# Sample Matching for Joint Extinction Gradient Estimation in Differentiable Volume Rendering

RUIHAN YU, The University of Tokyo, Japan and Tsinghua University, China
YU-CHEN WANG, University of California, Irvine, USA
JINGWANG LING, University of Illinois Urbana-Champaign, USA
FENG XU, Tsinghua University, China
SHUANG ZHAO, University of Illinois Urbana-Champaign, USA

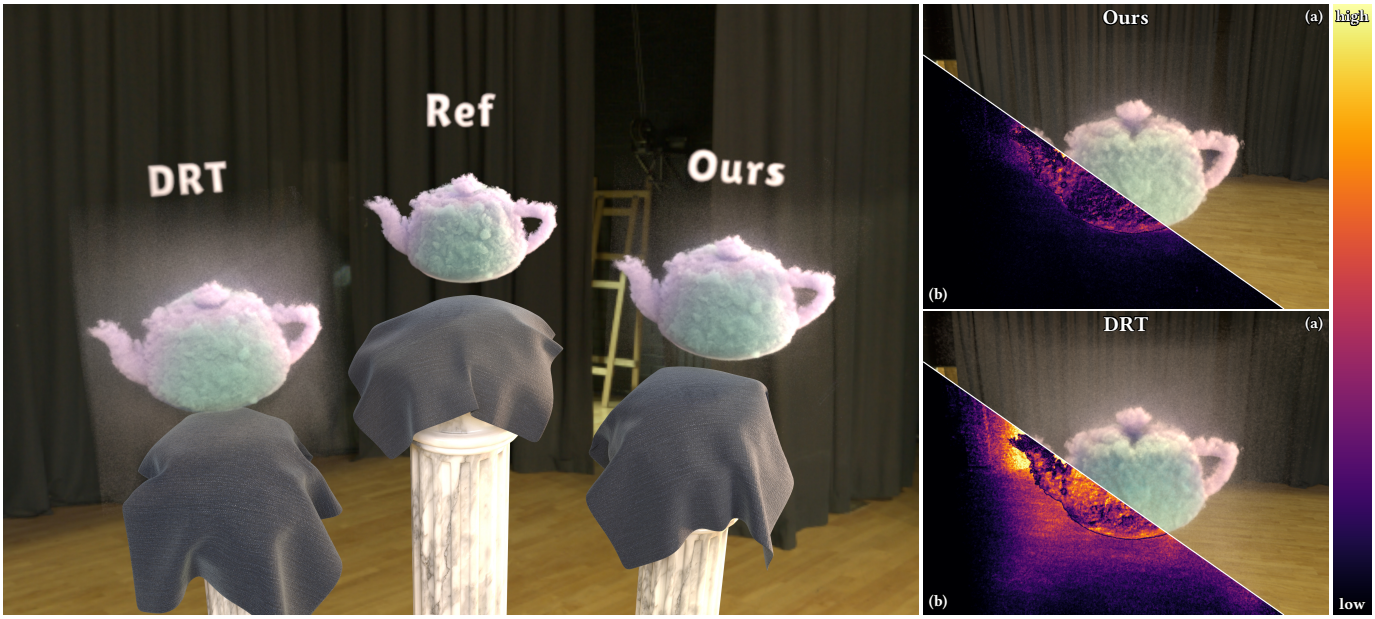


Fig. 1. Our differentiable volume rendering approach computes scattering and transmittance gradients jointly, achieving higher-quality reconstructions with lower variance. *Left:* a relighting scene comparing reconstructions from DRT and our method against the reference. *Right:* per-method detail panels for the same scene; each panel is split diagonally into (a) the rendered result and (b) the per-pixel error map against the reference.

Differentiable volume rendering enables gradient-based optimization of volumetric scenes, but unbiased estimators suffer from high gradient variance. We observe that the extinction gradients split into two components on structurally different integration domains: a scattering term evaluated at a single path vertex, and a transmittance term integrated along the ray segment. Because the domains are mismatched, existing estimators sample the two components at different locations, leaving the negative correlation between their opposite-signed contributions unexploited. We expose this overlooked correlation and exploit it through a principle we call *sample matching*: evaluate both components at shared sample locations. To enable this, we derive the first reformulation of the differential path integral that couples the two contributions within a single integrand, yielding an unbiased Monte Carlo estimator that ties them together by construction. For efficiency, the estimator reuses partially sampled light paths and amortizes in-scattering cost by evaluating gradients at multiple probe points per segment. On voxel-grid reconstruction, our estimator reduces gradient variance by up to 80% over differential ratio tracking (DRT), yielding faster convergence and higher reconstruction quality.

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

Additional Key Words and Phrases: Volume Rendering, Differentiable Rendering, Inverse Rendering, Variance Reduction

Authors' Contact Information: Ruihan Yu, auroraryan0301@gmail.com, The University of Tokyo, Japan and Tsinghua University, China; Yu-Chen Wang, yuchew23@uci.edu, University of California, Irvine, USA; Jingwang Ling, gerwang@outlook.com, University of Illinois Urbana-Champaign, USA; Feng Xu, xufeng2003@gmail.com, Tsinghua University, China; Shuang Zhao, shzhao@illinois.edu, University of Illinois Urbana-Champaign, USA.

## 1 Introduction

Volume inverse rendering has emerged as a powerful paradigm for reconstructing complex 3D scenes from images, with applications spanning graphics production and scientific imaging—from 3D asset generation [Leonard et al. 2025] and holographic manufacturing [Nicolet et al. 2024] to appearance-accurate 3D printing [Iser et al. 2025], plasma reconstruction [Öztürk et al. 2025], and turbid liquid characterization [Yamakoshi et al. 2025]. Unlike simplified volumetric representations such as NeRF [Mildenhall et al. 2021] or 3D Gaussians [Kerbl et al. 2023], which approximate light transport using a zero-bounce emissive field, methods based on the full radiative transfer theory (RTT) [Chandrasekhar 1960; Kajiya and Von Herzen

1984; Rushmeier and Torrance 1987] recover intrinsic material properties such as extinction and albedo coefficients, enabling faithful relighting and physically meaningful reconstruction of complex heterogeneous media.

Yet inverse rendering under the full radiative transfer theory remains challenging. Differentiable volume renderers estimate gradients via Monte Carlo sampling. Despite sustained progress, from free-flight sampling [Zhang et al. 2019] to differential ratio tracking (DRT) [Nimier-David et al. 2022], path-space formulations [Zhang et al. 2021b], and efficient gradient backpropagation through long differentiable path graphs [Nimier-David et al. 2020; Vicini et al. 2021], all existing unbiased extinction-gradient estimators exhibit variance high enough to slow convergence and degrade reconstruction quality. Reducing this variance without sacrificing unbiasedness is therefore the central challenge.

The difficulty traces to the structure of the radiative transfer formulation itself. Extinction appears both explicitly, as a local attenuation term, and implicitly through transmittance. Differentiating therefore decomposes the gradient into two components with structurally different integration domains: a *scattering component* evaluated at a single path vertex, and a *transmittance component* integrated over the segment interior. The two correspond to competing physical effects—raising extinction at a point enhances in-scattering there while attenuating radiance transmitted through it—and, crucially, they enter the gradient with *opposite signs*. This opposite-sign structure invites negative correlation, yet all existing estimators leave it unexploited: free-flight sampling [Zhang et al. 2019] places the two components at geometrically distinct sample locations, while differential ratio tracking (DRT) [Nimier-David et al. 2022] draws independent samples for each—forcing zero cross-covariance.

In this work, we expose this overlooked negative correlation and exploit it through a principle we call *sample matching*: evaluate the scattering and transmittance components at shared sample locations. To enable this, we derive the first reformulation of the differential path integral that couples the two contributions within a single integrand, yielding an unbiased estimator that ties them together by construction. The resulting negative covariance substantially reduces extinction-gradient variance without sacrificing unbiasedness.

Concretely, this paper makes the following contributions:

- We identify a negative correlation structure inherent to extinction gradients and introduce a **path-space reformulation** that exposes it (Section 4). To our knowledge, this is the first systematic exploitation of this correlation.
- Building on this formulation, we derive an **unbiased Monte Carlo estimator** for extinction gradients (Section 5). For efficiency, the estimator reuses partially sampled light paths and amortizes overhead by evaluating gradients at multiple probe points per segment.

We demonstrate the effectiveness of our approach against state-of-the-art baselines in Figure 5, Figure 6, and Figure 9, with ablation studies in Figure 7 and Figure 10.

## 2 Related Work

### 2.1 Differentiable Volume Rendering

Differentiating the path integral enables gradient-based optimization of volumetric media. Zhang et al. [2019] derive a differential theory of radiative transfer under free-flight sampling, in which the explicit extinction term cancels in the resulting gradient estimator. Differential ratio tracking (DRT) [Nimier-David et al. 2022] retains this term and reduces its variance by drawing scattering samples from a transmittance-proportional distribution via ratio tracking. Path-space differentiable rendering [Zhang et al. 2020; Zhao et al. 2020] was extended to participating media by Zhang et al. [2021b], which unifies interfacial and volumetric light transport in a single generalized differential path integral and thereby handles scenes that mix surfaces and participating media. Reconstruction of translucent objects [Deng et al. 2022] further enables differentiable rendering for subsurface-scattering media, and differential diffusion theory [Cen et al. 2023] provides a complementary formulation for dense media. Our work continues this lineage: we identify a negative-correlation structure in the extinction-gradient decomposition that all existing estimators fail to exploit, and reformulate the differential path integral to expose it; we implement the resulting estimator in the PRB form [Vicini et al. 2021] shared by DRT and other Mitsuba 3 [Jakob et al. 2022] integrators.

### 2.2 Variance Reduction in Differentiable Rendering

Zeltner et al. [2021] showed that estimators designed for the derivative integrand outperform naive reuse of primal samplers, motivating derivative-tailored sampling in the differentiable setting; this line has been refined primarily for surface material-parameter derivatives [Belhe et al. 2024; Fan et al. 2024]. Our reformulation extends it to differentiable volume rendering: expressing the extinction-gradient integrand as a single coupled integral opens up a new design space for derivative-specific importance sampling in the volumetric setting. A complementary family treats the gradient estimator as a black box and applies variance reduction at an outer level—recursive control variates [Nicolet et al. 2023], variance-aware optimization [Yan et al. 2024], differentiable regularization [Wu et al. 2024], and image-space adaptive batching [Yan et al. 2025]; these are orthogonal to our estimator-level reformulation.

A second complementary line exploits correlation between samples. Closest in spirit is antithetic sampling for differentiable rendering [Zhang et al. 2021a], which pairs samples to reduce the variance of scalar derivatives with respect to scene geometry; there, both paired samples live on the same integral measure and no positional mismatch arises. A broader family targets scalar radiometric quantities or their finite-difference gradients by constructing *positively* correlated sample pairs: shift mappings in gradient-domain rendering [Kettunen et al. 2015; Lehtinen et al. 2013], random-replay and primary-sample-space techniques [Kelemen et al. 2002; Vicini et al. 2021], and reservoir-based resampling in forward [Lin et al. 2021] and differentiable rendering [Wang et al. 2023]. Our setting differs in two structural ways. First, the scattering and transmittance gradient components are integrated over different domains (segment endpoints vs. segment interiors), so the negative correlation between them is not accessible by direct sample pairing—aligning positions

across these mismatched integrals is precisely what *sample matching* does. Second, the target is a high-dimensional gradient vector rather than a scalar, so the scalar-variance analyses underlying the above techniques do not directly apply.

### 2.3 Volumetric Path Construction

Heterogeneous volumes are typically traversed by delta tracking [Butcher and Messel 1958; Woodcock et al. 1965] via null collisions, as in DRT. Spectral and decomposition tracking [Kutz et al. 2017] reduce per-sample cost for chromatic media; the null-scattering path integral [Miller et al. 2019], building on path-space extensions [Hachisuka et al. 2017, 2012], makes null-collision PDFs explicit and enables MIS across distance, equiangular, and next-event estimation (NEE) strategies in spatially and spectrally varying media. Progressive null-tracking [Misso et al. 2023] extends this family to procedural media with unknown majorants. A comprehensive survey of Monte Carlo techniques for volumes is provided by Novák et al. [2018]. Our reformulation is derived in path space and composes directly with any of the above path-construction strategies.

### 2.4 Transmittance Estimators

When paths are pre-sampled under a different distribution, as in bidirectional and reservoir-based path construction [Lafortune and Willems 1996; Lin et al. 2021], the transmittance along each segment must be estimated independently. Analog estimators (e.g., score-based delta tracking) and non-analog estimators (ratio and residual-ratio tracking [Novák et al. 2014]) were unified under the integral analysis of Georgiev et al. [2019], which also covers power-series variants [Coleman 1968; Cramer 1978]. Recent additions include unbiased ray marching [Kettunen et al. 2021] and biased Jackknife estimation [Peters 2025]. These non-analog estimators are related to the family of track-length estimators in forward rendering, which originates in photon beams [Jarosz et al. 2011a,b, 2008] and were unified by Křivánek et al. [2014]. Our per-segment multi-probe estimator (Section 5.3) shares the "spread along a track" spirit with track-length estimators: we distribute $\Lambda$ probes along each segment and amortize a single recursive indirect-radiance trace across them. However, in a differentiable setting, our estimator targets a high-dimensional gradient vector while track-length estimators usually target a scalar radiometric quantity.

## 3 Preliminaries

*Note to the reader.* For generality, we develop the formulation directly in the path-space language [Veach 1997; Zhang et al. 2021b]. Readers who prefer to first see the same derivation and the same observation expressed in the more familiar RTE language may find the supplementary material a useful entry point—it gives an RTE-centric tour of sample matching that can be read in lieu of, or before, this section.

Before introducing our technique in Section 4 and Section 5, we briefly revisit the mathematical and algorithmic preliminaries it builds upon. Section 3.1 formulates volumetric light transport using generalized path integrals, Section 3.2 differentiates this formulation with respect to material optical properties, and Section 3.3 reviews Monte Carlo methods for estimating both quantities. Throughout, we follow the notation summarized in Table 1.

### 3.1 Volumetric Light Transport

Volumetric light transport concerns the interaction between light and microscopic particles distributed throughout space. In heterogeneous non-emissive media, these interactions are governed by a spatially varying **extinction coefficient** $\sigma_t$, **single-scattering albedo** $\alpha$, and **phase function** $\rho$. We adopt the path-integral formulation of Veach [1997], which provides a unified framework for path tracing and underpins the derivations throughout this work.

