\,\mathbb{E}\left[0 + \frac{(X_1 - \mathbb{E}[X])^2}{2!} + \cdots\right].\end{aligned} \tag{41}$$

This suggests that we can build an effective low-bias estimator by *always* truncating the series at the zeroth order, that is to say, evaluating only

$$e^{-\tau} \approx e^{X_1} \cdot 1 = e^{X_1}, \tag{42}$$

where X_1 is an unbiased estimator for the optical depth, which we obtain with combing and by using all of the transmittance budget to increase the tuple size M. We optionally couple this technique with our endpoint matching control variate (see Algorithm 6).

This again reduces to *random-offset ray marching* [Pauly et al. 2000] with endpoint matching and the following very important difference: our step size is calculated automatically and it adapts to the medium as required, rather than being left as an arbitrary user parameter. This yields an efficient and simple algorithm with minimum observable bias.

ALGORITHM 5: Unbiased ray marching

Input : Interval length ℓ; control optical thickness $\bar{\tau}$
Output: Transmittance T
M = DetermineTupleSize($\bar{\tau}$);
$N, (P_0, \ldots, P_N)$ = AggressiveBKRoulette($K = c = 2, p_Z = 90\%$);
μ_ℓ, μ_0 = EvalDensity(ℓ), EvalDensity(0); // optional
for $i = 1$ *to* $N + 1$ **do**
  u_i = rand();
  $X_i = -\frac{\ell}{M} \cdot \sum_{j=0}^{M-1} \text{EvalDensity}(\frac{\ell}{M}(u + j))$;
  $X_i = X_i - \frac{\ell}{M}\left(\frac{1}{2} - u_i\right)(\mu_\ell - \mu_0)$; // optional
end
T = 0;
for $i = 1$ *to* $N + 1$ **do**
  $m_0, \ldots, m_N$ = ElementaryMeans($X \setminus \{X_i\} - X_i$);
  $T = T + \frac{1}{N+1} e^{X_i} \sum_{k=0}^{N} \frac{m_k}{k! P_k}$;
end

ALGORITHM 6: Biased ray marching

Input : Interval length ℓ; control optical thickness $\bar{\tau}$
Output: Transmittance T
$M = \left\lceil \sqrt[3]{(0.015 + \bar{\tau})(0.65 + \bar{\tau})(60.3 + \bar{\tau})} \right\rceil$;
u = rand();
$X = -\frac{\ell}{M} \cdot \sum_{j=0}^{M-1} \text{EvalDensity}(\frac{\ell}{M}(u + i))$;
$X = X - \frac{\ell}{M}\left(\frac{1}{2} - u\right)\left(\text{EvalDensity}(\ell) - \text{EvalDensity}(0)\right)$; // optional
$T = e^X$;

5 RESULTS

In this section we compare our proposed unbiased and biased transmittance estimators to ratio tracking (RT) [Cramer 1978], residual ratio tracking (RRT) [Novák et al. 2014] and the p-series CMF [Georgiev et al. 2019] estimators in a variety of scenes featuring participating media. For the unbiased methods we report variance, and for our biased ray marching we measure mean squared error of one sample.

In Figure 1 and Figure 11 we study the performance of the individual estimators in a path tracer. All volumes in the figure are stored using the VDB data structure [Museth 2013] that additionally provides aggregate volumetric statistics (minimum, maximum, and mean density) over $8 \times 8 \times 8$ voxel regions—super voxels. We utilize the statistics for computing tighter (residual) majorants for tracking estimators. For p-series CMF and our estimators, we use the mean densities in super voxels to "warp" the lookups. Along each ray, we perform regular tracking [Amanatides and Woo 1987] through the super grid and build a piecewise-constant probability density function (PDF) from the super-voxel means. We then distribute the lookup points proportional to this PDF using the inversion method. For our estimators, specifically, we generate CP-rotated equidistant samples in the $[0, 1]$ primary interval and then transform them into a warped comb along the ray.

The insets in the figure show results for different estimators at one path sample per pixel. Since the efficiency of certain estimators improves with higher lookup counts, we normalize the comparison by adjusting them to yield an approximately equal number of density lookups *per transmittance estimate*. We use the p-series CMF estimator as the baseline and uniformly increase local (residual) majorants such that the tracking estimators utilize approximately the same number of lookups (predicted by Equation 39). For our methods we employ the automatic tuple size mechanism discussed in Section 4.6. This results in a relatively fair, architecture and implementation agnostic comparison.

In the following we list the specifics of individual scenes:

- Plume features absorptive smoke. Transmittance estimation is the only source of noise in this scene.
- Box features indirect illumination from an area light surrounded by an absorptive medium. Despite the extra noise from simulating up to four light bounces, the impact of the different transmittance estimators is still clearly visible.
- Cloud features single-scattering illumination due to two point lights. We use equiangular sampling [Kulla and Fajardo 2011] to sample collisions along primary rays. Transmittance along the primary ray and the shadow rays is estimated with the studied estimator. The improvement from the transmittance estimation is partially masked by the noise from distance sampling: only a certain amount of noise would be removed even with exact transmittance values.
- Glass with Smoke features frequency-dependent absorptive smoke in a reflective glass box and shows how improved transmittance estimation can affect the quality of volume rendering either directly or through reflections. To remove noise due to randomly selecting reflection/transmission interactions with the glass, we visualize and measure variance only on paths that undergo two consecutive transmission interactions; these contribute most of the color.
- Figure 1 features another frequency-dependent absorptive medium in a glass embedding, and the same selective variance measurement as the Glass with Smoke scene.

Our unbiased estimator obtains an MSE reduction between 1.5× and 13× across all scenes compared to the previous state-of-the-art method for each scene. Our biased estimator provides an *additional* improvement of 1.1× to 2× on top of that. Note that the MSE values include also other sources of noise (such as from global illumination in the Box scene, or single-scattering in the Cloud scene), which partly masks the improvements in transmittance estimation.

Scaling to higher quality. Figure 12 shows a simple test comparing the variance of our estimators to Georgiev et al.'s [2019] p-series CMF as a function of the number of density evaluations on a single example density that exhibits high frequencies and fractal behavior. For p-series CMF, we examined two methods of increasing the expected number of samples: by averaging multiple evaluations, and by multiplying the optical thickness (in this case the majorant) by a constant greater than 1. For our methods we increase the control optical thickness similarly, but it is only used for calculating the tuple size.

Increasing the majorant helps the p-series CMF estimator in this particular instance, but this is not always the case, as we will see later on. Our method always benefits from increasing the tuple size due to the improved pivot and lower-variance correction samples, but we see bumps in the convergence curves due to the non-uniform frequency response of our equidistant sampling combs. Despite the bumps, we always found equidistant sampling to perform better than breaking the frequency response with e.g. stratified (jittered) sampling or a low-discrepancy pattern.

