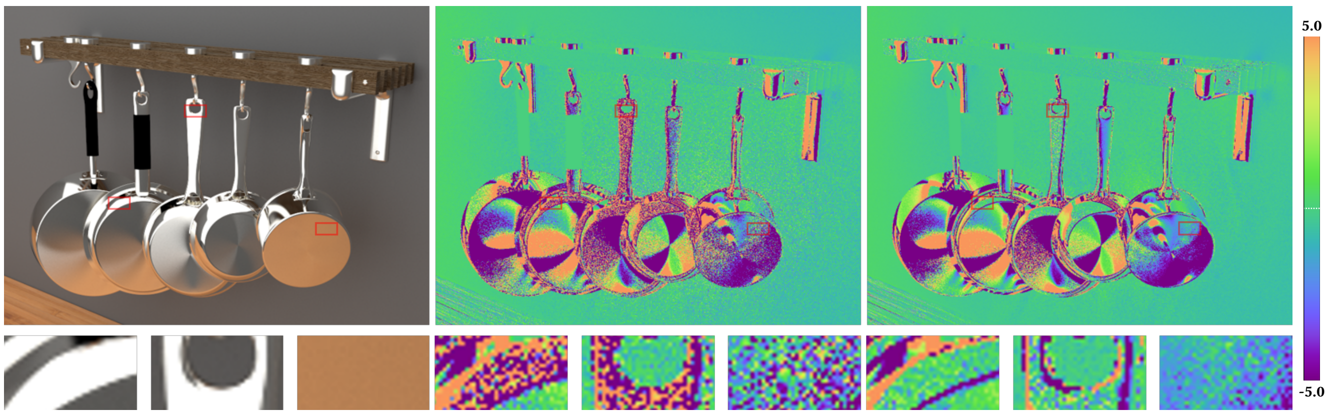
# Antithetic Sampling for Monte Carlo Differentiable Rendering

CHENG ZHANG, University of California, Irvine, USA and Facebook Reality Labs, USA
ZHAO DONG, Facebook Reality Labs, USA
MICHAEL DOGGETT, Lund University, Sweden and Facebook Reality Labs, USA
SHUANG ZHAO, University of California, Irvine, USA




(a) **Ordinary image** (b) **Deriv. image (w/o antithetic sampling)** (c) **Deriv. image (with antithetic sampling)**

Fig. 1. In physics-based differentiable rendering, previous sampling techniques developed for forward rendering have difficulties estimating geometric derivatives (i.e., those with respect to scene geometry), when the scene contains highly glossy or near-specular surfaces. To address this problem, we introduce antithetic sampling for Monte Carlo differentiable rendering. This example involves several pans exhibiting specular anisotropic reflection (a). When estimating derivatives with respect to the camera angle, state-of-the-art differentiable rendering methods produce high variance (b). By applying our technique to the same base algorithm, significant variance reduction can be achieved in equal time (c).

Stochastic sampling of light transport paths is key to Monte Carlo forward rendering, and previous studies have led to mature techniques capable of drawing high-contribution light paths in complex scenes. These sampling techniques have also been applied to differentiable rendering.

In this paper, we demonstrate that path sampling techniques developed for forward rendering can become inefficient for differentiable rendering of glossy materials—especially when estimating derivatives with respect to global scene geometries. To address this problem, we introduce *antithetic sampling* of BSDFs and light-transport paths, allowing significantly faster convergence and can be easily integrated into existing differentiable rendering pipelines. We validate our method by comparing our derivative estimates to those generated with existing unbiased techniques. Further, we demonstrate the effectiveness of our technique by providing equal-quality and equal-time comparisons with existing sampling methods.

CCS Concepts: • **Computing methodologies** → **Rendering**.

Authors' addresses: Cheng Zhang, chengz20@uci.edu, University of California, Irvine, USA and Facebook Reality Labs, USA; Zhao Dong, zhaodong@fb.com, Facebook Reality Labs, USA; Michael Doggett, Michael.Doggett@cs.lth.se, Lund University, Sweden and Facebook Reality Labs, USA; Shuang Zhao, shz@ics.uci.edu, University of California, Irvine, USA.