Specifically, a radiometric sensor measurement can be expressed as a **generalized path integral** [Pauly et al. 2000]

$$I = \int_{\Omega} f(\bar{\mathbf{x}})\, \mathrm{d}\bar{\mathbf{x}}, \tag{1}$$

where $\bar{\mathbf{x}} = (\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_K) \in \Omega$ is a light-transport path with $K \geq 1$ segments and $K+1$ vertices. The endpoints $\boldsymbol{x}_0$ and $\boldsymbol{x}_K$ lie on the sensor and a light source, respectively,[1] and the intermediate vertices $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{K-1}$ correspond to scattering events inside the medium volume $\mathcal{V} \subset \mathbb{R}^3$.

We assume no surface interactions inside the volume, so only $\boldsymbol{x}_0$ and $\boldsymbol{x}_K$ are surface vertices and the rest are volumetric. Thus,

$$\mathrm{d}\bar{\mathbf{x}} = \mathrm{d}A(\boldsymbol{x}_0) \left( \prod_{j=1}^{K-1} \mathrm{d}V(\boldsymbol{x}_j) \right) \mathrm{d}A(\boldsymbol{x}_K), \tag{2}$$

where $V$ and $A$ denote the volume and area measures, respectively.

The integrand $f$ is the **measurement contribution function** defined as

$$f(\bar{\mathbf{x}}) = \left( \prod_{j=0}^{K} v_j(\bar{\mathbf{x}}) \right) \left( \prod_{j=1}^{K} s_j(\bar{\mathbf{x}}) \right). \tag{3}$$

Here, each vertex factor $v_j$ is given by

$$v_j(\bar{\mathbf{x}}) \coloneqq \begin{cases} W_\mathrm{e}(\boldsymbol{x}_0 \to \boldsymbol{x}_1) & (j = 0), \\ \sigma_t(\boldsymbol{x}_j)\, \hat{\rho}(\boldsymbol{x}_{j+1} \to \boldsymbol{x}_j \to \boldsymbol{x}_{j-1}) & (0 < j < K), \\ L_\mathrm{e}(\boldsymbol{x}_K \to \boldsymbol{x}_{K-1}) & (j = K), \end{cases} \tag{4}$$

where $L_\mathrm{e}$ and $W_\mathrm{e}$ denote the emitted radiance and the sensor importance, respectively, and

$$\hat{\rho}(\boldsymbol{x}_{j+1} \to \boldsymbol{x}_j \to \boldsymbol{x}_{j-1}) \coloneqq \alpha(\boldsymbol{x}_j)\, \rho(\boldsymbol{x}_{j+1} \to \boldsymbol{x}_j \to \boldsymbol{x}_{j-1}), \tag{5}$$

is the **scaled phase function**. Each segment factor $s_j$ is given by

$$s_j(\bar{\mathbf{x}}) \coloneqq G(\boldsymbol{x}_{j-1} \leftrightarrow \boldsymbol{x}_j)\, T(\boldsymbol{x}_{j-1} \leftrightarrow \boldsymbol{x}_j), \tag{6}$$

where $G$ is the visibility-aware **geometric term**: for any two vertices $\boldsymbol{x}$ and $\boldsymbol{y}$,

$$G(\boldsymbol{x} \leftrightarrow \boldsymbol{y}) \coloneqq \mathbb{V}(\boldsymbol{x} \leftrightarrow \boldsymbol{y}) \frac{D(\boldsymbol{x} \to \boldsymbol{y})\, D(\boldsymbol{y} \to \boldsymbol{x})}{\|\boldsymbol{y} - \boldsymbol{x}\|^2}, \tag{7}$$

$\mathbb{V}$ is the mutual visibility function, and

$$D(\boldsymbol{x} \to \boldsymbol{y}) \coloneqq \begin{cases} \left| \boldsymbol{n}(\boldsymbol{x}) \cdot \overrightarrow{\boldsymbol{xy}} \right|, & (\boldsymbol{x} \text{ is a surface vertex}), \\ 1, & (\text{otherwise}), \end{cases} \tag{8}$$

[1] Conventional path-integral formulations place $\boldsymbol{x}_0$ on the light source and $\boldsymbol{x}_K$ on the sensor. We flip the indexing to better fit unidirectional-path-tracing estimators.

with $\boldsymbol{n}(\boldsymbol{x})$ denoting the surface normal at $\boldsymbol{x}$ and $\overrightarrow{\boldsymbol{xy}}$ the unit direction from $\boldsymbol{x}$ to $\boldsymbol{y}$. For the volume-only scenes considered in this work, the visibility reduces to $\mathbb{V} \equiv 1$; we therefore omit $\mathbb{V}$ from all subsequent equations. Lastly, $T$ is the **transmittance** given by

$$T(\boldsymbol{x}_{j-1} \leftrightarrow \boldsymbol{x}_j) \coloneqq \exp\left(-\int_{\overline{\boldsymbol{x}_{j-1}\boldsymbol{x}_j}} \sigma_t(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right), \tag{9}$$

with $\overline{\boldsymbol{x}_{j-1}\,\boldsymbol{x}_j}$ denoting the line segment between $\boldsymbol{x}_{j-1}$ and $\boldsymbol{x}_j$.

## 3.2 Differential Volumetric Light Transport

We focus on differentiating the path integral in Equation 1 with respect to parameters $\boldsymbol{\theta} \in \mathbb{R}^m$ that control[2] the spatially varying extinction coefficient $\sigma_t$. In this setting, both $\Omega$ and $\mu$ are independent of $\boldsymbol{\theta}$, so the derivative can be taken inside the path integral:

$$\partial_{\boldsymbol{\theta}} I \coloneqq \frac{\partial I}{\partial \boldsymbol{\theta}} = \int_{\Omega} \partial_{\boldsymbol{\theta}} f(\bar{\mathbf{x}})\,\mathrm{d}\bar{\mathbf{x}}, \tag{10}$$

where $f$ is defined in Equation 3. Applying the product rule to $f$ and grouping terms by whether the derivative falls on a vertex factor $v_j$ or a segment factor $s_j$ yields

$$\partial_{\boldsymbol{\theta}} f(\bar{\mathbf{x}}) = \underbrace{\sum_{j=1}^{K-1} \partial_{\boldsymbol{\theta}} v_j(\bar{\mathbf{x}})\, s_j(\bar{\mathbf{x}})\, f_{\setminus j}(\bar{\mathbf{x}})}_{\text{scattering}} + \underbrace{\sum_{j=1}^{K} v_j(\bar{\mathbf{x}})\, \partial_{\boldsymbol{\theta}} s_j(\bar{\mathbf{x}})\, f_{\setminus j}(\bar{\mathbf{x}})}_{\text{transmittance}}, \tag{11}$$

where

$$f_{\setminus j}(\bar{\mathbf{x}}) \coloneqq \left(\prod_{k \in \{0,1,\ldots,K\}\setminus\{j\}} v_k(\bar{\mathbf{x}})\right)\left(\prod_{k \in \{1,2,\ldots,K\}\setminus\{j\}} s_k(\bar{\mathbf{x}})\right) \tag{12}$$

captures the measurement contributions from the remaining vertices and segments.

The *scattering component* of Equation 11 collects derivatives of $\sigma_t$ evaluated at the interior vertices (via $v_j$), while the *transmittance component* collects derivatives of $\sigma_t$ accumulated along each segment (via $s_j$). The scattering sum runs only from $j = 1$ to $K - 1$ because $L_e$ and $W_e$ are assumed to be independent of $\boldsymbol{\theta}$, so $\partial_{\boldsymbol{\theta}} v_0(\bar{\mathbf{x}}) = \partial_{\boldsymbol{\theta}} v_K(\bar{\mathbf{x}}) = \mathbf{0}$.

## 3.3 Monte Carlo Methods for Volumetric Light Transport

Monte Carlo estimation of Equation 1 and Equation 10 requires drawing light-transport paths $\bar{\mathbf{x}}$ stochastically from the path space $\Omega$. A widely adopted technique is volumetric path tracing (VPT), which builds a path starting from the sensor vertex $\boldsymbol{x}_0$ and appends the remaining vertices $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ one at a time.

Given the previous vertex $\boldsymbol{x}_{j-1}$, VPT obtains the next vertex $\boldsymbol{x}_j$ by first sampling a direction $\boldsymbol{\omega}_j \in \mathcal{S}^2$ (typically based on the phase function $\rho$ at $\boldsymbol{x}_{j-1}$) and then a **free-flight distance** $t_j \in \mathbb{R}_{>0}$, and setting $\boldsymbol{x}_j = \boldsymbol{x}_{j-1} + t_j\,\boldsymbol{\omega}_j$.

To estimate Equation 1, the free-flight distance $t_j$ is ideally drawn from an exponential distribution with probability density function (PDF)

$$p(t_j) = \sigma_t(\boldsymbol{x}_j)\, T(\boldsymbol{x}_{j-1} \leftrightarrow \boldsymbol{x}_j). \tag{13}$$

[2]We make no assumptions on the exact parameterization of $\boldsymbol{\theta}$. Concrete examples include per-voxel values of a 3D grid and the weights of a multi-layer perceptron (MLP).

Table 1. Summary of the main symbols used in this paper.

| Symbol | Definition | Ref. |
|---|---|---|
| | *Volume properties* | |
| $\mathcal{V}$ | Medium volume | §3.1 |
| $\sigma_t$ | Extinction coefficient | §3.1 |
| $\hat{\rho}$ | Scaled phase function ($\alpha\,\rho$) | (5) |
| $T$ | Transmittance | (9) |
| | *Path spaces* | |
| $\Omega$ | Path space | (1) |
| $\Omega_K$ | Subspace of paths with exactly $K$ segments | (14) |
| $\Omega^{\mathrm{pre}}$ | Prefix path space (sensor-rooted) | (16) |
| $\Omega^{\mathrm{suf}}$ | Suffix path space (light-rooted) | (17) |
| | *Path contributions* | |
| $f$ | Measurement contribution | (3) |
| $v_j$ | Vertex factor at $\boldsymbol{x}_j$ | (4) |
| $s_j$ | Segment factor for $\overline{\boldsymbol{x}_{j-1}\,\boldsymbol{x}_j}$ | (6) |
| $f_{\setminus j}$ | $f$ with the $j$-th vertex/segment factors removed | (12) |
| $f_j^{\mathrm{ins}}$ | Alternate measurement contribution (internal) | (24) |
| $f^{\mathrm{out}}$ | Alternate measurement contribution (light) | (29) |
| $I_{\mathrm{ins}}$ | In-scattered light from $\boldsymbol{x}_K$ toward $\boldsymbol{x}_{K-1}$ | (31) |
| $\mathcal{E}_j$ | Per-segment joint integrand | (25) |
| $\bar{\mathbf{y}}$ | Path with $j$-th vertex replaced by probe $\boldsymbol{y}$ | §4.1 |

In homogeneous media, this is done via the inversion method. In heterogeneous media, delta tracking [Butcher and Messel 1958; Woodcock et al. 1965] is typically used to obtain unbiased distance samples.

To estimate Equation 10, many prior methods reuse the primal PDF in Equation 13. This has two problems. First, the scattering gradient is biased: wherever $\sigma_t = 0$ the sampling PDF vanishes but $\partial_{\boldsymbol{\theta}} v_j$ in Equation 11 can stay nonzero [Zhang et al. 2019], so empty regions are missed. Second, the scattering term evaluates $\partial_{\boldsymbol{\theta}} \sigma_t$ at the vertices while the transmittance term evaluates it at points between adjacent vertices. Both probe the same underlying extinction field along the same ray, yet do so at uncorrelated points, leaving any statistical coupling between them unexploited.

Nimier-David et al. [2022] eliminate the bias of the scattering gradient by drawing the scattering vertex from a PDF proportional to the transmittance $T(\boldsymbol{x}_{j-1} \leftrightarrow \boldsymbol{x}_j)$ via ratio tracking, an approach known as differential ratio tracking (DRT). The transmittance component, however, is still estimated independently. Our method is the first to close this gap, jointly estimating both gradient components at shared sample positions (Figure 2).

# 4 Our Formulation

*Note to the reader.* The same reformulation, presented in classical RTE language and along a single ray segment, is given in the supplementary material; readers more comfortable with RTE intuition than with path-space measure theory may prefer that route.

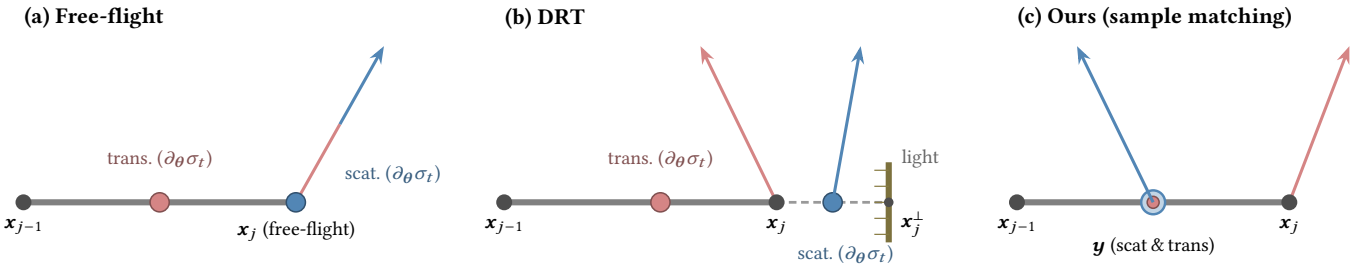


Fig. 2. **Differentiation strategies on a host segment $\overline{\boldsymbol{x}_{j-1}\,\boldsymbol{x}_j}$.** Colored dots mark where $\partial_{\boldsymbol{\theta}}\sigma_t$ is evaluated — blue for the scattering term, rose for the transmittance term — and the colored arrows are the in-scattering sub-paths whose radiance multiplies $\partial_{\boldsymbol{\theta}}\sigma_t$. **(a) Free-flight**: a single primal path is drawn with a two-color arrow that feeds the transmittance derivative and the scattering derivative simultaneously, but $\partial_{\boldsymbol{\theta}}\sigma_t$ is evaluated at *different* positions ($\boldsymbol{x}_j$ versus a sample on the segment). **(b) DRT [Nimier-David et al. 2022]**: the two terms use *independent* sub-paths drawn from *different* positions; the scattering sample is drawn from $\overline{\boldsymbol{x}_{j-1}\,\boldsymbol{x}_j^{\perp}}$. **(c) Ours**: a single matched probe $\boldsymbol{y}$ on the segment carries both terms, so $\partial_{\boldsymbol{\theta}}\sigma_t$ is evaluated at the *same* position for both.

We now reformulate the differential path integral in Equation 10 so that the derivatives arising from the scattering and transmittance components become aligned.