Our methods clearly show a higher rate of convergence which continually increases their lead by orders of magnitude when targeting noise-free transmittance estimates, with the biased variant featuring slightly lower MSE at the cost of a small amount of bias.

Quantifying the bias. Figure 16 studies the bias of our biased method in the Plume scene, which features a wide range of optical thicknesses. The figure also includes a variant of our method that is otherwise the same, but uses stratified (jittered) sampling instead of equidistant sampling combs. As predicted, our equidistant combs compute the optical thicknesses with less variance, which leads to less bias and variance in the transmittance estimates. The bias from only evaluating the zeroth-order term is often relatively invisible to the eye.

Endpoint matching. Figure 13 shows another test where we analyze the behavior of our estimators with and without endpoint matching on two different densities. The top of the figure shows a case where the density is very different at the two endpoints (e.g., when a ray starts outside and terminates inside a medium). This difference creates a strong discontinuity in the periodic extension of the function. Our endpoint-matching control variate removes this discontinuity, greatly reducing the variance and improving the convergence rate. The effect is particularly big when the discontinuity is large compared to the other variation in the density function (as

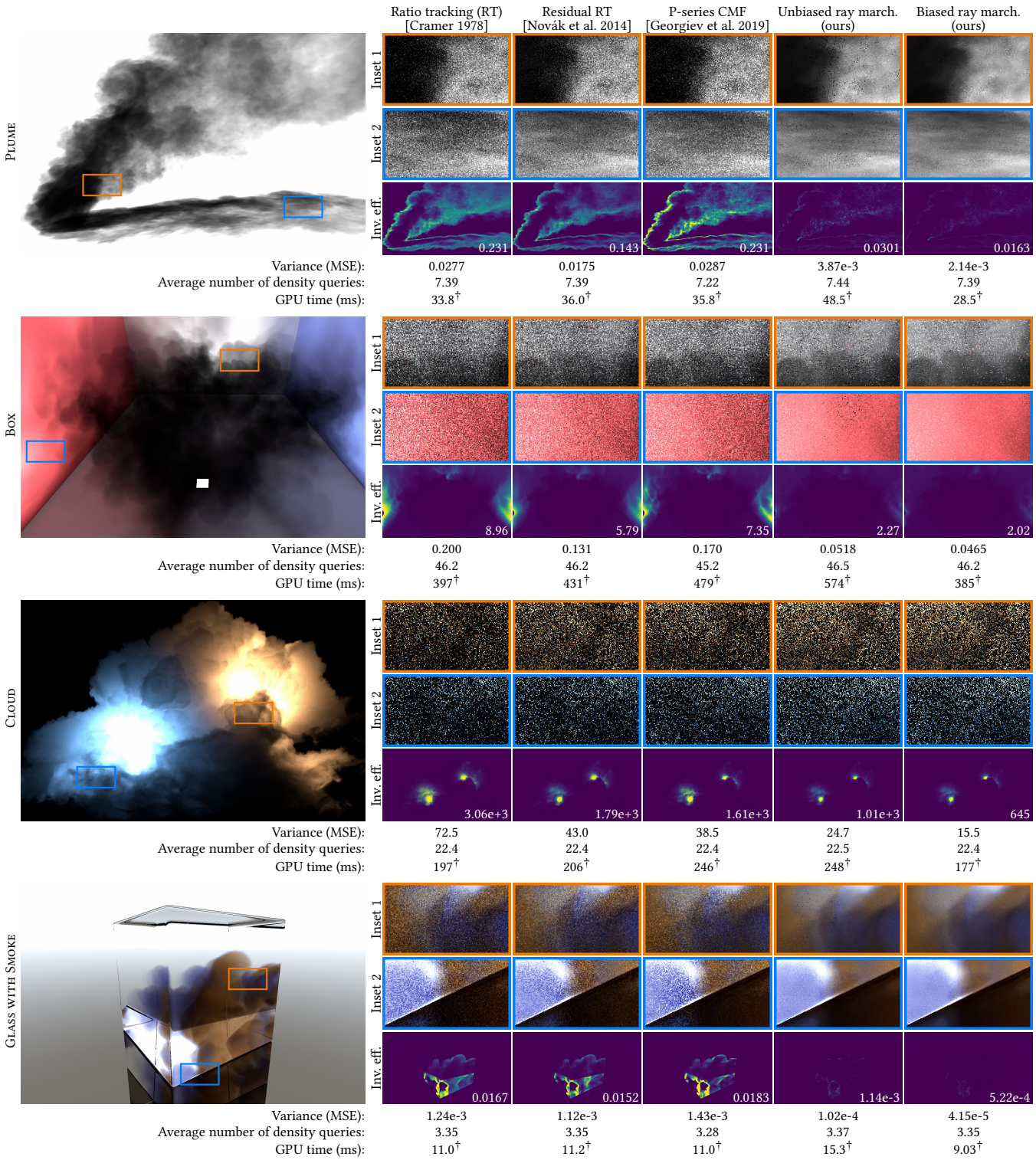

Fig. 11. A comparison of our unbiased and biased estimators (two rightmost columns) to ratio tracking, residual ratio tracking and the p-series CMF estimators on a variety of rendered content featuring participating media. The Plume, Box and Glass with Smoke scenes contain purely absorptive media, while the Cloud scene shows single-scattering illumination by point lights, rendered using equiangular sampling to sample collisions along primary rays. †: The GPU times are heavily influenced by SIMT and roulette-based branching in our non-wavefront GPU renderer. See the end of Section 5 for discussion.

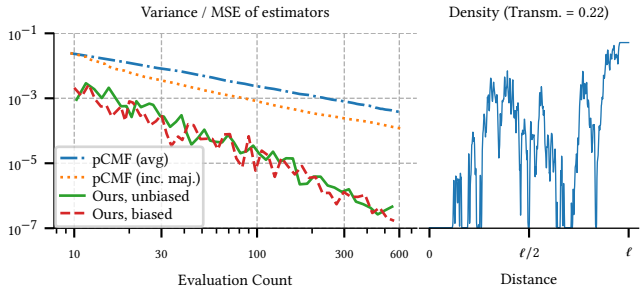

Fig. 12. A graph of variance (respectively MSE) of our unbiased and biased estimators as well as that of p-series CMF [Georgiev et al. 2019] as a function of sample count. For p-series CMF, we display two methods of increasing the expected sample count: the first (dashed blue line) is by averaging multiple evaluations, the second (dashed yellow line) is increasing the control optical thickness (in this case the majorant). Both our estimators display a faster convergence rate. The right-hand-side shows the density profile of the integrated medium.

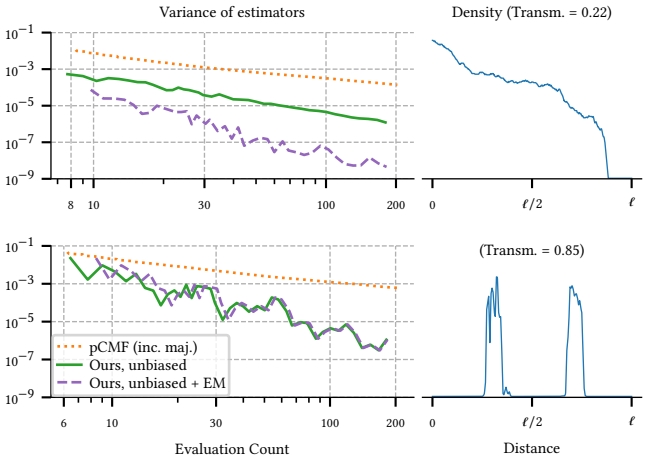

Fig. 13. Endpoint matching (top, dashed purple) may improve the convergence rate over the base method (green) when the interval ends are at very different densities compared to the general density variation (top right). Endpoint matching is not beneficial when the ends are at similar densities (bottom).

in this example). The bottom plot shows a counter example where the control variate does not yield any improvement. The plot shows that the overhead of performing the extra lookups at the endpoints is relatively low; we see only mild reduction in efficiency, especially when targeting high-quality transmittance estimates.

Pure transmittance estimation. Figure 14 analyzes the impact of gradually enabling some of our proposed techniques.

We make the following observations:

- Using tight per-pixel majorants causes the p-series CMF estimator to take discrete jumps in the base number of terms evaluated, due to it activating RR after reaching 99% mass only–this appears as blocky variations in variance/efficiency.
- With enough samples (e.g. with the global majorant), our equidistant sampling combs coupled with the symmetrization provided by U-statistics already provide a significant efficiency improvement.
- Enabling the endpoint matching control variate allows in some areas a relatively large variance reduction, but the largest improvement is obtained by combining the previous techniques with our aggressive roulette, that allows using even large tuples by sampling fewer orders.
- In regions with a low-frequency density function, we obtain up to 5 orders of magnitude improvements in efficiency. With higher frequencies our final estimator achieves 2 to 3 orders of magnitude lower variance.

Since our estimator gains efficiency with larger and larger tuples, we also compared to the p-series CMF estimator with varying majorants $\bar{\mu}$ (1, 10, and 100 times larger than the tight per-pixel majorant) against single evaluations of our estimators with tuple size M = DetermineTupleSize($\bar{\mu}$); see Figure 15. This comparison reveals that using larger majorants with the p-series CMF estimator can be very detrimental at low optical thicknesses; the majorant effectively acts as a worse and worse pivot. This leads to increased variance from its stochastic Taylor approximations: the Russian roulette continuation probability after the CMF threshold of 99% approaches zero with increasing pivots, while there is still 1% of the contributions left to integrate. In the second and third row, the increasing majorant appears to squeeze the large variance (low efficiency) bump at the center of the first row (using the tight majorant) towards the bottom, where the transmittance T approaches 1.

Implementation. Our unidirectional GPU path tracer is implemented in the open source Falcor [Benty et al. 2020] framework. We report the rendering times of the 1080p images on an NVIDIA Quadro RTX 6000 graphics card with the following disclaimer: Our implementation does not properly balance the workloads of the transmittance estimates inside the thread groups. This results in a SIMT-based slowdown for methods that rely on Russian roulette for cost amortization, such as our unbiased ray marching. The poor utilization of thread groups explains the higher GPU times of unbiased ray marching in Figure 11: synchronizing the evaluation orders makes the unbiased ray marcher as fast as the others, but causes occasional correlation artifacts. Using queues and persistent threads might offer a better solution.

6 DISCUSSION

In Section 5, we have seen how our novel unbiased ray-marching estimator provides a major efficiency improvement across all our tests compared to previous state of the art, and how the biased ray-marching solution reaches even lower MSE at equal cost.

In the following we discuss a different perspective on our U-statistics estimator as well as alternative strategies to equidistant combing and connections to the more general theme of sample stratification.

Complex factorization of the truncated Taylor polynomial. Another path to obtaining our U-statistics estimator is to use the complex