To begin, we decompose the path space $\Omega$ into disjoint subspaces of paths with fixed lengths via $\Omega := \cup_{K\geq 1}\Omega_K$, where $\Omega_K$ denotes the space of paths with exactly $K$ segments. This allows us to rewrite Equation 10 as an infinite sum of integrals:

$$\partial_{\boldsymbol{\theta}} I = \sum_{K=1}^{\infty}\sum_{j=1}^{K}(\partial_{\boldsymbol{\theta}} I)_{K,j}, \tag{14}$$

where

$$(\partial_{\boldsymbol{\theta}} I)_{K,j} := \int_{\Omega_K}\left(\partial_{\boldsymbol{\theta}} v_j(\bar{\mathbf{x}})\, s_j(\bar{\mathbf{x}}) + v_j(\bar{\mathbf{x}})\,\partial_{\boldsymbol{\theta}} s_j(\bar{\mathbf{x}})\right) f_{\setminus j}(\bar{\mathbf{x}})\,\mathrm{d}\bar{\mathbf{x}} \tag{15}$$

captures the contribution from the segment $\overline{\boldsymbol{x}_{j-1}\,\boldsymbol{x}_j}$ (via $\partial_{\boldsymbol{\theta}} v_j$ and $\partial_{\boldsymbol{\theta}} s_j$).

In what follows, we first reformulate the per-segment terms $(\partial_{\boldsymbol{\theta}} I)_{K,j}$ from Equation 15 in Section 4.1, then combine the results to obtain the final formulation (Equation 30) in Section 4.2.

*Prefix and suffix path spaces.* To facilitate the derivations, we define the **prefix path space** as

$$\Omega^{\mathrm{pre}} := \cup_{K=1}^{\infty}\Omega_K^{\mathrm{pre}}, \tag{16}$$

where $\Omega_K^{\mathrm{pre}}$ consists of "prefix" light paths $(\boldsymbol{x}_0^{\mathrm{pre}}, \boldsymbol{x}_1^{\mathrm{pre}}, \ldots, \boldsymbol{x}_K^{\mathrm{pre}})$ with $K$ segments (and $K+1$ vertices) such that $\boldsymbol{x}_0^{\mathrm{pre}}$ lies on the sensor and the remaining vertices correspond to volumetric scattering events.

Similarly, we define the **suffix path space**

$$\Omega^{\mathrm{suf}} := \cup_{K=0}^{\infty}\Omega_K^{\mathrm{suf}}, \tag{17}$$

where $\Omega_K^{\mathrm{suf}}$ consists of "suffix" light paths $(\boldsymbol{x}_0^{\mathrm{suf}}, \boldsymbol{x}_1^{\mathrm{suf}}, \ldots, \boldsymbol{x}_K^{\mathrm{suf}})$ such that $\boldsymbol{x}_K^{\mathrm{suf}}$ lies on a light source and the remaining vertices correspond to volumetric scattering events.

## 4.1 Per-Segment Gradients

We now reformulate the per-segment contributions $(\partial_{\boldsymbol{\theta}} I)_{K,j}$ from Equation 15. To this end, we consider two complementary cases: (i) when $\overline{\boldsymbol{x}_{j-1}\,\boldsymbol{x}_j}$ is an *internal* segment (i.e., $0 < j < K$); and (ii) when $\overline{\boldsymbol{x}_{j-1}\,\boldsymbol{x}_j}$ is a *light* segment (i.e., $j = K$).

*Internal segments.* When $0 < j < K$, we parameterize the volumetric vertex as $\boldsymbol{x}_j = \boldsymbol{x}_{j-1} + t\,\boldsymbol{\omega}$, where $\boldsymbol{\omega} \in \mathcal{S}^2$ is the scattering direction and $t \in (0, t_{\max})$ is the free-flight distance, with $t_{\max}$ being the minimum distance from $\boldsymbol{x}_{j-1}$ to a surface along $\boldsymbol{\omega}$. Since $\mathrm{d}V(\boldsymbol{x}_j) = t^2\,\mathrm{d}t\,\mathrm{d}\boldsymbol{\omega}$ and $s_j(\bar{\mathbf{x}}) = T(\boldsymbol{x}_{j-1}\leftrightarrow\boldsymbol{x}_j)/t^2$, Equation 15 can be rewritten as

$$(\partial_{\boldsymbol{\theta}} I)_{K,j} = \int_{\Omega_{K-1}}\int_{\mathcal{S}^2} g(\bar{\mathbf{x}}_{\setminus j}, \boldsymbol{\omega})\,\mathrm{d}\boldsymbol{\omega}\,\mathrm{d}\bar{\mathbf{x}}_{\setminus j}, \tag{18}$$

where $\bar{\mathbf{x}}_{\setminus j}$ denotes the path with the vertex $\boldsymbol{x}_j$ removed, and

$$g(\bar{\mathbf{x}}_{\setminus j}, \boldsymbol{\omega}) := \int_0^{t_{\max}}\left(\partial_{\boldsymbol{\theta}} v_j(t)\,T(t) + v_j(t)\,\partial_{\boldsymbol{\theta}} T(t)\right) f_{\setminus j}(\bar{\mathbf{x}})\,\mathrm{d}t \tag{19}$$

with $v_j(t) := v_j(\bar{\mathbf{x}})$ and $T(t) := T(\boldsymbol{x}_{j-1}\leftrightarrow\boldsymbol{x}_j)$.

Differentiating the transmittance $T$ defined in Equation 9 gives

$$\partial_{\boldsymbol{\theta}} T(t) = -\left(\int_0^t \partial_{\boldsymbol{\theta}}\sigma_t(s)\,\mathrm{d}s\right) T(t). \tag{20}$$

It follows that Equation 19 decomposes into a scattering and a transmittance part:

$$g(\bar{\mathbf{x}}_{\setminus j}, \boldsymbol{\omega}) = \underbrace{\int_0^{t_{\max}} \partial_{\boldsymbol{\theta}} v_j(t)\,T(t)\,f_{\setminus j}(\bar{\mathbf{x}})\,\mathrm{d}t}_{\text{scattering}} - \underbrace{\int_0^{t_{\max}}\int_0^{t} v_j(t)\,\partial_{\boldsymbol{\theta}}\sigma_t(s)\,T(t)\,f_{\setminus j}(\bar{\mathbf{x}})\,\mathrm{d}s\,\mathrm{d}t}_{\text{transmittance}}. \tag{21}$$

The two components on the right-hand side of Equation 21 evaluate the derivative $\partial_{\boldsymbol{\theta}}\sigma_t$ at different positions: the scattering term at the segment endpoint $\boldsymbol{x}_j$ (through $\partial_{\boldsymbol{\theta}} v_j(t)$), and the transmittance term at interior points with $s < t$ (through $\partial_{\boldsymbol{\theta}}\sigma_t(s)$). This misalignment is present in all previous techniques [Nimier-David et al. 2022; Zhang et al. 2019].

We overcome this challenge by reformulating Equation 21 using the definition of transmittance and Fubini's theorem as follows:

$$g(\bar{\mathbf{x}}_{\setminus j}, \boldsymbol{\omega}) = \int_0^{t_{\max}} T(t) \int_0^t \left( \partial_{\boldsymbol{\theta}} v_j(s)\, \sigma_t(t)\, f_{\setminus j}(\bar{\mathbf{y}}) - v_j(t)\, \partial_{\boldsymbol{\theta}} \sigma_t(s)\, f_{\setminus j}(\bar{\mathbf{x}}) \right) \mathrm{d}s\, \mathrm{d}t + \int_0^{t_{\max}} T(t_{\max})\, \partial_{\boldsymbol{\theta}} v_j(t)\, f_{\setminus j}(\bar{\mathbf{x}})\, \mathrm{d}t. \quad (22)$$

Please see Section A.1 for a detailed derivation.

Substituting Equation 22 into Equation 18 and changing the integration variables $(\boldsymbol{\omega}, t)$ back to $\boldsymbol{x}_j$ yields

$$\boxed{(\partial_{\boldsymbol{\theta}} I)_{K,j} = \int_{\Omega_K} s_j(\bar{\mathbf{x}})\, \mathcal{E}_j(\bar{\mathbf{x}})\, \mathrm{d}\bar{\mathbf{x}} + \int_{\Omega_K} \partial_{\boldsymbol{\theta}} v_j(\bar{\mathbf{x}})\, f_j^{\mathrm{ins}}(\bar{\mathbf{x}})\, \mathrm{d}\bar{\mathbf{x}},} \quad (23)$$

where

$$f_j^{\mathrm{ins}}(\bar{\mathbf{x}}) \coloneqq T(\boldsymbol{x}_j \leftrightarrow \boldsymbol{x}_j^{\perp})\, s_j(\bar{\mathbf{x}})\, f_{\setminus j}(\bar{\mathbf{x}}) \quad (24)$$

with $\boldsymbol{x}_j^{\perp}$ denoting the intersection between the medium's boundary and the ray from $\boldsymbol{x}_{j-1}$ through $\boldsymbol{x}_j$.

Additionally,

$$\begin{aligned} \mathcal{E}_j(\bar{\mathbf{x}}) &\coloneqq \int_{\overline{\boldsymbol{x}_{j-1}\boldsymbol{x}_j}} \left( \partial_{\boldsymbol{\theta}} v_j(\bar{\mathbf{y}})\, \sigma_t(\boldsymbol{x}_j)\, f_{\setminus j}(\bar{\mathbf{y}}) - v_j(\bar{\mathbf{x}})\, \partial_{\boldsymbol{\theta}} \sigma_t(\boldsymbol{y})\, f_{\setminus j}(\bar{\mathbf{x}}) \right) \mathrm{d}\boldsymbol{y} \\ &= \int_{\overline{\boldsymbol{x}_{j-1}\boldsymbol{x}_j}} \Big( \hat{\rho}(\boldsymbol{x}_{j+1} \to \boldsymbol{y} \to \boldsymbol{x}_{j-1})\, f_{\setminus j}(\bar{\mathbf{y}}) \\ &\quad - \hat{\rho}(\boldsymbol{x}_{j+1} \to \boldsymbol{x}_j \to \boldsymbol{x}_{j-1})\, f_{\setminus j}(\bar{\mathbf{x}}) \Big)\, \sigma_t(\boldsymbol{x}_j)\, \partial_{\boldsymbol{\theta}} \sigma_t(\boldsymbol{y})\, \mathrm{d}\boldsymbol{y}, \end{aligned} \quad (25)$$

is an integral over the line segment $\overline{\boldsymbol{x}_{j-1}\boldsymbol{x}_j}$. In this equation, the light path $\bar{\mathbf{y}}$ is obtained by replacing the $j$-th vertex $\boldsymbol{x}_j$ of $\bar{\mathbf{x}}$ with $\boldsymbol{y}$, and $\hat{\rho}$ denotes the scaled phase function defined in Equation 5.

The second equality in Equation 25 holds because

$$\partial_{\boldsymbol{\theta}} v_j(\bar{\mathbf{y}}) = \hat{\rho}(\boldsymbol{x}_{j+1} \to \boldsymbol{y} \to \boldsymbol{x}_{j-1})\, \partial_{\boldsymbol{\theta}} \sigma_t(\boldsymbol{y}), \quad (26)$$

which follows from the assumption that $\boldsymbol{\theta}$ affects only the extinction coefficient $\sigma_t$.

The first term on the right-hand side of Equation 23, as illustrated in Figure 3, integrates the contribution difference between an ordinary light path $\bar{\mathbf{x}}$ and an altered version $\bar{\mathbf{y}}$ that differs from $\bar{\mathbf{x}}$ in only one vertex. Crucially, this term evaluates $\partial_{\boldsymbol{\theta}} \sigma_t$ at the same point $\boldsymbol{y}$ for both the scattering and transmittance contributions (via Equation 25), achieving the desired alignment.

The second integral in Equation 23 will later be combined with the reformulated $(\partial_{\boldsymbol{\theta}} I)_{K,K}$—presented next—to produce the final derivative $\partial_{\boldsymbol{\theta}} I$ in Section 4.2.

*Light segments.* We have derived the reformulated $(\partial_{\boldsymbol{\theta}} I)_{K,j}$ in Equation 23 for all $0 < j < K$. We now focus on the case $j = K$:

$$(\partial_{\boldsymbol{\theta}} I)_{K,K} = \int_{\Omega_K} v_K(\bar{\mathbf{x}})\, \partial_{\boldsymbol{\theta}} s_K(\bar{\mathbf{x}})\, f_{\setminus K}(\bar{\mathbf{x}})\, \mathrm{d}\bar{\mathbf{x}}, \quad (27)$$

where the $\partial_{\boldsymbol{\theta}} v_K(\bar{\mathbf{x}})\, s_K(\bar{\mathbf{x}})$ term vanishes because $\partial_{\boldsymbol{\theta}} v_K(\bar{\mathbf{x}}) = \mathbf{0}$.

As illustrated in Figure 4, by reparameterizing the light path as a prefix path whose endpoint $\boldsymbol{x}_K$ lies along the light segment,

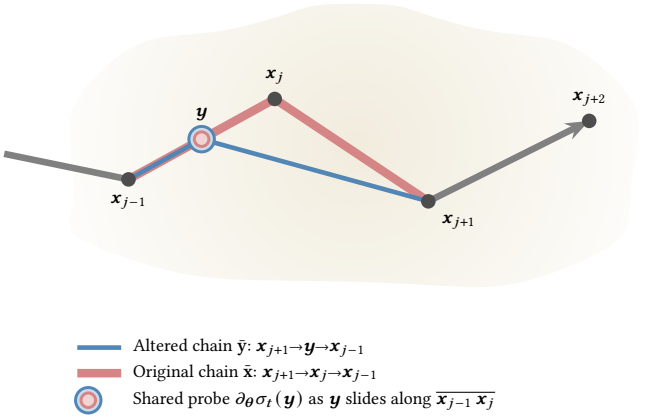


Fig. 3. **Geometric interpretation of the first term of Equation 23.** A four-vertex window of the host light path $\bar{\mathbf{x}}$ is shown traversing the volumetric medium. $\mathcal{E}_j$ (Equation 25) integrates the difference between the rose chain (host, through $\boldsymbol{x}_j$) and the blue chain (altered $\bar{\mathbf{y}}$, through a shared gradient probe $\boldsymbol{y}$).

Equation 27 can be rewritten as a path integral over the prefix path space $\Omega_K^{\mathrm{pre}}$ (defined in Equation 16):

$$\boxed{(\partial_{\boldsymbol{\theta}} I)_{K,K} = -\int_{\Omega_K^{\mathrm{pre}}} \partial_{\boldsymbol{\theta}} \sigma_t(\boldsymbol{x}_K)\, f^{\mathrm{out}}(\bar{\mathbf{x}})\, \mathrm{d}\bar{\mathbf{x}},} \quad (28)$$

where

$$f^{\mathrm{out}}(\bar{\mathbf{x}}) \coloneqq L_{\mathrm{e}}(\boldsymbol{x}_K^{\perp} \to \boldsymbol{x}_K)\, T(\boldsymbol{x}_K \leftrightarrow \boldsymbol{x}_K^{\perp})\, s_K(\bar{\mathbf{x}})\, f_{\setminus K}(\bar{\mathbf{x}}), \quad (29)$$

is the measurement contribution $f(\bar{\mathbf{x}})$ with the last vertex term $v_K(\bar{\mathbf{x}})$ replaced by $L_{\mathrm{e}}(\boldsymbol{x}_K^{\perp} \to \boldsymbol{x}_K)\, T(\boldsymbol{x}_K \leftrightarrow \boldsymbol{x}_K^{\perp})$.