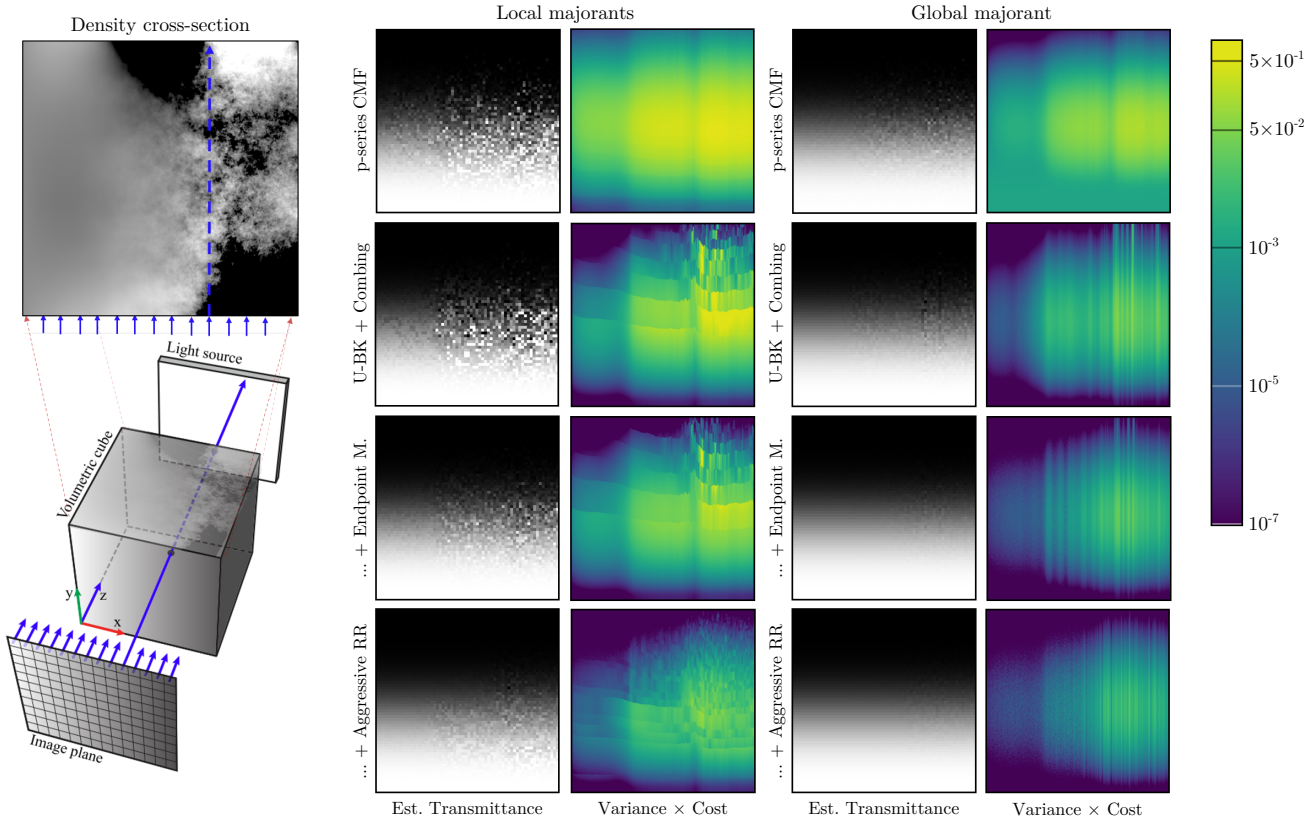

Fig. 14. An ablation study of the unbiased ray marching estimator by simulating transmittance through a 3D slab with varying density (bottom left). The slab is lit from behind with a directional light of intensity 1 such that the pixels of the image plane capture the transmittance through the volumetric cube. The density cross-section of the top plane is shown in the top left: the fractal dimension and variance increase towards the right. The cross-section is the same at every height, but the average density is varied such that transmittance is 1 at the bottom of the image plane, and roughly 5×10^{-5} at the top.

Rows: The rows compare (1) the p-series CMF estimator; (2) our U-BK estimator ($c = 2$) with sampled pivot and combing; (3) the same with our endpoint matching control variate added; and (4) our full unbiased ray marching method which adds the aggressive roulette scheme (Section 4.5). All U-BK variants employ automatic tuple size deduction (Section 4.6) to match the expected sample count to p-series CMF.

Columns: The two first columns show equal sample count results with tight per-pixel majorants, whereas the two last columns show equal sample count results using a single global majorant. The odd columns show the result of a single evaluation, while the even columns show the inverse efficiency (lower is better). Note the logarithmic color scale: our final estimators result in two to five orders of magnitude better efficiency than previous state-of-the-art.

factorization of the truncated power-series polynomial:

$$\sum_{k=1}^{N} \frac{1}{k!\,P_k} \prod_{i}^{k} X_i = c_0 \prod_{i=1}^{N} (X_i - c_i) \tag{43}$$

and apply the generic estimator of products of unbiased estimators recently suggested by Lee et al. [2019, Eq.(6)]. The resulting permuted estimator matches exactly our U-statistics estimator, and despite the presence of complex coefficients, the imaginary part cancels out. Importantly though, our algorithms can exploit the structure of the truncated Taylor series to evaluate all combinations in $O(NZ)$ time, whereas the direct evaluation of Lee et al. [2019, Eq.(6)] is #P-hard.

Connections of combing to stratified sampling. Combing can be seen as a form of stratified sampling applied to each *individual* estimate of the integral of the null density. It is important to note that the separate integral estimates are uncorrelated. Using a single stratified set of random numbers across all orders is not possible, as that would result in correlated integral estimates whose products would result in biased estimates of the powers of $-\tau$. Georgiev et al. [2019] had previously suggested another form of stratification, *across multiple evaluations* of the transmittance integral. This form of stratification is orthogonal and can be combined with our approach: it is sufficient to stratify the Cranley-Patterson rotations $(x_0, ..., x_N)$ across different evaluations of the estimator, for example using Latin hypercube sampling or some other randomized QMC sequence.

Interpretation as non-local tracking estimators. Given the equivalence of certain p-series and tracking estimators highlighted earlier [Georgiev et al. 2019], it is worth noting that combing can be applied also to, e.g., (residual) ratio tracking. Instead of performing a single lookup at each point of the PPP, the tracker would compute

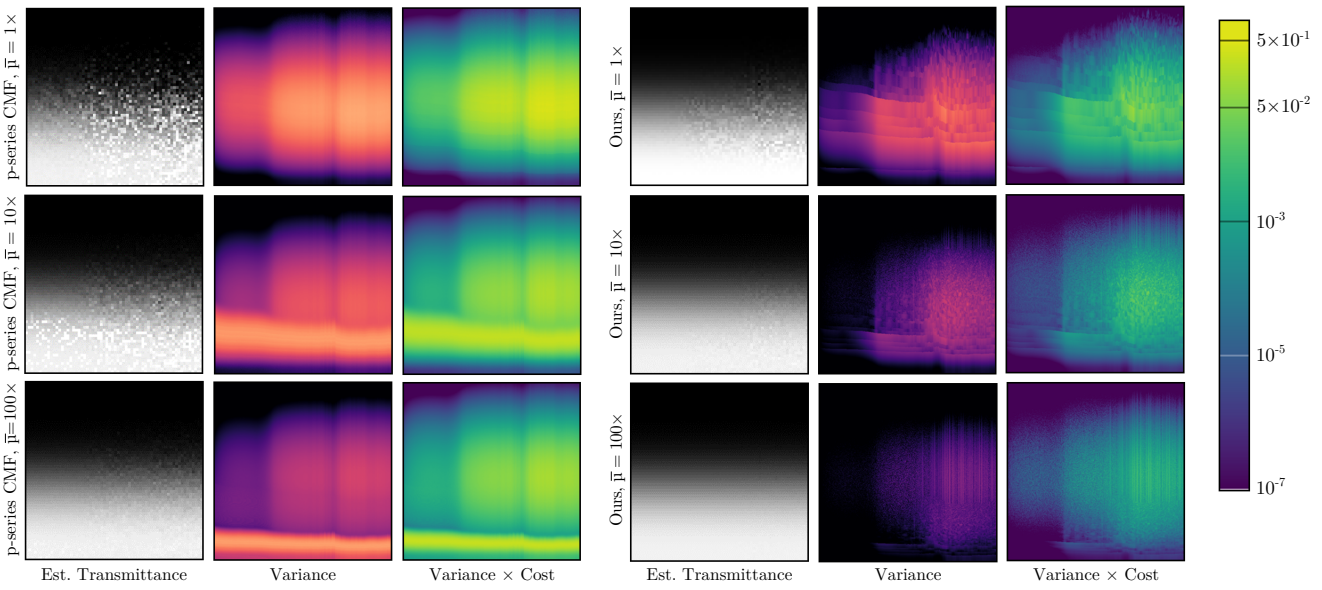

Fig. 15. This figure uses the setup of Figure 14 to compare the variance and inverse efficiency of the p-series CMF estimator (left 3 columns) to that of our unbiased estimator (right 3 columns) when the majorant is raised respectively by 1, 10, and 100 times compared to the tight per-pixel majorant, increasing the number of density evaluations. Notice how at low optical thickness values the original p-series CMF estimator can even suffer from raising the majorant above a certain point due to the use of a very bad pivot. Our estimator is able to use all the available density evaluations to reduce variance and improve efficiency.