Please refer to Section A.2 for a rigorous derivation.

## 4.2 Full Gradient

Substituting the reformulated contributions—Equation 23 for $0 < j < K$ and Equation 28 for $j = K$—into Equation 14 yields our new formulation of $\partial_{\boldsymbol{\theta}} I$ (see Section A.3 for the full derivation):

$$\boxed{\partial_{\boldsymbol{\theta}} I = \underbrace{\int_{\Omega^{\mathrm{pre}}} s_K(\bar{\mathbf{x}}) \int_{\Omega^{\mathrm{suf}}} \mathcal{E}_K(\bar{\mathbf{x}} + \bar{\mathbf{x}}^{\mathrm{suf}})\, \mathrm{d}\bar{\mathbf{x}}^{\mathrm{suf}}\, \mathrm{d}\bar{\mathbf{x}}}_{\text{internal}} + \underbrace{\int_{\Omega^{\mathrm{pre}}} \partial_{\boldsymbol{\theta}} \sigma_t(\boldsymbol{x}_K) \left( I_{\mathrm{ins}}(\bar{\mathbf{x}}) - f^{\mathrm{out}}(\bar{\mathbf{x}}) \right) \mathrm{d}\bar{\mathbf{x}}}_{\text{light}},} \quad (30)$$

where $\mathcal{E}_K$ is defined in Equation 25, and, for any prefix path $\bar{\mathbf{x}} = (\boldsymbol{x}_0, \ldots, \boldsymbol{x}_K)$,

$$I_{\mathrm{ins}}(\bar{\mathbf{x}}) \coloneqq \int_{\Omega^{\mathrm{suf}}} \hat{\rho}(\boldsymbol{x}_0^{\mathrm{suf}} \to \boldsymbol{x}_K \to \boldsymbol{x}_{K-1})\, f_K^{\mathrm{ins}}(\bar{\mathbf{x}} + \bar{\mathbf{x}}^{\mathrm{suf}})\, \mathrm{d}\bar{\mathbf{x}}^{\mathrm{suf}} \quad (31)$$

captures the in-scattered light from $\boldsymbol{x}_K$ toward $\boldsymbol{x}_{K-1}$.

In Equation 30 and Equation 31, $\Omega^{\mathrm{pre}}$ and $\Omega^{\mathrm{suf}}$ are the prefix and suffix path spaces defined in Equation 16 and Equation 17, respectively, and $\bar{\mathbf{x}} + \bar{\mathbf{x}}^{\mathrm{suf}}$ denotes the path obtained by concatenating the prefix $\bar{\mathbf{x}}$ with the suffix $\bar{\mathbf{x}}^{\mathrm{suf}}$.

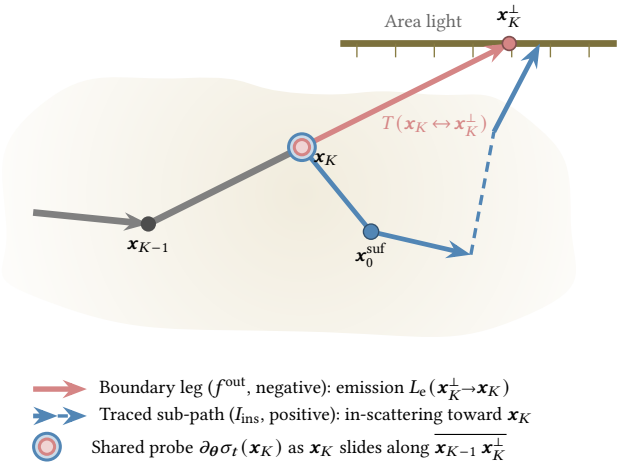


Fig. 4. **Geometric layout of the light component of Equation 30.** The light path is reparameterized so that its endpoint $\boldsymbol{x}_K$ slides along the segment $\overline{\boldsymbol{x}_{K-1}\, \boldsymbol{x}_K^\perp}$, with $\boldsymbol{x}_K^\perp$ residing on a light source surface (here drawn as an area light, illustrating that the light is in general independent of the medium boundary). At $\boldsymbol{x}_K$, the integrand is the difference between two contributions sharing the same $\partial_{\boldsymbol{\theta}} \sigma_t$ probe: the in-scattering $I_{\text{ins}}$ (blue, positive) is sampled by tracing a sub-path from $\boldsymbol{x}_K$, while the boundary emission $f^{\text{out}}$ (rose, negative) is delivered along the residual leg from $\boldsymbol{x}_K$ to $\boldsymbol{x}_K^\perp$ together with the transmittance $T(\boldsymbol{x}_K \leftrightarrow \boldsymbol{x}_K^\perp)$.

*Discussion.* Conceptually, our new formulation in Equation 30 captures two types of contribution differences. The *internal* component, as illustrated in Figure 3 (with $j = K$), integrates the difference between the in-scatterings at $\boldsymbol{y}$ and $\boldsymbol{x}_K$ toward $\boldsymbol{x}_{K-1}$. The *light* component, in turn, integrates the difference between the in-scattering at $\boldsymbol{x}_K$ and the emission at $\boldsymbol{x}_K^\perp$, both as observed from $\boldsymbol{x}_{K-1}$.

## 5 Our Estimator

With our new formulation in Equation 14 established, we now introduce Monte Carlo estimators for the derivative $\partial_{\boldsymbol{\theta}} I$. We first discuss estimation of the internal component in Section 5.1, followed by the light component in Section 5.2. Lastly, we provide implementation details in Section 5.3.

### 5.1 Internal Component

To estimate the *internal* component of Equation 30, our method begins by drawing a prefix light path $\bar{\mathbf{x}} = (\boldsymbol{x}_0, \ldots, \boldsymbol{x}_K)$ using unidirectional volumetric path tracing (VPT) with standard delta tracking [Woodcock et al. 1965], together with one extra *probe* point $\boldsymbol{y}$ sampled uniformly over the segment $\overline{\boldsymbol{x}_{K-1}\, \boldsymbol{x}_K}$ of $\bar{\mathbf{x}}$.

With the prefix path $\bar{\mathbf{x}}$ and probe $\boldsymbol{y}$ drawn, we propose two approaches to estimate the inner integral over the suffix path space $\boldsymbol{\Omega}^{\text{suf}}$.

*Single suffix path.* Our first approach samples a single suffix path $\bar{\mathbf{x}}^{\text{suf}}$ by continuing the already-sampled path $\bar{\mathbf{x}}$ via VPT. This produces a full path $\bar{\mathbf{x}} + \bar{\mathbf{x}}^{\text{suf}}$ drawn with a PDF $p^{\text{int}}_{\text{single}}$ that is approximately proportional to the measurement contribution $f(\bar{\mathbf{x}} + \bar{\mathbf{x}}^{\text{suf}})$, leading to the segment estimator:

$$\Big(\sigma_t(\boldsymbol{x}_K)\, \hat{\rho}(\boldsymbol{x}_0^{\text{suf}} \to \boldsymbol{y} \to \boldsymbol{x}_{K-1})\, s_K(\bar{\mathbf{x}})\, f_{\backslash K}(\bar{\mathbf{y}} + \bar{\mathbf{x}}^{\text{suf}}) - f(\bar{\mathbf{x}} + \bar{\mathbf{x}}^{\text{suf}})\Big) \frac{\|\boldsymbol{x}_K - \boldsymbol{x}_{K-1}\|}{p^{\text{int}}_{\text{single}}}\, \partial_{\boldsymbol{\theta}} \sigma_t(\boldsymbol{y}), \quad (32)$$

where $\boldsymbol{x}_0^{\text{suf}}$ denotes the first vertex of the suffix path $\bar{\mathbf{x}}^{\text{suf}}$, and the $\|\boldsymbol{x}_K - \boldsymbol{x}_{K-1}\|$ factor in the numerator is the inverse PDF of the uniform sampling of $\boldsymbol{y}$ on the segment.

Most numerator/denominator factors in Equation 32 cancel. The exception is the next-segment factor: $\bar{\mathbf{y}}$ has $\boldsymbol{y}$ in place of $\boldsymbol{x}_K$, so $\overline{\boldsymbol{y}\, \boldsymbol{x}_0^{\text{suf}}}$ replaces $\overline{\boldsymbol{x}_K\, \boldsymbol{x}_0^{\text{suf}}}$. This requires evaluating the ratio between segment factors:

$$\frac{s_{K+1}(\bar{\mathbf{y}} + \bar{\mathbf{x}}^{\text{suf}})}{s_{K+1}(\bar{\mathbf{x}} + \bar{\mathbf{x}}^{\text{suf}})} = \frac{G(\boldsymbol{y} \leftrightarrow \boldsymbol{x}_0^{\text{suf}})}{G(\boldsymbol{x}_K \leftrightarrow \boldsymbol{x}_0^{\text{suf}})} \frac{T(\boldsymbol{y} \leftrightarrow \boldsymbol{x}_0^{\text{suf}})}{T(\boldsymbol{x}_K \leftrightarrow \boldsymbol{x}_0^{\text{suf}})}. \quad (33)$$

While the first ratio of geometric terms on the right-hand side is easy to evaluate, the second ratio of transmittances requires special treatment: dividing two independently estimated transmittance values introduces bias. Instead, we collapse this ratio into a single exponential by reparameterizing both segments with a shared parameter $u \in [0, 1]$:

$$\frac{T(\boldsymbol{y} \leftrightarrow \boldsymbol{x}_0^{\text{suf}})}{T(\boldsymbol{x}_K \leftrightarrow \boldsymbol{x}_0^{\text{suf}})} = \exp\left( - \int_0^1 \Big[ \|\boldsymbol{x}_0^{\text{suf}} - \boldsymbol{y}\|\, \sigma_t(\boldsymbol{y} + u(\boldsymbol{x}_0^{\text{suf}} - \boldsymbol{y})) - \|\boldsymbol{x}_0^{\text{suf}} - \boldsymbol{x}_K\|\, \sigma_t(\boldsymbol{x}_K + u(\boldsymbol{x}_0^{\text{suf}} - \boldsymbol{x}_K)) \Big] \mathrm{d}u \right), \quad (34)$$

which is the exponential of a *single* integral over $u$ rather than a quotient of two transmittances. A null-collision estimator applied to Equation 34 with collision positions $\{u_k\}$ *shared* between the two media queries is unbiased, and its integrand vanishes pointwise as $\boldsymbol{y} \to \boldsymbol{x}_K$, so the estimator's variance scales as $\|\boldsymbol{y} - \boldsymbol{x}_K\|^2$.

*Dual suffix paths.* Alternatively, we propose a second approach in which two independent suffix paths $\bar{\mathbf{x}}^{\text{suf1}}$ and $\bar{\mathbf{x}}^{\text{suf2}}$ are sampled starting from $\boldsymbol{y}$ and $\boldsymbol{x}_K$, respectively. In this way, all transmittance terms cancel cleanly, avoiding the need to estimate the ratio in Equation 33. And this yields a different estimator:

$$\left( \frac{\sigma_t(\boldsymbol{x}_K)\, \hat{\rho}(\boldsymbol{x}_0^{\text{suf}} \to \boldsymbol{y} \to \boldsymbol{x}_{K-1})\, s_K(\bar{\mathbf{x}})\, f_{\backslash K}(\bar{\mathbf{y}} + \bar{\mathbf{x}}^{\text{suf1}})}{p^{\text{int}}_{\text{dual1}}} - \frac{f(\bar{\mathbf{x}} + \bar{\mathbf{x}}^{\text{suf2}})}{p^{\text{int}}_{\text{dual2}}} \right) \|\boldsymbol{x}_K - \boldsymbol{x}_{K-1}\|\, \partial_{\boldsymbol{\theta}} \sigma_t(\boldsymbol{y}), \quad (35)$$

where $p^{\text{int}}_{\text{dual1}}$ and $p^{\text{int}}_{\text{dual2}}$ denote, respectively, the PDFs for drawing the full paths $\bar{\mathbf{y}} + \bar{\mathbf{x}}^{\text{suf1}}$ and $\bar{\mathbf{x}} + \bar{\mathbf{x}}^{\text{suf2}}$.