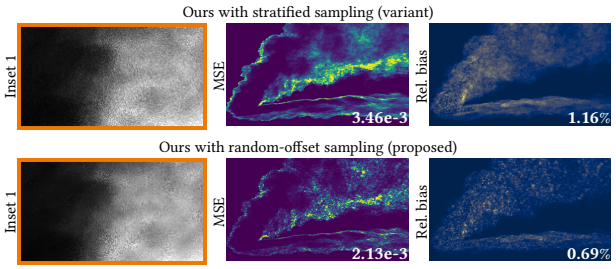

Fig. 16. A comparison of our biased method to a variant that uses stratified ("jittered") sampling. Random-offset equidistant sampling (below) calculates lower-variance optical thickness estimates, which results in less bias and variance when applied through the exponential function. Without sampling the higher orders, both variants result in a relatively small amount of bias. The numbers are averages over the full image.

weights (or collision probabilities) using a combed lookup. Such a *non-local* tracking algorithm would feature lower variance and could be further improved using endpoint matching and sampled pivots. However, given that it would still lack the advantages of the truncated estimators discussed in Section 4.2.1, and the benefits of our U-statistics approach, we draw this connection primarily for didactic purposes and completeness.

Alternative strategies for reducing Y-variance. An equidistant sampling comb works well under the assumption that the density has bounded slope: in this case it can potentially reduce the integration error to $O(1/M)$ or less.

For highly discontinuous densities, or densities with a very high fractal dimension, this might no longer be the case. An alternative in these extreme cases could be using CP-rotated low discrepancy blue-noise combs that react less to the spectrum of the integrand. An example of such a comb can be easily obtained using $u_j = \ell \cdot \phi \cdot j$, where ϕ is the well known golden ratio.

In practice, however, we have found equidistant sampling to always outperform any other low-discrepancy set we have tried. This might be related to the observations of Ramamoorthi et al. [2012].

6.1 Future Work

As briefly discussed in Section 4.6, it would be desirable to have a scheme for deriving the tuple sizes adaptively when other sources of variance are present. The superlinear convergence of our estimators might benefit from drawing more samples in relevant regions of path space, and fewer elsewhere; we leave these investigations for future work. Another potentially related point that deserves attention is a more thorough investigation of the bias/variance tradeoff of our biased and unbiased estimators.

There may be scenarios where negative transmittance estimates are undesired or the sample budget is fixed independent of a majorant optical depth. What estimator performs best in these cases remains an open question. Finally, power series estimation of zero-order probabilities for random media with non-exponential transmission laws, where $\mu(x)$ is a random variable, is an interesting open area [Bitterli et al. 2018; d'Eon 2019; Jarabo et al. 2018], and some steps in this direction using the Master equation for binary mixtures has already been made [Longo 2002].

7 CONCLUSION

We presented a novel in-depth variance analysis (Section 3) of existing unbiased transmittance estimators, revealing weaknesses and areas for improvement. We then proposed a series of techniques (Section 4) exploiting these insights, specifically:

- We have presented a novel power series estimator utilizing all samples efficiently using U-statistics, a recipe for evaluating the estimator in quadratic time, and a numerically robust, incremental elementary symmetric means algorithm.
- We have demonstrated how to further reduce variance by using sampled mean pivots instead of majorant-derived ones; a development enabled by the U-statistics.
- We have described a combed estimator for evaluating optical depth using M rotated equidistant samples and proposed an affine CV to preserve its superlinear convergence rate.
- We have proposed to alter the BK roulette and make it vastly more aggressive, enabling us to use larger tuples and attain even higher overall efficiency.

Since the zeroth-order term of our final power series estimator is analogous to the classical ray marching solution (with the addition of our endpoint matching control variate), we refer to the novel estimator as *unbiased ray marching*. In our tests, unbiased ray marching was universally more effective than any of the previously known unbiased estimators, and often offered several orders of magnitude lower variance at equal sample count. Moreover, we have shown that stopping the power series evaluation at the zeroth order, and effectively getting back to simple ray marching, results in a low-bias estimator with our tuple size deduction approach. This biased estimator attains lower MSE than the unbiased estimator, even at relatively low sample counts, and represents a viable option for real-time rendering and other applications where unbiasedness is not crucial.

ACKNOWLEDGMENTS

We thank Matt Pharr and the anonymous reviewers for their suggestions, which improved the paper.

REFERENCES

John Amanatides and Andrew Woo. 1987. A Fast Voxel Traversal Algorithm for Ray Tracing. In *EG 1987-Technical Papers*. Eurographics Association. https://doi.org/10.2312/egtp.19871000

Nir Benty, Kai-Hwa Yao, Petrik Clarberg, Lucy Chen, Simon Kallweit, Tim Foley, Matthew Oakes, Conor Lavelle, and Chris Wyman. 2020. The Falcor Rendering Framework. https://github.com/NVIDIAGameWorks/Falcor

H. W. Bertini. 1963. *Monte Carlo simulations on intranuclear cascades.* Technical Report ORNL–3383. Oak Ridge National Laboratory, Oak Ridge, TN, USA. https://doi.org/10.2172/4692927

Alexandros Beskos, Omiros Papaspiliopoulos, Gareth O Roberts, and Paul Fearnhead. 2006. Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68, 3 (2006), 333–382. https://doi.org/10.1111/j.1467-9868.2006.00552.x

Gyan Bhanot and Anthony D Kennedy. 1985. Bosonic lattice gauge theory with noise. *Physics letters B* 157, 1 (1985), 70–76. https://doi.org/10.1016/0370-2693(85)91214-6

Benedikt Bitterli, Srinath Ravichandran, Thomas Müller, Magnus Wrenninge, Jan Novák, Steve Marschner, and Wojciech Jarosz. 2018. A radiative transfer framework for non-exponential media. *ACM Transactions on Graphics* 37, 6 (2018). https://doi.org/10.1145/3272127.3275103

Thomas E Booth. 2007. Unbiased Monte Carlo estimation of the reciprocal of an integral. *Nuclear science and engineering* 156, 3 (2007), 403–407. https://doi.org/10.13182/NSE07-A2707

NP Buslenko, DI Golenko, Yu A Shreider, I.M. Sobol', , and VG Sragovich. 1966. *The Monte Carlo method: the method of statistical trials.* Vol. 87. Pergamon.

J. C. Butcher and H. Messel. 1958. Electron Number Distribution in Electron-Photon Showers. *Phys. Rev.* 112 (Dec. 1958), 2096–2106. Issue 6. https://doi.org/10.1103/PhysRev.112.2096

RH Cameron. 1954. The generalized heat flow equation and a corresponding Poisson formula. *Annals of Mathematics* (1954), 434–462. https://doi.org/10.2307/1969711

LL Carter, ED Cashwell, and WM Taylor. 1972. Monte Carlo sampling with continuously varying cross sections along flight paths. *Nucl. Sci. Eng* 48 (1972), 403–411. https://doi.org/10.13182/NSE72-1

Subrahmanyan Chandrasekhar. 1960. *Radiative Transfer.* Dover.

Nan Chen and Zhengyu Huang. 2012. Brownian meanders, importance sampling and unbiased simulation of diffusion extremes. *Operations research letters* 40, 6 (2012), 554–563. https://doi.org/10.1016/j.orl.2012.09.010

W. A. Coleman. 1968. Mathematical Verification of a Certain Monte Carlo Sampling Technique and Applications of the Technique to Radiation Transport Problems. *Nuclear Science and Engineering* 32, 1 (1968), 76–81. https://doi.org/10.13182/NSE68-1

DR Cox and PAW Lewis. 1966. *The statistical analysis of Series of Events.* Wiley.

SN Cramer. 1978. Application of the fictitious scattering radiation transport model for deep-penetration Monte Carlo calculations. *Nuclear Science and Engineering* 65, 2 (1978), 237–253. https://doi.org/10.13182/NSE78-A27154

Eugene d'Eon. 2019. A reciprocal formulation of nonexponential radiative transfer. 1: Sketch and motivation. *Journal of Computational and Theoretical Transport* 48, 6 (2019), 201–262. https://doi.org/10.1080/23324309.2018.1481433

Mouna El Hafi, Stephane Blanco, Jérémi Dauchet, Richard Fournier, Mathieu Galtier, Loris Ibarrart, Jean-Marc Tregan, and Najda Villefranque. 2021. Three viewpoints on null-collision Monte Carlo algorithms. *Journal of Quantitative Spectroscopy and Radiative Transfer* 260, 107402. https://doi.org/10.1016/j.jqsrt.2020.107402

Paul Fearnhead, Omiros Papaspiliopoulos, and Gareth O Roberts. 2008. Particle filters for partially observed diffusions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70, 4 (2008), 755–777. https://doi.org/10.1111/j.1467-9868.2008.00661.x

M. Galtier, S. Blanco, Cyril Caliot, C. Coustet, J. Dauchet, Mouna El-Hafi, Vincent Eymet, R. Fournier, J. Gautrais, A. Khuong, B. Piaud, and Guillaume Terrée. 2013. Integral formulation of null-collision Monte Carlo algorithms. *Journal of Quantitative Spectroscopy and Radiative Transfer* 125 (Aug. 2013), 57–68. https://doi.org/10.1016/j.jqsrt.2013.04.001

Iliyan Georgiev, Zackary Misso, Toshiya Hachisuka, Derek Nowrouzezahrai, Jaroslav Křivánek, and Wojciech Jarosz. 2019. Integral formulations of volumetric transmittance. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia)* 38, 6 (Nov. 2019). https://doi.org/10/dffn

Mark Girolami, Anne-Marie Lyne, Heiko Strathmann, Daniel Simpson, and Yves Atchade. 2013. Playing Russian roulette with intractable likelihoods. *arXiv preprint arXiv:1306.4032* (2013). https://arxiv.org/abs/1306.4032

Gerald J Glasser. 1962. Minimum variance unbiased estimators for Poisson probabilities. *Technometrics* 4, 3 (1962), 409–418. https://doi.org/10.1080/00401706.1962.10490021

Paul R Halmos. 1946. The theory of unbiased estimation. *The Annals of Mathematical Statistics* (1946), 34–43. https://doi.org/10.1214/aoms/1177731020

Pierre E Jacob, Alexandre H Thiery, et al. 2015. On nonnegative unbiased estimators. *The Annals of Statistics* 43, 2 (2015), 769–784. https://doi.org/10.1214/15-AOS1311

Adrian Jarabo, Carlos Aliaga, and Diego Gutierrez. 2018. A Radiative Transfer Framework for Spatially-Correlated Materials. *ACM Transactions on Graphics* 37, 4 (2018), 14. https://doi.org/10.1145/3197517.3201282

NL Johnson. 1951. Estimators of the probability of the zero class in Poisson and certain related populations. *The Annals of Mathematical Statistics* 22, 1 (1951), 94–101. https://doi.org/10.1214/aoms/1177729696

Daniel Jonsson, Joel Kronander, Jonas Unger, Thomas B Schon, and Magnus Wrenninge. 2020. Direct Transmittance Estimation in Heterogeneous Participating Media Using Approximated Taylor Expansions. *IEEE Transactions on Visualization and Computer Graphics* (2020). https://doi.org/10.1109/TVCG.2020.3035516

Robert W Klein and Stephen D Roberts. 1984. A time-varying Poisson arrival process generator. *Simulation* 43, 4 (1984), 193–195. https://doi.org/10.1177/003754978404300406

Christopher Kulla and Marcos Fajardo. 2011. Importance Sampling of Area Lights in Participating Media. *ACM SIGGRAPH 2011 Talks, SIGGRAPH'11*, 55. https://doi.org/10.1145/2037826.2037899

Anthony Lee, Simone Tiberi, and Giacomo Zanella. 2019. Unbiased approximations of products of expectations. *Biometrika* 106, 3 (Sept. 2019). https://doi.org/10.1093/biomet/asz008

A J Lee. 1990. *U-statistics: Theory and Practice.* Routledge.

David Legrady, Balazs Molnar, Milan Klausz, and Tibor Major. 2017. Woodcock tracking with arbitrary sampling cross section using negative weights. *Annals of Nuclear Energy* 102 (04 2017), 116–123. https://doi.org/10.1016/j.anucene.2016.12.003

L Lin, K F Liu, and J Sloan. 2000. A noisy Monte Carlo algorithm. *Physical Review. D, Particles Fields* 61 (Apr 2000). Issue 7. https://doi.org/10.1103/PhysRevD.61.074505

Savino Longo. 2002. Direct derivation of Skullerud's Monte Carlo method for charged particle transport from the linear Boltzmann equation. *Physica A: Statistical Mechanics and its Applications* 313, 3-4 (2002), 389–396. https://doi.org/10.1016/S0378-4371(02)01007-5

Anne-Marie Lyne, Mark Girolami, Yves Atchadé, Heiko Strathmann, Daniel Simpson, et al. 2015. On Russian roulette estimates for Bayesian inference with doubly-intractable likelihoods. *Statistical science* 30, 4 (2015), 443–467. https://doi.org/10.1214/15-STS523

D. G. Mead. 1992. Newton's Identities. *The American Mathematical Monthly* 99, 8 (1992), 749–751. http://www.jstor.org/stable/2324242

GA Mikhailov. 1970. A method for simulating the mean free path of a particle. *Soviet Atomic Energy* 28, 2 (1970), 224–225. https://doi.org/10.1007/BF01162640

Gennadii A. Mikhailov. 1992. *Optimization of weighted Monte Carlo methods.* Springer. https://www.springer.com/gp/book/9783642759833