### 5.2 Light Component

To estimate the second integral in Equation 30, our method again begins by sampling a prefix path $\bar{\mathbf{x}}$. However, unlike the internal component (Section 5.1), additional care is needed when drawing this path because the last vertex $\boldsymbol{x}_K$ now plays a role analogous to the probe $\boldsymbol{y}$ in the internal component. Specifically, we obtain $\bar{\mathbf{x}}$ by

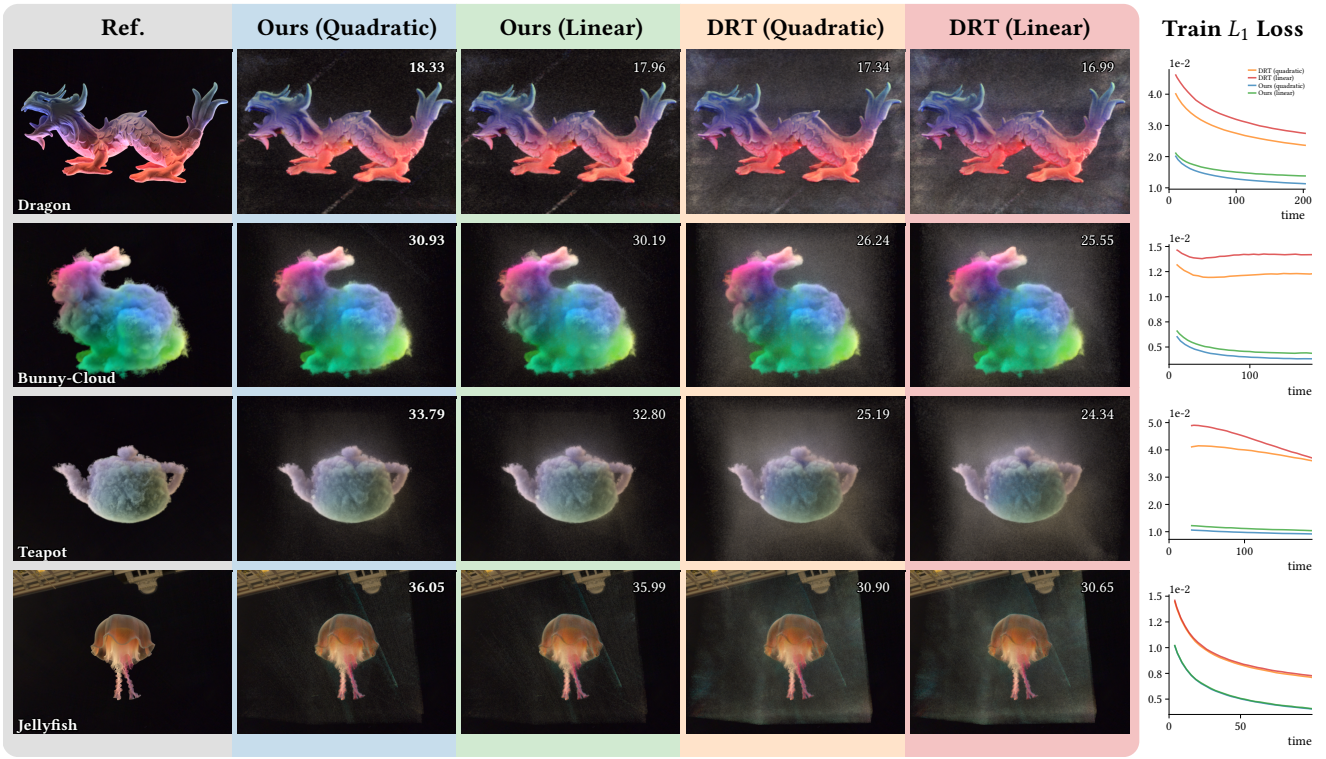


Fig. 5. **Reconstruction Results (part 1/2).** Comparison of DRT and our method for the linear and quadratic complexity variants. Numbers in the top-right corner of each image indicate PSNR values (dB) for reference; visual differences, especially in volumetric details, are more indicative of quality. The rightmost column plots training $L_1$ loss as a function of optimization time in minutes; the "1e−$N$" above each panel's y-axis is the shared order-of-magnitude factor for the tick values. Curves: DRT (quadratic), DRT (linear), Ours (quadratic), Ours (linear). Both of our variants reach a lower training $L_1$ than the corresponding DRT variant at every point during optimization. See Fig. 6 for the remaining four scenes.

drawing a full path $(\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{K-1}, \boldsymbol{x}_K^{\perp})$ using standard VPT (with $\boldsymbol{x}_K^{\perp}$ located on a light source) and then sampling $\boldsymbol{x}_K$ uniformly over the last segment $\overline{\boldsymbol{x}_{K-1}\boldsymbol{x}_K^{\perp}}$.

With the prefix path $\bar{\mathbf{x}}$ sampled, we construct a suffix path $\bar{\mathbf{x}}^{\mathrm{suf}}$ from $\boldsymbol{x}_K$ using standard VPT. This yields the segment estimator:

$$\left(\frac{\hat{p}(\boldsymbol{x}_0^{\mathrm{suf}} \to \boldsymbol{x}_K \to \boldsymbol{x}_{K-1})\, f_K^{\mathrm{ins}}(\bar{\mathbf{x}} + \bar{\mathbf{x}}^{\mathrm{suf}})}{p_{\mathrm{pre}}^{\mathrm{light}}\, p_{\mathrm{suf}}^{\mathrm{light}}} - \frac{f^{\mathrm{out}}(\bar{\mathbf{x}})}{p_{\mathrm{pre}}^{\mathrm{light}}}\right)\sigma_t(\boldsymbol{x}_K), \quad (36)$$

where $p_{\mathrm{pre}}^{\mathrm{light}}$ and $p_{\mathrm{suf}}^{\mathrm{light}}$ denote the PDFs for drawing the prefix path $\bar{\mathbf{x}}$ and suffix path $\bar{\mathbf{x}}^{\mathrm{suf}}$, respectively.

The estimators in Section 5.1 and Section 5.2 are unbiased; we provide a brief argument in the supplementary material.

## 5.3 Implementation

We now provide additional details on implementing the estimators introduced in Section 5.1 and Section 5.2.

*Correlated prefix paths.* The internal and light prefixes share $\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{K-1}$ via one VPT pass; only the last vertex differs ($\boldsymbol{x}_K^{\mathrm{int}}$ by free-flight, $\boldsymbol{x}_K^{\mathrm{light}}$ by the boundary procedure of Section 5.2).

*Single vs. dual suffix paths.* Given the sampled prefix path and probe point $\boldsymbol{y}$, we still need to construct suffix paths to estimate the internal component. In Section 5.1, we proposed two variants of this process. Although the single-suffix-path variant is appealing, we find it nontrivial to implement efficiently on the GPU. First, the reconnection from $\boldsymbol{y}$ to $\boldsymbol{x}_0^{\mathrm{suf}}$ requires an additional ray traversal along a different geometric chord for every probe, with branch-divergent control flow whenever probes terminate at different occluders. Second, the transmittance ratio in Equation 34 requires its own null-collision estimator with positions coupled across two slightly different media queries, introducing a second tightly synchronized tracker inside the inner loop.

We therefore leave efficient GPU implementation of the single-suffix-path variant to future work and adopt the dual-suffix-path variant. For better performance, we further employ path-replay backpropagation [Vicini et al. 2021] to evaluate Equation 35. In practice, the resulting estimator can be obtained by modifying only **60–70 lines of code** on top of an existing DRT [Nimier-David et al. 2022] implementation, making the technique easy to adopt in differentiable volume renderers like Mitsuba 3 that already support DRT.


Fig. 6. **Reconstruction Results (part 2/2).** Continuation of Fig. 5 on the remaining four scenes. Column conventions, PSNR overlays, and the training $L_1$ loss plots follow the same format.

*Linear and quadratic variants.* Similar to Nimier-David et al. [2022], our estimator admits two per-sample complexity variants. The *quadratic* variant places a probe and evaluates the suffix path contribution at each segment, accumulating $O(K^2)$ work per primal path. The *linear* variant selects a single segment by weighted reservoir sampling, evaluates only that segment's contribution, and reweights by the inverse selection probability for unbiasedness, yielding $O(K)$ work. Both variants are unbiased and differ only in how many segments contribute per primal path. We implement both and compare their performance in Section 6.

*Amortizing multiple probes.* Since our estimators evaluate derivatives only at the probe points (namely, $\boldsymbol{y}$ for the internal component and $\boldsymbol{x}_K$ for the light component), it is desirable to use $\Lambda > 1$ probes to amortize the cost of constructing the prefix paths. However, because we also need to construct a suffix path for each probe (to estimate in-scattering at those points), a large $\Lambda$ can quickly become inefficient.

To address this issue, we decompose the in-scattering at each probe into a *direct* and an *indirect* component. We then estimate the *direct* contribution at all $\Lambda$ probes, and the *indirect* contribution at only one of them (see Figure 8). This direct/indirect split is conceptually related to the decoupled ray-marching strategy of Kulla [2011], which similarly separates transmittance and in-scattering evaluation along a segment; here the decomposition controls the cost of multi-probe *gradient* estimation rather than forward radiance estimation.

In practice, we find that $\Lambda = 4$ yields strong performance, and it matches the setting used by Nimier-David et al. [2022], enabling a fair comparison.

## 6 Experiments

We demonstrate the effectiveness of our method by comparing it to DRT [Nimier-David et al. 2022], the state-of-the-art unbiased extinction-gradient estimator for the differential path integral. We implement our method and baseline in Mitsuba 3. All experiments are conducted on an NVIDIA RTX 4090 GPU.

### 6.1 Validations and Comparisons

*1D validation.* We first validate the variance reduction of our method in a synthetic 1D setting. As shown in Figure 9, we consider a piecewise-linear extinction field $\sigma_t(t)$ on $t \in [0, 1]$, and compare the gradient vectors produced by our method and DRT [Nimier-David et al. 2022]. Both methods follow the expected Monte Carlo convergence rate of $O(1/N)$, yet our estimator attains an MSE roughly 6× lower than DRT across all sample counts.

*Differentiable and inverse rendering setup.* We compare the performance of our method and baselines across 8 different scenes.

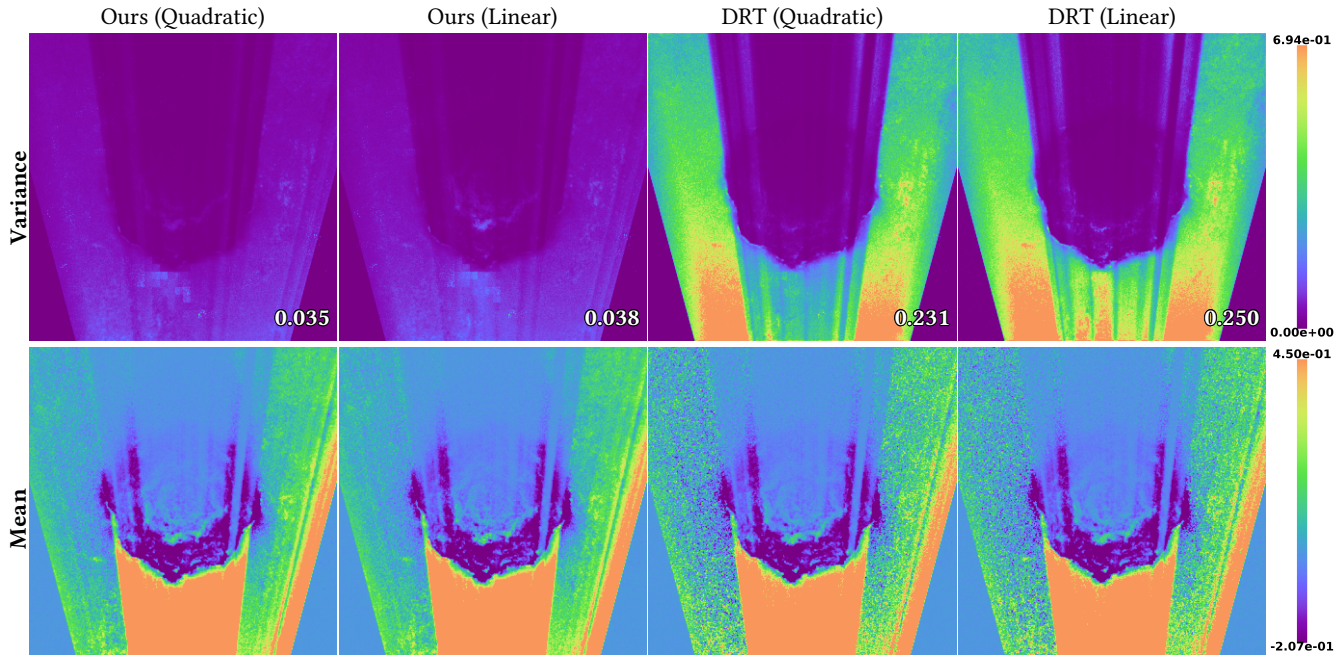


Fig. 7. **Variance and Mean Comparison for the Jellyfish Scene.** Top row visualizes per-pixel variance and bottom row shows mean values under equal-time rendering. The lower-right number reports the average value of the slice variance.

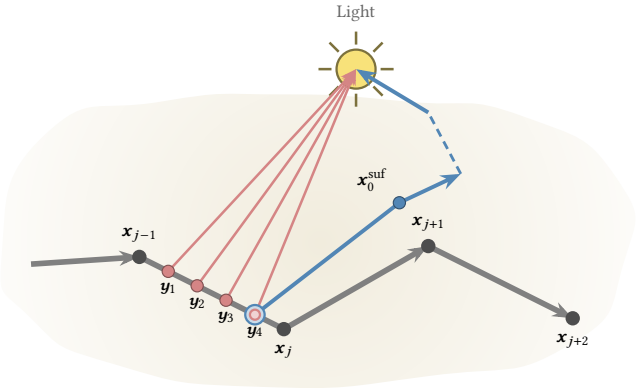


Fig. 8. **Per-segment direct/indirect split inside the dual-suffix-path estimator (Equation 35).** $\Lambda$ probes $\{y_i\}$ are drawn uniformly on each interior segment. Indirect in-scattering is traced once from a single chosen probe (blue, through $x_0^{\text{suf}}$); direct in-scattering is evaluated at each probe via a direct-lighting query (rose).

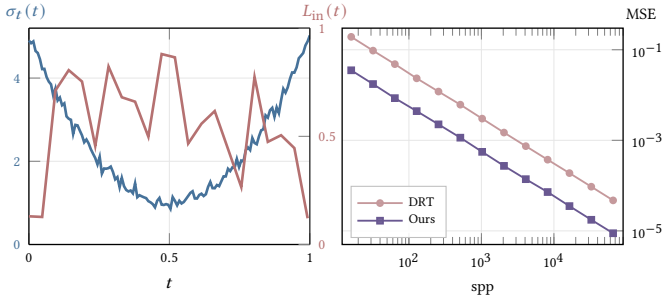


Fig. 9. **1D Differentiable Rendering Comparison.** (a) A piecewise-linear extinction coefficient $\sigma_t(t)$ (left axis) and in-scattering radiance $L_{\text{in}}(t)$ (right axis), each with 256 parameters. (b) MSE of the extinction gradient as a function of spp. Our method achieves 6× lower variance through joint estimation of the extinction gradient.

All scenes contain an environment map and a high-resolution heterogeneous volume grid. We use Adam for optimization in inverse rendering. We use $\Lambda = 4$ probes per segment in our method and the same number of points in the transmittance-component estimation of the baseline methods.

*Differentiable rendering comparisons.* To directly validate our variance analysis, we visualize voxel-wise gradient variance of Jellyfish in Fig. 7; all methods are evaluated with the same volumetric grids. Compared to DRT, our method reduces extinction gradient variance by up to 80%, with the strongest reduction occurring in empty regions where $\sigma_t \approx 0$—explaining the substantial suppression of floating artifacts observed in reconstructions. Quadratic variants exhibit lower variance than linear variants for both methods, yet our method maintains significantly lower variance regardless of algorithmic complexity, confirming the importance of introducing negative covariance. The mean gradient values across the 2D slice spatially match DRT's, verifying that our estimator remains unbiased.

*Inverse rendering comparisons.* We compare our method against DRT under an equal computational budget of 30–90 minutes of optimization per scene per method, evaluating both linear- and quadratic-complexity variants of each. Fig. 5 and Fig. 6 show relighting results across all scenes: our method produces consistently cleaner empty regions, whereas DRT exhibits noticeable floating artifacts across different scene types; for some heterogeneous volumes (Teapot and Bunny-Cloud), our method also recovers finer geometric and appearance details, which we attribute to more accurate radiance estimation enabled by fewer floating artifacts. Quantitatively, both the linear- and quadratic-complexity variants of our method outperform their corresponding DRT counterparts in PSNR, and our linear variant often matches or exceeds the reconstruction quality of DRT's quadratic variant—suggesting that variance reduction via induced covariance is more critical than increasing the raw sample count. Table 2 reports the full set of equal-time metrics (PSNR, SSIM, LPIPS, RMSE) for all eight scenes; for mean value, our quadratic variant is the top performer on every metric.