Rupert G Miller. 1974. The jackknife-a review. *Biometrika* 61, 1 (1974), 1–15. https://doi.org/10.1093/biomet/61.1.1

Sarat Babu Moka, Dirk P Kroese, and Sandeep Juneja. 2019. Unbiased estimation of the reciprocal mean for non-negative random variables. In *2019 Winter Simulation Conference (WSC)*. IEEE, 404–415. https://doi.org/10.1109/WSC40007.2019.9004815

Ken Museth. 2013. VDB: High-Resolution Sparse Volumes with Dynamic Topology. *ACM Trans. Graph.* 32, 3, Article 27 (July 2013), 22 pages. https://doi.org/10.1145/2487228.2487235

Adolfo Muñoz. 2014. Higher Order Ray Marching. *Computer Graphics Forum* 33, 8 (2014), 167–176. https://doi.org/10.1111/cgf.12424

Jerzy Neyman and Elizabeth L Scott. 1960. Correction for bias introduced by a transformation of variables. *The Annals of Mathematical Statistics* 31, 3 (1960), 643–655. https://www.jstor.org/stable/2237574

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/10.1111/cgf.13383

Jan Novák, Andrew Selle, and Wojciech Jarosz. 2014. Residual ratio tracking for estimating attenuation in participating media. *ACM Trans. Graph.* 33, 6 (2014), 179–1. https://doi.org/10.1145/2661229.2661292

Omiros Papaspiliopoulos. 2011. *Monte Carlo probabilistic inference for diffusion processes: a methodological framework.* Cambridge University Press, 82–103. https://doi.org/10.1017/CBO9780511984679.005

Raghu Pasupathy. 2011. *Generating Nonhomogeneous Poisson Processes.* American Cancer Society. https://doi.org/10.1002/9780470400531.eorms0356

Mark Pauly, Thomas Kollig, and Alexander Keller. 2000. Metropolis Light Transport for Participating Media. *Rendering Techniques* 2000 (11 2000). https://doi.org/10.1007/978-3-7091-6303-0_2

Matthias Raab, Daniel Seibert, and Alex Keller. 2006. Unbiased Global Illumination with Participating Media. In *Monte Carlo and Quasi Monte Carlo Methods 2006.* Springer, 591–601. https://doi.org/10.1007/978-3-540-74496-2_35

Ravi Ramamoorthi, John Anderson, Mark Meyer, and Derek Nowrouzezahrai. 2012. A Theory of Monte Carlo Visibility Sampling. *ACM Trans. Graph.* 31, 5, Article 121 (Sept. 2012). https://doi.org/10.1145/2231816.2231819

H. R. Skullerud. 1968. The stochastic computer simulation of ion motion in a gas subjected to a constant electric field. *Journal of Physics D: Applied Physics* 1, 11 (1968), 1567–1568. https://doi.org/10.1088/0022-3727/1/11/423

László Szirmay-Kalos, Balázs Tóth, and Milán Magdics. 2011. Free path sampling in high resolution inhomogeneous participating media. In *Computer Graphics Forum*, Vol. 30. Wiley Online Library, 85–97. https://doi.org/10.1111/j.1467-8659.2010.01831.x

Wolfgang Wagner. 1987. Unbiased Monte Carlo evaluation of certain functional integrals. *J. Comput. Phys.* 71, 1 (1987), 21–33. https://doi.org/10.1016/0021-9991(87)90017-9

Wolfgang Wagner. 1988. Monte Carlo evaluation of functionals of solutions of stochastic differential equations. Variance reduction and numerical examples. *Stochastic Analysis and Applications* 6, 4 (1988), 447–468. 10.1080/07362998808809161

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.* Technical Report ANL-7050. Argonne National Laboratory.

C. D. Zerby, R. B. Curtis, and H. W. Bertini. 1961. *The relativistic doppler problem.* Technical Report ORNL-61-7-20. Oak Ridge National Laboratory, Oak Ridge, TN, USA. https://doi.org/10.2172/4836227

A OPTIMALITY OF THE MEAN PIVOT

Earlier, we discussed how non-symmetric power series estimators benefit from a majorant pivot and how U-statistics changes this behaviour. With this important difference, we perform a similar analysis for the pivot as in earlier work (e.g. Georgiev et al. [2019]): We analyze the sum of the absolute values of the different order contributions in the Taylor series of $e^{\mathbb{E}[X]}$ with different pivots p. The sum of the absolute values of the contributions from all orders with pivot p is

$$e^p\left(1 + \left|\mathbb{E}[X] - p\right| + \frac{\left|\mathbb{E}[X] - p\right|^2}{2!} + \cdots\right) = e^{p + \left|\mathbb{E}[X] - p\right|}. \tag{44}$$

This says that in some sense, it is optimal to use any pivot less than the expectation, $p \leq \mathbb{E}[x]$, as any such pivot minimizes the above expression. In terms of positive densities, this says that the control density should be at least as high as the mean density. However, the control density does not need to be greater than all of the density samples—there is no need to use a majorant.

However, the picture changes drastically when we take Russian roulette into account: As noted earlier, moving the pivot closer to $-\tau$ implies faster convergence for the Taylor series. This means that we need to evaluate fewer orders of the power series for good estimates, which means that we can employ more aggressive Russian roulette.

Interestingly, with a good pivot, even aggressive Russian roulette will not make efficiency worse: for N total evaluations of an estimator X, returning the roulette-compensated variable with probability p and otherwise zero, the actual number of evaluations is Np. The inverse efficiency of the estimator is thus proportional to

$$\begin{aligned} p \operatorname{Var}\left[\frac{RX}{p}\right] &= p\left(\frac{\mathbb{E}[R^2]\,\mathbb{E}[X^2]}{p^2} - \frac{\mathbb{E}[R]^2\,\mathbb{E}[X]^2}{p^2}\right) \\ &= \mathbb{E}[X^2] - p\,\mathbb{E}[X]^2 = \operatorname{Var}[X] + (1-p)\,\mathbb{E}[X]^2, \end{aligned} \tag{45}$$

where R is the random binary choice variable. This says that the efficiency of the roulette is maximized when $\mathbb{E}[X] = 0$, that is, when we use the theoretical mean pivot. The efficiency of the rouletted estimator decreases as the pivot moves farther from the real expectation.

Therefore, using an approximate mean pivot allows the use of a more aggressive roulette.

The resulting lower mean estimation order from more aggressive roulette means that we now need a smaller number of independent samples, and we can use our density evaluation budget to make those samples higher-quality by performing variance reduction techniques such as stratification or numerical integration rules, as discussed in Section 4.

B ELEMENTARY SYMMETRIC MEANS

In this appendix we derive the elementary symmetric means formulas that lead to Algorithm 1.