### 6.2 Ablation Study on the Number of Gradient Samples

We study how the number of gradient samples $\Lambda$ per segment affects reconstruction quality. We test on the Dust-Devil and Teapot scenes under equal-time training, varying $\Lambda$ while keeping all other settings unchanged. As shown in Figure 10, PSNR improves monotonically as $\Lambda$ increases from 1 to 8, with the largest gain from $\Lambda = 1$ to $\Lambda = 2$ and saturation beyond $\Lambda = 4$. Two factors limit further gains. First, only the direct-lighting contribution scales with $\Lambda$; the indirect-lighting contribution is evaluated recursively only once per segment regardless of $\Lambda$, so its effective sample count is bounded above by one. Second, each additional probe adds work per segment, increasing per-iteration runtime and reducing the number of optimization steps that fit into the equal-time budget. Even so, modest sample counts ($\Lambda = 2 \sim 4$) already provide substantial benefits, and we set $\Lambda = 4$ as our default to match DRT's transmittance-component sample count for a fair comparison.

## 7 Discussion and Conclusion

*Limitations and future work.* Our method assumes physics-based light transport; non-physical models such as linear transmittance or emission-only representations fall outside its scope.

Our implementation uses the dual-suffix-path method (Equation 35) for GPU efficiency. The single-suffix-path variant (Equation 32) would in principle yield further variance reduction, but its divergent reconnection ray and tightly coupled transmittance-ratio tracker break a single-megakernel layout. A megakernel-friendly implementation of this variant is worth future exploration.

Our per-segment probe averaging (Section 5.3) also carries an asymmetry: only one probe per segment recursively traces the indirect contribution, while the remaining $\Lambda - 1$ probes evaluate the direct contribution alone. Reducing this indirect sampling cost—via neural radiance caching or hierarchical sample reuse—is another natural avenue.

Finally, extending sample matching to volume-rendering-based neural methods (e.g., NeRF and 3DGS) offers a further direction.

*Conclusion.* We presented *sample matching*, a principle for differentiable volume rendering grounded in a structural observation about the radiative transfer formulation: differentiating with respect to extinction produces two components—scattering and transmittance—on mismatched integration domains that enter the gradient with opposite signs, inducing a negative correlation that prior estimators leave unexploited by sampling the two independently. To activate this correlation, we derived a reformulation of the differential path integral that couples both contributions within a single integrand, and from it obtained an unbiased Monte Carlo estimator whose scattering and transmittance components share sample positions by construction. To our knowledge, this is the first systematic exploitation of the negative correlation inherent to extinction gradients. Empirically, the resulting estimator reduces gradient variance by up to 80% and suppresses floating artifacts in inverse rendering.

## Acknowledgments

We thank the anonymous reviewers for their insightful feedback. We thank Wenzel Jakob for early discussions and for his advice on system implementation, and Nobuyuki Umetani for his advice and support throughout this work. We also thank Haolin Lu for help with the dataset. The environment maps and fabric used in the teaser were provided by PolyHaven; the marble pillar in the teaser is from Sketchfab by Louqeekon. The first author, Ruihan Yu, is deeply grateful to everyone around him who provided resources and personal support during this research.

## References

Yash Belhe, Bing Xu, Sai Praveen Bangaru, Ravi Ramamoorthi, and Tzu-Mao Li. 2024. Importance Sampling BRDF Derivatives. *ACM Trans. Graph.* 43, 3 (2024). doi:10.1145/3648611

John C Butcher and Harry Messel. 1958. Electron number distribution in electron-photon showers. *Physical Review* 112, 6 (1958), 2096.

Yunchi Cen, Chen Li, Frederick W. B. Li, Bailin Yang, and Xiaohui Liang. 2023. A Differential Diffusion Theory for Participating Media. *Computer Graphics Forum* 42, 7 (2023), e14956. doi:10.1111/cgf.14956

Subrahmanyan Chandrasekhar. 1960. *Radiative Transfer.* Dover Publications, New York. Originally published by Oxford University Press, 1950.

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 (April 1968), 76–81. doi:10.13182/NSE68-1

S. N. 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. doi:10.13182/NSE78-A27154

Xi Deng, Fujun Luan, Bruce Walter, Kavita Bala, and Steve Marschner. 2022. Reconstructing Translucent Objects using Differentiable Rendering. In *ACM SIGGRAPH 2022 Conference Proceedings (SIGGRAPH '22).* ACM, New York, NY, USA, Article 38, 10 pages. doi:10.1145/3528233.3530714

Zhimin Fan, Pengcheng Shi, Mufan Guo, Ruoyu Fu, Yanwen Guo, and Jie Guo. 2024. Conditional Mixture Path Guiding for Differentiable Rendering. *ACM Trans. Graph.* 43, 4 (2024). doi:10.1145/3658133

Iliyan Georgiev, Zackary Misso, Toshiya Hachisuka, Derek Nowrouzezahrai, Jaroslav Křivánek, and Wojciech Jarosz. 2019. Integral Formulations of Volumetric Transmittance. *ACM Trans. Graph.* 38, 6 (Nov. 2019), 154:1–154:17. doi:10.1145/3355089.3356559

Toshiya Hachisuka, Iliyan Georgiev, Wojciech Jarosz, Jaroslav Křivánek, and Derek Nowrouzezahrai. 2017. Extended Path Integral Formulation for Volumetric Transport. In *Eurographics Symposium on Rendering – Experimental Ideas & Implementations.* Eurographics Association.

Toshiya Hachisuka, Jacopo Pantaleoni, and Henrik Wann Jensen. 2012. A Path Space Extension for Robust Light Transport Simulation. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia)* 31, 6 (2012), 191:1–191:10.

Tomáš Iser, Tobias Rittig, and Alexander Wilkie. 2025. Scattering-Aware Color Calibration for 3D Printers Using a Simple Calibration Target. *ACM Transactions on*

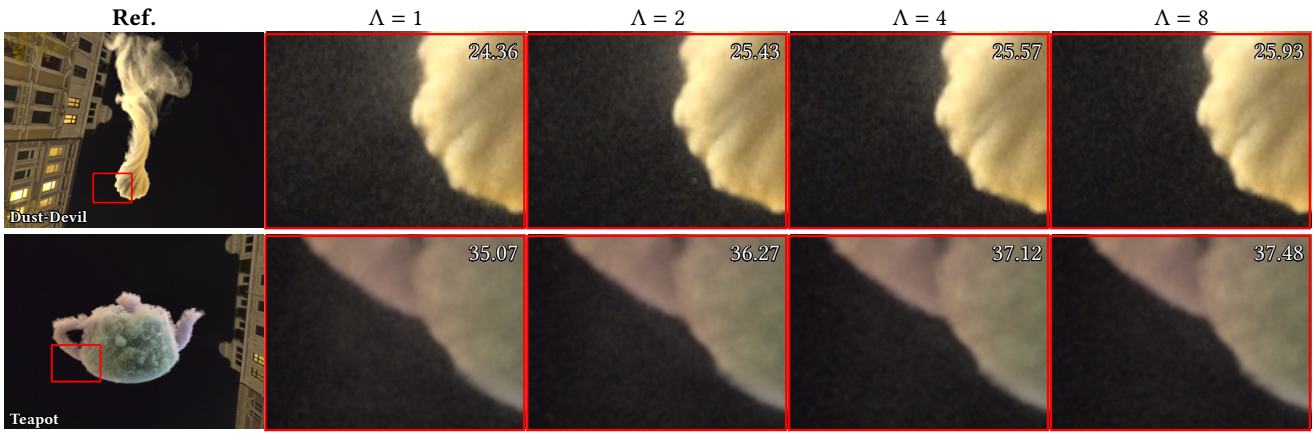


Fig. 10. **Ablation Study on the Number of Gradient Samples (Quadratic Complexity).** Effect of the number of samples (Λ = 1, 2, 4, 8) on relighting quality. Numbers denote PSNR values.

Table 2. **Per-scene equal-time metrics for the relighting comparison in Figure 5 and Figure 6.** For each of the eight scenes and each of the four method variants, we report PSNR (↑), SSIM (↑), LPIPS (↓), and RMSE (↓). PSNR and RMSE are evaluated in linear HDR space, and SSIM and LPIPS are evaluated on tone-mapped sRGB images. The best result among the four methods in each metric block is shown in bold. **O-Q** = Ours (Quadratic), **O-L** = Ours (Linear), **D-Q** = DRT (Quadratic), **D-L** = DRT (Linear).

| | PSNR ↑ | | | | SSIM ↑ | | | | LPIPS ↓ | | | | RMSE ↓ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Scene** | O-Q | O-L | D-Q | D-L | O-Q | O-L | D-Q | D-L | O-Q | O-L | D-Q | D-L | O-Q | O-L | D-Q | D-L |
| Dragon | **18.33** | 17.96 | 17.34 | 17.00 | **0.593** | 0.564 | 0.558 | 0.531 | **0.445** | 0.468 | 0.453 | 0.487 | **0.121** | 0.127 | 0.136 | 0.141 |
| Bunny-Cloud | **30.93** | 30.19 | 26.24 | 25.55 | **0.753** | 0.732 | 0.568 | 0.554 | **0.287** | 0.310 | 0.423 | 0.433 | **0.028** | 0.031 | 0.049 | 0.053 |
| Teapot | **33.79** | 32.80 | 25.19 | 24.34 | **0.677** | 0.657 | 0.535 | 0.532 | **0.401** | 0.419 | 0.529 | 0.541 | **0.020** | 0.023 | 0.055 | 0.061 |
| Jellyfish | **36.05** | 36.00 | 30.90 | 30.65 | **0.587** | 0.584 | 0.447 | 0.443 | **0.477** | 0.478 | 0.605 | 0.606 | **0.016** | 0.016 | 0.029 | 0.029 |
| Dust-Devil | **31.70** | 31.04 | 27.10 | 26.39 | **0.785** | 0.779 | 0.725 | 0.718 | **0.258** | 0.263 | 0.320 | 0.326 | **0.026** | 0.028 | 0.044 | 0.048 |
| Scarf | **22.37** | 22.24 | 21.88 | 21.87 | **0.760** | 0.753 | 0.644 | 0.643 | **0.330** | 0.335 | 0.439 | 0.434 | **0.076** | 0.077 | 0.081 | 0.081 |
| Rover | 18.86 | **18.87** | 18.80 | 18.85 | **0.556** | 0.553 | 0.537 | 0.534 | **0.439** | 0.444 | 0.469 | 0.469 | 0.114 | **0.114** | 0.115 | 0.114 |
| Astronaut-Rotated | **19.04** | 17.89 | 18.17 | 17.25 | **0.689** | 0.657 | 0.588 | 0.561 | **0.348** | 0.376 | 0.378 | 0.404 | **0.112** | 0.127 | 0.123 | 0.137 |
| **Mean** | **26.38** | 25.87 | 23.20 | 22.74 | **0.675** | 0.660 | 0.575 | 0.565 | **0.373** | 0.387 | 0.452 | 0.463 | **0.064** | 0.068 | 0.079 | 0.083 |

*Graphics (TOG)* 44, 6 (2025), 1–14.

Wenzel Jakob, Sébastien Speierer, Nicolas Roussel, Merlin Nimier-David, Delio Vicini, Tizian Zeltner, Baptiste Nicolet, Miguel Crespo, Vincent Leroy, and Ziyi Zhang. 2022. Mitsuba 3 Renderer. https://mitsuba-renderer.org.

Wojciech Jarosz, Derek Nowrouzezahrai, Iman Sadeghi, and Henrik Wann Jensen. 2011a. A Comprehensive Theory of Volumetric Radiance Estimation using Photon Points and Beams. *ACM Transactions on Graphics* 30, 1 (2011), 5:1–5:19.

Wojciech Jarosz, Derek Nowrouzezahrai, Robert Thomas, Peter-Pike Sloan, and Matthias Zwicker. 2011b. Progressive Photon Beams. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia)* 30, 6 (2011), 181:1–181:12.

Wojciech Jarosz, Matthias Zwicker, and Henrik Wann Jensen. 2008. The Beam Radiance Estimate for Volumetric Photon Mapping. *Computer Graphics Forum (Proceedings of Eurographics)* 27, 2 (2008), 557–566.

James T Kajiya and Brian P Von Herzen. 1984. Ray tracing volume densities. *ACM SIGGRAPH computer graphics* 18, 3 (1984), 165–174.

Csaba Kelemen, László Szirmay-Kalos, György Antal, and Ferenc Csonka. 2002. A Simple and Robust Mutation Strategy for the Metropolis Light Transport Algorithm. *Computer Graphics Forum* 21, 3 (2002), 531–540. doi:10.1111/1467-8659.00703

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 2023. 3D Gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.* 42, 4 (2023), 139–1.

Markus Kettunen, Eugene D'Eon, Jacopo Pantaleoni, and Jan Novák. 2021. An unbiased ray-marching transmittance estimator. *ACM Transactions on Graphics (TOG)* 40, 4 (2021), 1–20.

Markus Kettunen, Marco Manzi, Miika Aittala, Jaakko Lehtinen, Frédo Durand, and Matthias Zwicker. 2015. Gradient-Domain Path Tracing. *ACM Trans. Graph.* 34, 4 (2015), 123:1–123:13. doi:10.1145/2766997

Jaroslav Křivánek, Iliyan Georgiev, Toshiya Hachisuka, Petr Vévoda, Martin Šik, Derek Nowrouzezahrai, and Wojciech Jarosz. 2014. Unifying Points, Beams, and Paths in Volumetric Light Transport Simulation. *ACM Trans. Graph.* 33, 4, Article 103 (July 2014), 103:1–103:13 pages. doi:10.1145/2601097.2601110

Christopher Kulla. 2011. Decoupled Ray-Marching of Heterogeneous Participating Media. In *ACM SIGGRAPH 2011 Talks (SIGGRAPH '11)*. ACM, New York, NY, USA, Article 56, 1 pages. doi:10.1145/2037826.2037882

Peter Kutz, Ralf Habel, Yining Karl Li, and Jan Novák. 2017. Spectral and Decomposition Tracking for Rendering Heterogeneous Volumes. *ACM Trans. Graph.* 36, 4, Article 111 (July 2017), 111:1–111:16 pages. doi:10.1145/3072959.3073665

Eric P. Lafortune and Yves D. Willems. 1996. Rendering Participating Media with Bidirectional Path Tracing. In *Rendering Techniques '96 (Proceedings of the Eurographics Workshop on Rendering)*. Springer, Vienna, 91–100. doi:10.1007/978-3-7091-7484-5_10

Jaakko Lehtinen, Tero Karras, Samuli Laine, Miika Aittala, Frédo Durand, and Timo Aila. 2013. Gradient-Domain Metropolis Light Transport. *ACM Trans. Graph.* 32, 4 (2013), 95:1–95:12. doi:10.1145/2461912.2461943

Ludwic Leonard, Nils Thurey, and Rüdiger Westermann. 2025. Light Transport-aware Diffusion Posterior Sampling for Single-View Reconstruction of 3D Volumes. In *Proceedings of the Computer Vision and Pattern Recognition Conference*. 16163–16174.

Daqi Lin, Chris Wyman, and Cem Yuksel. 2021. Fast Volume Rendering with Spatiotemporal Reservoir Resampling. *ACM Trans. Graph.* 40, 6, Article 279 (Dec. 2021),

279:1–279:18 pages. doi:10.1145/3478513.3480499
Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Commun. ACM* 65, 1 (2021), 99–106.
Bailey Miller, Iliyan Georgiev, and Wojciech Jarosz. 2019. A Null-Scattering Path Integral Formulation of Light Transport. *ACM Trans. Graph.* 38, 4 (July 2019), 44:1–44:13. doi:10.1145/3306346.3323025
Zackary Misso, Yining Karl Li, Brent Burley, Daniel Teece, and Wojciech Jarosz. 2023. Progressive Null-Tracking for Volumetric Rendering. In *ACM SIGGRAPH 2023 Conference Proceedings (SIGGRAPH '23)*. ACM, New York, NY, USA, Article 31, 10 pages. doi:10.1145/3588432.3591557
Baptiste Nicolet, Fabrice Rousselle, Jan Novak, Alexander Keller, Wenzel Jakob, and Thomas Müller. 2023. Recursive control variates for inverse rendering. *ACM Transactions on Graphics (TOG)* 42, 4 (2023), 1–13.
Baptiste Nicolet, Felix Wechsler, Jorge Madrid-Wolff, Christophe Moser, and Wenzel Jakob. 2024. Inverse rendering for tomographic volumetric additive manufacturing. *ACM Transactions on Graphics (TOG)* 43, 6 (2024), 1–17.
Merlin Nimier-David, Thomas Müller, Alexander Keller, and Wenzel Jakob. 2022. Unbiased Inverse Volume Rendering with Differential Trackers. *ACM Trans. Graph.* 41, 4, Article 44 (July 2022), 20 pages. doi:10.1145/3528223.3530073
Merlin Nimier-David, Sébastien Speierer, Benoît Ruiz, and Wenzel Jakob. 2020. Radiative Backpropagation: An Adjoint Method for Lightning-Fast Differentiable Rendering. *ACM Transactions on Graphics (Proceedings of SIGGRAPH)* 39, 4 (2020). doi:10.1145/3386569.3392406
Jan Novák, Iliyan Georgiev, Johannes Hanika, Jaroslav K˘ rivánek, and Wojciech Jarosz. 2018. Monte Carlo methods for physically based volume rendering. In *ACM SIGGRAPH Courses*. doi:10/c5fj
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.
Ekin Öztürk, Rob Akers, Stanislas Pamela, Pieter Peers, Abhijeet Ghosh, MAST Team, et al. 2025. Inverse rendering of fusion plasmas: inferring plasma composition from imaging systems. *Nuclear Fusion* 65, 2 (2025), 026020.
Mark Pauly, Thomas Kollig, and Alexander Keller. 2000. Metropolis Light Transport for Participating Media. In *Rendering Techniques 2000*. Springer, 11–22.
Christoph Peters. 2025. Jackknife Transmittance and MIS Weight Estimation. *ACM Transactions on Graphics (TOG)* 44, 6 (2025), 1–16.
Holly E Rushmeier and Kenneth E Torrance. 1987. The zonal method for calculating light intensities in the presence of a participating medium. *ACM SIGGRAPH Computer Graphics* 21, 4 (1987), 293–302.
Eric Veach. 1997. *Robust Monte Carlo Methods for Light Transport Simulation*. Ph. D. Dissertation. Stanford University.
Delio Vicini, Sébastien Speierer, and Wenzel Jakob. 2021. Path Replay Backpropagation: Differentiating Light Paths using Constant Memory and Linear Time. *ACM Trans. Graph.* 40, 4 (Aug. 2021), 108:1–108:14. doi:10.1145/3450626.3459804
Yu-Chen Wang, Chris Wyman, Lifan Wu, and Shuang Zhao. 2023. Amortizing Samples in Physics-Based Inverse Rendering Using ReSTIR. *ACM Trans. Graph.* 42, 6 (2023), 214:1–214:17. doi:10.1145/3618331
E Woodcock, T Murphy, P Hemmings, and S Longworth. 1965. Techniques used in the GEM code for Monte Carlo neutronics calculations in reactors and other systems of complex geometry. In *Proc. Conf. Applications of Computing Methods to Reactor Problems*, Vol. 557. Argonne National Laboratory.
Wei Wu, Bo Wang, Miloš Hašan, Liwen Zhang, Zhenzhong Jin, and Ling-Qi Yan. 2024. Efficient participating media rendering with differentiable regularization. *Computational Visual Media* 10, 5 (Oct. 2024), 937–948. doi:10.1007/s41095-023-0372-2
Ryoichi Yamakoshi, Tatsushi Kaneko, Guoqing Hao, Yonghao Yue, and Kazuhiko Sumi. 2025. Efficient inverse volume rendering via path filtering for optical property estimation in scattering media. *Optics Express* 33, 22 (2025), 45539–45550.
Kai Yan, Vincent Pegoraro, Marc Droske, Jiří Vorba, and Shuang Zhao. 2024. Differentiating Variance for Variance-Aware Inverse Rendering. In *SIGGRAPH Asia 2024 Conference Papers*. 1–10.
Kai Yan, Cheng Zhang, Sébastien Speierer, Guangyan Cai, Yufeng Zhu, Zhao Dong, and Shuang Zhao. 2025. Image-space Adaptive Sampling for Fast Inverse Rendering. In *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers*. 1–11.
Tizian Zeltner, Sébastien Speierer, Iliyan Georgiev, and Wenzel Jakob. 2021. Monte Carlo estimators for differential light transport. *ACM Transactions on Graphics (TOG)* 40, 4 (2021), 1–16.
Cheng Zhang, Zhao Dong, Michael Doggett, and Shuang Zhao. 2021a. Antithetic sampling for Monte Carlo differentiable rendering. *ACM Transactions on Graphics (TOG)* 40, 4 (2021), 1–12.
Cheng Zhang, Bailey Miller, Kai Yan, Ioannis Gkioulekas, and Shuang Zhao. 2020. Path-Space Differentiable Rendering. *ACM Trans. Graph.* 39, 4 (2020), 143:1–143:19. doi:10.1145/3386569.3392383
Cheng Zhang, Lifan Wu, Changxi Zheng, Ioannis Gkioulekas, Ravi Ramamoorthi, and Shuang Zhao. 2019. A differential theory of radiative transfer. *ACM Transactions on Graphics (TOG)* 38, 6 (2019), 1–16.
Cheng Zhang, Zihan Yu, and Shuang Zhao. 2021b. Path-Space Differentiable Rendering of Participating Media. *ACM Trans. Graph.* 40, 4 (2021), 76:1–76:15. doi:10.1145/3450626.3459782
Shuang Zhao, Wenzel Jakob, and Tzu-Mao Li. 2020. Physics-based Differentiable Rendering: A Comprehensive Introduction. In *ACM SIGGRAPH 2020 Courses*. https://courses.shuangz.com/pbdr-course-sg20/.

# A Detailed Derivations

In what follows, we present detailed derivations for the key results in Section 4.

## A.1 Derivation for Equation 22

In the following, we derive Equation 22 in Section 4.1—a new formulation for Equation 21 that allows the derivatives to be evaluated at identical locations.

To start, we note that

$$T(t) = \int_t^{t_{\max}} \sigma_t(s)\, T(s)\, \mathrm{d}s + T(t_{\max}), \tag{37}$$

for any $0 < t < t_{\max}$, since both sides equal the probability that an exponentially distributed random variable exceeds $t$.

Using this relation, the scattering component of Equation 21 can be rewritten as

$$\begin{aligned} &\int_0^{t_{\max}} \partial_{\boldsymbol{\theta}} v_j(t)\, T(t)\, f_{\setminus j}(\bar{\mathbf{x}})\, \mathrm{d}t \\ &= \int_0^{t_{\max}} \int_t^{t_{\max}} \partial_{\boldsymbol{\theta}} v_j(t)\, \sigma_t(s)\, T(s)\, f_{\setminus j}(\bar{\mathbf{x}})\, \mathrm{d}s\, \mathrm{d}t \\ &\quad + T(t_{\max}) \int_0^{t_{\max}} \partial_{\boldsymbol{\theta}} v_j(t)\, f_{\setminus j}(\bar{\mathbf{x}})\, \mathrm{d}t. \end{aligned} \tag{38}$$

We now rewrite the double integral in Equation 38 so that it can be merged with the transmittance term in Equation 21. To this end, we swap the names of $s$ and $t$, yielding:

$$\begin{aligned} &\int_0^{t_{\max}} \int_t^{t_{\max}} \partial_{\boldsymbol{\theta}} v_j(t)\, \sigma_t(s)\, T(s)\, f_{\setminus j}(\bar{\mathbf{x}})\, \mathrm{d}s\, \mathrm{d}t \\ &= \int_0^{t_{\max}} \int_s^{t_{\max}} \partial_{\boldsymbol{\theta}} v_j(s)\, \sigma_t(t)\, T(t)\, f_{\setminus j}(\bar{\mathbf{y}})\, \mathrm{d}t\, \mathrm{d}s \\ &= \int_0^{t_{\max}} \int_0^{t} \partial_{\boldsymbol{\theta}} v_j(s)\, \sigma_t(t)\, T(t)\, f_{\setminus j}(\bar{\mathbf{y}})\, \mathrm{d}s\, \mathrm{d}t, \end{aligned} \tag{39}$$

where $\bar{\mathbf{y}}$ is identical to $\bar{\mathbf{x}}$ except that its $j$-th vertex is $\boldsymbol{x}_{j-1} + s\boldsymbol{\omega}$ instead of $\boldsymbol{x}_{j-1} + t\boldsymbol{\omega}$, and the last equality follows from Fubini's theorem.

Substituting Equation 39 back into Equation 38 and combining with the transmittance term of Equation 21 produces Equation 22.

## A.2 Derivation for Equation 28

We now derive Equation 28 in Section 4.1.

To start, we rewrite Equation 27 by reparameterizing the light vertex $\boldsymbol{x}_K$ as $\boldsymbol{x}_K = \boldsymbol{x}_{K-1} + t_{\max}\,\boldsymbol{\omega}$, where $t_{\max}$ is the distance from

$\boldsymbol{x}_{K-1}$ to the first surface intersection along $\boldsymbol{\omega} \in \mathcal{S}^2$:

$$\begin{aligned}(\partial_{\boldsymbol{\theta}} I)_{K,K} &= \int_{\Omega^{\text{pre}}_{K-1}} \int_{\mathcal{S}^2} v_K(\bar{\mathbf{x}})\, \partial_{\boldsymbol{\theta}} T(t_{\max})\, f_{\backslash K}(\bar{\mathbf{x}})\, \mathrm{d}\boldsymbol{\omega}\, \mathrm{d}\bar{\mathbf{x}}_{\backslash K} \\ &= -\int_{\Omega^{\text{pre}}_{K-1}} \int_{\mathcal{S}^2} \int_0^{t_{\max}} v_K(\bar{\mathbf{x}})\, \partial_{\boldsymbol{\theta}} \sigma_t(t)\, T(t_{\max}) \\ &\qquad\qquad f_{\backslash K}(\bar{\mathbf{x}})\, \mathrm{d}t\, \mathrm{d}\boldsymbol{\omega}\, \mathrm{d}\bar{\mathbf{x}}_{\backslash K},\end{aligned} \tag{40}$$

where $\Omega^{\text{pre}}_{K-1}$ is the prefix path space defined in Equation 16, and the last equality follows from Equation 20.

By setting $\boldsymbol{y} = \boldsymbol{x}_{K-1} + t\,\boldsymbol{\omega}$ and combining the inner integrals with respect to $\boldsymbol{\omega}$ and $t$ into a single integral over the scene volume $\mathcal{V}$, we obtain

$$\begin{aligned}(\partial_{\boldsymbol{\theta}} I)_{K,K} = -\int_{\Omega^{\text{pre}}_{K-1}} \int_{\mathcal{V}} & L_{\text{e}}(\boldsymbol{x}_K \to \boldsymbol{y})\, \partial_{\boldsymbol{\theta}} \sigma_t(\boldsymbol{y}) \\ & T(\boldsymbol{x}_{K-1} \leftrightarrow \boldsymbol{x}_K)\, G(\boldsymbol{x}_{K-1} \leftrightarrow \boldsymbol{y})\, f_{\backslash K}(\bar{\mathbf{x}})\, \mathrm{d}\boldsymbol{y}\, \mathrm{d}\bar{\mathbf{x}}_{\backslash K}.\end{aligned} \tag{41}$$

Relabeling $\boldsymbol{y}$ as $\boldsymbol{x}_K$ (and the original $\boldsymbol{x}_K$ as $\boldsymbol{x}_K^{\perp}$) produces Equation 28.