Elementary symmetric sums $e_k = e_k(x_1, \cdots, x_n)$ are defined as

$$e_k = \sum_{1 \leq i_1 < \cdots < i_k \leq n} x_{i_1} \cdots x_{i_k}, \tag{46}$$

with $e_0 = 1$. To distinguish between different numbers of parameters, in this appendix we denote the elementary sums of n variables, $x_1, \ldots, x_n$, by

$$e_k^n = e_k(x_1, \cdots, x_n), \tag{47}$$

and elementary symmetric means by

$$m_k^n = m_k(x_1, \cdots, x_n) = \frac{e_k(x_1, \cdots, x_n)}{\binom{n}{k}}. \tag{48}$$

A simple derivation leads to a formula for iteratively constructing elementary symmetric sums:

$$\begin{aligned} e_k^{n+1} &= \sum_{1\le i_1<\cdots<i_k\le n+1} x_{i_1}\cdots x_{i_k} \\ &= \sum_{1\le i_1<\cdots<i_k\le n} x_{i_1}\cdots x_{i_k} + \sum_{1\le i_1<\cdots<i_k=n+1} x_{i_1}\cdots x_{i_k} \\ &= e_k^n + \left(\sum_{1\le i_1<\cdots<i_{k-1}\le n} x_{i_1}\cdots x_{i_{k-1}}\right) x_{n+1} \\ &= e_k^n + e_{k-1}^n x_{n+1}. \end{aligned} \tag{49}$$

Then, by substituting the definition of elementary symmetric means, we reach the recurrence formula

$$m_k^{n+1} = m_k^n + \frac{k}{n+1}\left(m_{k-1}^n x_{n+1} - m_k^n\right). \tag{50}$$

Observing the directions of the dependencies in this formula leads to Algorithm 1.

C EFFICIENCY DERIVATIONS

In this appendix we review known analytic results for the variance and cost of transmittance estimators as well as present some new derivations for power series estimators.

C.1 Costs

Tracking estimators. The expected number $\mathbb{E}[N]$ of optical-depth estimates for a tracking estimator follows from the mean of the Poisson distribution, which is simply the rate, $\mathbb{E}[N] = \lambda_\ell = \bar{\tau}_\mathrm{r}$. Therefore, for residual ratio tracking with query size M,

$$\mathrm{Cost}[\widehat{T}_{rrt}] = M\,\bar{\tau}_\mathrm{r}. \tag{51}$$

Bhanot and Kennedy roulette. Using the continuation probabilities of the generalized BK estimator (Equation 19) we find the probability $Q_{\mathrm{BK}}(N)$ of evaluating term N to be

$$Q_{\mathrm{BK}}(N) = \frac{c}{K+1}\frac{c}{K+2}\cdots\frac{c}{K+N} = \frac{c^{N-K}}{N!/K!}, \quad N > K = \lfloor c \rfloor, \tag{52}$$

and 1 otherwise. The expected number of evaluated orders is thus

$$\mathbb{E}[N_{BK}] = K + \sum_{N=K+1}^{\infty}\frac{c^{N-K}}{N!/K!} = K + \frac{K!}{c^K}\left(e^c - \sum_{N=0}^{K}\frac{c^N}{N!}\right). \tag{53}$$

C.2 Variances

Delta-tracking. The exact variance of delta-tracking is known [Glasser 1962] and agrees with a derivation for the special case of $n = 1$ and a uniform-medium [Georgiev et al. 2019]

$$\mathrm{Var}[\widehat{T}_{dt}] = e^{-\tau} - e^{-2\tau}. \tag{54}$$

This result is exact for any input, and generalizes [Glasser 1962] to Johnson's $n > 1$ estimator (Section 2.3):

$$\mathrm{Var}[\widehat{T}_J] = e^{-2\tau}\left(e^{\frac{\tau}{n}} - 1\right). \tag{55}$$

We investigate the efficiency of Johnson's estimator in the supplementary material.

Residual ratio tracking. The exact variance for residual ratio tracking is also known. The rate of the Poisson process follows from the difference of the known optical depths, $\lambda = \bar{\tau} - \tau_\mathrm{c}$, which are the integrals of the upper $\bar{\mu}(x)$ and lower $\mu_\mathrm{c}(x)$ control variates. The variance is then [Papaspiliopoulos 2011, Equation 4.16]

$$\mathrm{Var}[\widehat{T}_{rrt}] = e^{-2\,\bar{\tau}+\lambda+\frac{V}{\lambda}} - e^{-2\tau}, \quad \text{where} \tag{56}$$

$$V = \frac{1}{(b-a)}\int_a^b \left((b-a)(\bar{\mu}-\mu(x))\right)^2 dx. \tag{57}$$

C.2.1 Truncated estimators.

Roulette variance. With uniform density, the negative residual optical depth Y is estimated with zero variance and the full variance of the generalized BK estimator is (see supplemental material)

$$\mathrm{Var}[\widehat{T}_{BK}] = e^{-2\tau_\mathrm{c}}\sum_{j=0}^{\infty}\frac{c^j\left(1-\frac{c}{j+K+1}\right)}{(K+1)_j}\left(\sum_{n=0}^{K}\frac{Y^n}{n!} + \sum_{i=1}^{j}\frac{Y^{K+i}}{c^i K!}\right)^2 - T^2.$$

Additional derivations. In the supplemental material we provide three additional variance derivations for BK estimators in the special case that the pivot is optimal: the variances of generalized BK and U-BK estimators with fixed expansion ($K > 0, c = 0$), and the variance of generalized BK with roulette ($c > 0$).

D ENDPOINT MATCHING FORMULAS

Integration with endpoint matching and equidistant combs can be further simplified. As the order of the samples in a CP-rotated equidistant comb does not matter, we can write the optical thickness estimate X_i of the endpoint-matched volume as

$$X_i = -\frac{\ell}{M}\sum_{i=0}^{M-1}\mu^\star\left(\frac{\ell}{M}(u+i)\right), \tag{58}$$

where u is a uniform random number in $[0, 1)$. With Equation 35 this directly simplifies into

$$X_i = \underbrace{-\frac{\ell}{M}\sum_{i=0}^{M-1}\mu\left(\frac{\ell}{M}(u+i)\right)}_{\text{original estimate}} \underbrace{-\frac{\ell}{M}\left(\frac{1}{2}-u\right)\left(\mu(\ell)-\mu(0)\right)}_{\text{endpoint matching}}. \tag{59}$$

The left-hand-side term is the integral estimate without endpoint matching, and the right-hand-side is the zero-expectation term from endpoint matching that often improves the convergence rate.

This reshuffling can cause the resulting integrand to assume negative values. If non-negativity is a constraint, an alternate option is to symmetrize the estimator over the interval by using the mean of mirrored lookups [Buslenko et al. 1966] (p.106). This is equivalent to blending the interval of scattering material with its reversed copy.