### A.3 Derivation for Equation 30

Substituting Equation 23 and Equation 28 into Equation 14 gives

$$\begin{aligned}\partial_{\boldsymbol{\theta}} I = & \sum_{K=2}^{\infty} \sum_{j=1}^{K-1} \int_{\Omega_K} s_j(\bar{\mathbf{x}})\, \mathcal{E}_j(\bar{\mathbf{x}})\, \mathrm{d}\bar{\mathbf{x}} + \\ & \sum_{K=2}^{\infty} \sum_{j=1}^{K-1} \int_{\Omega_K} \partial_{\boldsymbol{\theta}} v_j(\bar{\mathbf{x}})\, f_j^{\text{ins}}(\bar{\mathbf{x}})\, \mathrm{d}\bar{\mathbf{x}} \\ & - \sum_{K=1}^{\infty} \int_{\Omega^{\text{pre}}_K} \partial_{\boldsymbol{\theta}} \sigma_t(\boldsymbol{x}_K)\, f^{\text{out}}(\bar{\mathbf{x}})\, \mathrm{d}\bar{\mathbf{x}}.\end{aligned} \tag{42}$$

We first rewrite the first sum on the right-hand side:

$$\begin{aligned} & \sum_{K=2}^{\infty} \sum_{j=1}^{K-1} \int_{\Omega_K} s_j(\bar{\mathbf{x}})\, \mathcal{E}_j(\bar{\mathbf{x}})\, \mathrm{d}\bar{\mathbf{x}} \\ &= \sum_{j=1}^{\infty} \sum_{i=0}^{\infty} \int_{\Omega^{\text{pre}}_j} \int_{\Omega^{\text{suf}}_i} s_j(\bar{\mathbf{x}})\, \mathcal{E}_j(\bar{\mathbf{x}} + \bar{\mathbf{x}}^{\text{suf}})\, \mathrm{d}\bar{\mathbf{x}}^{\text{suf}}\, \mathrm{d}\bar{\mathbf{x}} \\ &= \sum_{j=1}^{\infty} \int_{\Omega^{\text{pre}}_j} s_j(\bar{\mathbf{x}}) \left( \sum_{i=0}^{\infty} \int_{\Omega^{\text{suf}}_i} \mathcal{E}_j(\bar{\mathbf{x}} + \bar{\mathbf{x}}^{\text{suf}})\, \mathrm{d}\bar{\mathbf{x}}^{\text{suf}} \right) \mathrm{d}\bar{\mathbf{x}} \\ &= \sum_{K=1}^{\infty} \int_{\Omega^{\text{pre}}_K} s_K(\bar{\mathbf{x}}) \int_{\Omega^{\text{suf}}} \mathcal{E}_K(\bar{\mathbf{x}} + \bar{\mathbf{x}}^{\text{suf}})\, \mathrm{d}\bar{\mathbf{x}}^{\text{suf}}\, \mathrm{d}\bar{\mathbf{x}},\end{aligned} \tag{43}$$

where $\Omega^{\text{pre}}_K$ is the prefix path space defined in Equation 16, $\Omega^{\text{suf}}_i$ and $\Omega^{\text{suf}}$ denote the suffix path spaces defined in Equation 17, and $\bar{\mathbf{x}} + \bar{\mathbf{x}}^{\text{suf}}$ denotes the path obtained by concatenating the prefix $\bar{\mathbf{x}}$ with the suffix $\bar{\mathbf{x}}^{\text{suf}}$.

We now rewrite the second sum on the right-hand side so that the derivatives $\partial_{\boldsymbol{\theta}} v_j(\bar{\mathbf{x}})$ and $\partial_{\boldsymbol{\theta}} \sigma_t(\boldsymbol{x}_K)$ are evaluated at identical locations:

$$\begin{aligned} & \sum_{K=2}^{\infty} \sum_{j=1}^{K-1} \int_{\Omega_K} \partial_{\boldsymbol{\theta}} v_j(\bar{\mathbf{x}})\, f_j^{\text{ins}}(\bar{\mathbf{x}})\, \mathrm{d}\bar{\mathbf{x}} \\ &= \sum_{j=1}^{\infty} \sum_{i=0}^{\infty} \int_{\Omega^{\text{pre}}_j} \int_{\Omega^{\text{suf}}_i} \partial_{\boldsymbol{\theta}} \sigma_t(\boldsymbol{x}_j)\, \hat{\rho}(\boldsymbol{x}_0^{\text{suf}} \to \boldsymbol{x}_j \to \boldsymbol{x}_{j-1}) \\ &\quad \cdot f_j^{\text{ins}}(\bar{\mathbf{x}} + \bar{\mathbf{x}}^{\text{suf}})\, \mathrm{d}\bar{\mathbf{x}}^{\text{suf}}\, \mathrm{d}\bar{\mathbf{x}} \\ &= \sum_{j=1}^{\infty} \int_{\Omega^{\text{pre}}_j} \partial_{\boldsymbol{\theta}} \sigma_t(\boldsymbol{x}_j) \left( \sum_{i=0}^{\infty} \int_{\Omega^{\text{suf}}_i} \hat{\rho}(\boldsymbol{x}_0^{\text{suf}} \to \boldsymbol{x}_j \to \boldsymbol{x}_{j-1}) \right. \\ &\quad \left. \cdot f_j^{\text{ins}}(\bar{\mathbf{x}} + \bar{\mathbf{x}}^{\text{suf}})\, \mathrm{d}\bar{\mathbf{x}}^{\text{suf}} \right) \mathrm{d}\bar{\mathbf{x}} \\ &= \sum_{K=1}^{\infty} \int_{\Omega^{\text{pre}}_K} \partial_{\boldsymbol{\theta}} \sigma_t(\boldsymbol{x}_K)\, I_{\text{ins}}(\bar{\mathbf{x}})\, \mathrm{d}\bar{\mathbf{x}},\end{aligned} \tag{44}$$

where $I_{\text{ins}}$ is defined in Equation 31.

Finally, substituting Equation 43 and Equation 44 into Equation 42 produces Equation 30 in Section 4.2.

## B Vector Variance Analysis for Extinction Gradients

Following the main text, we analyze the variance of extinction coefficient gradients for a single segment for simplicity. Without loss of generality, we consider the behavior of the extinction gradient vector along a single ray segment. The extinction coefficients $\sigma_t(\boldsymbol{x})$ are parameterized using a $d$-dimensional vector $\boldsymbol{\theta} \in \mathbb{R}^d$ (e.g., voxel coefficients). We denote by $\theta_k$ the $k$-th element of $\boldsymbol{\theta}$.

Since the extinction gradient is a vector, the estimator can be decomposed into two vector estimators, denoted as $\langle \mathbf{A} \rangle$ and $\langle \mathbf{B} \rangle$:

$$\left\langle \frac{\partial L}{\partial \boldsymbol{\theta}} \right\rangle \coloneqq \left\langle \left( \frac{\partial I}{\partial \boldsymbol{\theta}} \right)_{K,j} \right\rangle = \langle \mathbf{A} \rangle + \langle \mathbf{B} \rangle, \tag{45}$$

where $\langle \mathbf{A} \rangle$ and $\langle \mathbf{B} \rangle$ correspond to estimators for the transmittance and scattering components, respectively. For the $k$-th parameter, we denote the scalar estimator as:

$$\langle A_k \rangle \coloneqq \left\langle \frac{\partial L}{\partial \theta_k} \right\rangle_{trans} \quad \text{and} \quad \langle B_k \rangle \coloneqq \left\langle \frac{\partial L}{\partial \theta_k} \right\rangle_{scat}. \tag{46}$$

### B.1 Parameter-wise Stochastic Structure

Due to the sparsity induced by Monte Carlo sampling, the parameter vector $\boldsymbol{\theta}$ is effectively sparse, with many entries being zero. We can express the component estimators $\langle A_k \rangle$ and $\langle B_k \rangle$ as:

$$\langle A_k \rangle = \frac{\mathbf{1}_{k,\text{trans}}}{P_{k,\text{trans}}} \tilde{g}_{k,\text{trans}}, \quad \langle B_k \rangle = \frac{\mathbf{1}_{k,\text{scat}}}{P_{k,\text{scat}}} \tilde{g}_{k,\text{scat}}, \tag{47}$$

where $\mathbf{1}_{k,\cdot} \in \{0, 1\}$ indicates whether parameter $k$ receives a gradient during the backward pass. $P_{k,\text{trans}}$ and $P_{k,\text{scat}}$ are the corresponding sampling probabilities, and $\tilde{g}_{k,\text{trans}}$, $\tilde{g}_{k,\text{scat}}$ are unbiased conditional gradient estimators, which give the gradient values conditioned on parameter $k$ being sampled ($\mathbf{1}_k = 1$).

## B.2 Variance Decomposition

The vector variance is usually defined as the trace of the covariance matrix:

$$\mathrm{Var}\left[\left\langle\frac{\partial L}{\partial\boldsymbol{\theta}}\right\rangle\right] \coloneqq \mathrm{Tr}\left[\mathrm{Cov}\left[\left\langle\frac{\partial L}{\partial\boldsymbol{\theta}}\right\rangle\right]\right] = \mathbb{E}\left[\left\|\left\langle\frac{\partial L}{\partial\boldsymbol{\theta}}\right\rangle\right\|^2\right] - \left\|\mathbb{E}\left[\left\langle\frac{\partial L}{\partial\boldsymbol{\theta}}\right\rangle\right]\right\|^2. \tag{48}$$

For the dual estimator $\left\langle\frac{\partial L}{\partial\boldsymbol{\theta}}\right\rangle = \langle\mathbf{A}\rangle + \langle\mathbf{B}\rangle$, this decomposes into:

$$\mathrm{Var}\left[\left\langle\frac{\partial L}{\partial\boldsymbol{\theta}}\right\rangle\right] = \mathrm{Var}\left[\langle\mathbf{A}\rangle\right] + \mathrm{Var}\left[\langle\mathbf{B}\rangle\right] + 2\mathrm{Tr}\left[\mathrm{Cov}\left[\langle\mathbf{A}\rangle, \langle\mathbf{B}\rangle\right]\right]. \tag{49}$$

Using the structure in Equation 47 and $\mathbf{1}_k^2 = \mathbf{1}_k$, the individual variances are:

$$\mathrm{Var}\left[\langle\mathbf{A}\rangle\right] = \sum_{k=1}^{d}\left(\frac{\mathbb{E}\left[\tilde{g}_{k,\mathrm{trans}}^2\right]}{P_{k,\mathrm{trans}}} - \left(\mathbb{E}\left[\tilde{g}_{k,\mathrm{trans}}\right]\right)^2\right), \tag{50}$$

and similarly for $\mathrm{Var}\left[\langle\mathbf{B}\rangle\right]$.

The coupling term is:

$$\mathrm{Tr}\left[\mathrm{Cov}\left[\langle\mathbf{A}\rangle, \langle\mathbf{B}\rangle\right]\right] = \sum_{k=1}^{d}\left(\mathbb{E}\left[\langle A_k\rangle\ \langle B_k\rangle\right] - \mathbb{E}\left[\langle A_k\rangle\right]\ \mathbb{E}\left[\langle B_k\rangle\right]\right). \tag{51}$$

Assuming the conditional estimators $\tilde{g}_{k,\mathrm{scat}}$ and $\tilde{g}_{k,\mathrm{trans}}$ are independent given the sample location, let

$$P_{k,\mathrm{joint}} = \mathbb{E}\left[\mathbf{1}_{k,\mathrm{scat}}\mathbf{1}_{k,\mathrm{trans}}\right]$$

denote the joint sampling probability. Then:

$$\mathbb{E}\left[\langle A_k\rangle\ \langle B_k\rangle\right] = \frac{P_{k,\mathrm{joint}}}{P_{k,\mathrm{scat}}\,P_{k,\mathrm{trans}}}\mathbb{E}\left[\tilde{g}_{k,\mathrm{trans}}\right]\ \mathbb{E}\left[\tilde{g}_{k,\mathrm{scat}}\right]. \tag{52}$$

Substituting into Equation 51 yields:

$$\begin{aligned}&2\mathrm{Tr}\left[\mathrm{Cov}\left[\langle\mathbf{A}\rangle, \langle\mathbf{B}\rangle\right]\right]\\ &\quad= 2\sum_{k=1}^{d}\left(\frac{P_{k,\mathrm{joint}}}{P_{k,\mathrm{scat}}\,P_{k,\mathrm{trans}}} - 1\right)\mathbb{E}\left[\tilde{g}_{k,\mathrm{trans}}\right]\ \mathbb{E}\left[\tilde{g}_{k,\mathrm{scat}}\right].\end{aligned} \tag{53}$$

## B.3 Sample Matching for Maximum Correlation

Proposition B.1. *The absolute value of the coupling term Equation 53 is maximized when both estimators are evaluated at identical sample locations, i.e., when $P_{k,scat} = P_{k,trans} = P_{k,joint} = P_k$ for all $k$.*

Proof. The coupling coefficient $\frac{P_{k,\mathrm{joint}}}{P_{k,\mathrm{scat}}\,P_{k,\mathrm{trans}}}$ is maximized when $P_{k,\mathrm{joint}}$ is maximized. For fixed marginals $P_{k,\mathrm{scat}}$ and $P_{k,\mathrm{trans}}$, the joint probability satisfies $P_{k,\mathrm{joint}} \le \min(P_{k,\mathrm{scat}}, P_{k,\mathrm{trans}})$. Equality occurs when both sources sample identical locations: $\mathbf{1}_{k,\mathrm{scat}} = \mathbf{1}_{k,\mathrm{trans}}$ almost surely, which yields $P_{k,\mathrm{joint}} = P_{k,\mathrm{scat}} = P_{k,\mathrm{trans}} = P_k$. This gives $\frac{P_{k,\mathrm{joint}}}{P_{k,\mathrm{scat}}\,P_{k,\mathrm{trans}}} = \frac{1}{P_k}$, the maximum achievable value. □

Our *sample matching* strategy enables $P_{k,\mathrm{scat}} = P_{k,\mathrm{trans}} = P_{k,\mathrm{joint}} = P_k$ for all $k$, which maximizes the absolute value of the coupling term. Then the total variance becomes:

$$\begin{aligned}\mathrm{Var}\left[\left\langle\frac{\partial L}{\partial\boldsymbol{\theta}}\right\rangle\right] &= \mathrm{Var}\left[\langle\mathbf{A}\rangle\right] + \mathrm{Var}\left[\langle\mathbf{B}\rangle\right]\\ &\quad+ 2\sum_{k=1}^{d}\left(\frac{1}{P_k} - 1\right)\mathbb{E}\left[\tilde{g}_{k,\mathrm{trans}}\right]\ \mathbb{E}\left[\tilde{g}_{k,\mathrm{scat}}\right].\end{aligned} \tag{54}$$

Since $\mathbb{E}\left[\tilde{g}_{k,\mathrm{trans}}\right] < 0$ and $\mathbb{E}\left[\tilde{g}_{k,\mathrm{scat}}\right] > 0$ (opposite signs), the coupling term in Equation 53 reduces overall variance, especially for a high-resolution volume representation where $P_k$ is small.

In contrast, for DRT [Nimier-David et al. 2022], the coupling term vanishes:

$$\mathrm{Tr}\left[\mathrm{Cov}\left[\langle\mathbf{A}\rangle, \langle\mathbf{B}\rangle\right]\right] = 0. \tag{55}$$

Consequently, the total variance of the DRT estimator is simply the sum of individual variances $\mathrm{Var}\left[\langle\mathbf{A}\rangle\right] + \mathrm{Var}\left[\langle\mathbf{B}\rangle\right]$. Without the variance reduction benefit from the negative coupling term in Equation 54, the estimator suffers from severe noise.

It is worth mentioning that our analysis does not impose any restriction on the specific parameterization of the extinction field $\sigma_t(\boldsymbol{x})$. While our experiments primarily adopt a voxel-grid representation, the derivation is agnostic to the particular interpolation scheme of the voxel grid and can extend to alternative volumetric representations beyond voxel grids.