0730-0301/2021/8-ART77
https://doi.org/10.1145/3450626.3459783

Additional Key Words and Phrases: Differentiable rendering, antithetic sampling, glossy materials

**ACM Reference Format:**
Cheng Zhang, Zhao Dong, Michael Doggett, and Shuang Zhao. 2021. Antithetic Sampling for Monte Carlo Differentiable Rendering. *ACM Trans. Graph.* 40, 4, Article 77 (August 2021), 12 pages. https://doi.org/10.1145/3450626.3459783

## 1 INTRODUCTION

*Forward rendering* numerically estimates responses of radiometric detectors given virtual scenes with fully specified object geometries and optical material properties. *Differentiable rendering*, on the contrary, focuses on computing derivatives of radiometric detector responses (with respect to differential changes of virtual scenes) and have applications in many areas such as computational fabrication, computational imaging, and remote sensing.

Recently, great progresses have been made in physics-based differentiable rendering theory, algorithms, and systems [Li et al. 2018; Zhang et al. 2019; Loubet et al. 2019; Nimier-David et al. 2019; Zhang et al. 2020; Bangaru et al. 2020]. Consequently, it is now possible to differentiate with respect to arbitrary scene parameters including those controlling global geometries (e.g., the global orientation of an object or the position of a mesh vertex). It has been shown that differentiable rendering typically amounts to estimating *interior* and *boundary* integrals. The latter is unique to differentiable rendering and defined on discontinuity boundaries of the ordinary rendering integrals. To handle these *boundary* terms, several new techniques—such as Monte Carlo edge sampling [Li et al. 2018], reparameterization of the ordinary rendering integral [Loubet et al. 2019; Bangaru et al. 2020], and differential path integrals [Zhang et al. 2020]—have been introduced.

The *interior* integrals, on the other hand, share the same domain as those for forward rendering. To estimate these terms, previous differentiable rendering techniques have relied on existing stochastic sampling strategies developed for forward rendering that typically draw light paths with probability densities (approximately) proportional to their measurement contributions. Although this works adequately for relatively rough scenes, the sampling efficiency can be unsatisfactory for glossy scenes—especially when differentiating with respect to scene geometries—resulting in high variance. With near-specular reflection and refraction, the estimated derivatives can even have unbounded variance.

In this paper, we introduce new Monte Carlo sampling methods that leverage **antithetic sampling** [Geweke 1988]—a classic variance reduction technique—to efficiently estimate the *interior* integrals. Concretely, our contributions include:

- Introducing antithetic sampling for Monte Carlo differentiable rendering of glossy and near-specular BSDFs (§3). Our technique is applicable to most, if not all, differentiable-rendering formulations (such as differentiable path tracing [Li et al. 2018] and path-space differentiable rendering [Zhang et al. 2020]).
- Generalizing the BSDF antithetic sampling framework to handle full light transport paths (§4).

Physics-based differentiable rendering algorithms, when coupled with our antithetic-sampling technique, can have greatly improved efficiency when handling glossy materials. We demonstrate this by comparing derivatives estimated with and without our sampling technique in Figures 1, 7 and 8. Additionally, we show inverse-rendering comparisons in Figures 9, 10, and 11.

## 2 RELATED WORK AND PRELIMINARIES

We now briefly revisit physics-based forward and differentiable rendering in §2.1 and §2.2, respectively. Further, we present the basics of antithetic sampling in §2.3.

### 2.1 Forward Rendering

At the core of physics-based forward rendering is to solve the **rendering equation**,[1] an integral equation governing the steady-state radiance $L$:

$$L(\boldsymbol{x}, \boldsymbol{\omega}_\mathrm{o}) = \int_{\mathbb{S}^2} L_\mathrm{i}(\boldsymbol{x}, \boldsymbol{\omega}_\mathrm{i})\, f_\mathrm{s}(\boldsymbol{x}, \boldsymbol{\omega}_\mathrm{i}, \boldsymbol{\omega}_\mathrm{o})\, |\boldsymbol{n}(\boldsymbol{x}) \cdot \boldsymbol{\omega}_\mathrm{i}|\, \mathrm{d}\sigma(\boldsymbol{\omega}_\mathrm{i}), \tag{1}$$

where $L_\mathrm{i}$ indicates incident radiance; $f_\mathrm{s}$ is the bidirectional scattering distribution function (BSDF); $\boldsymbol{n}(\boldsymbol{x})$ denotes the (unit) surface normal at $\boldsymbol{x}$; "$\cdot$" indicates the dot (scalar) product operator; and $\sigma$ is the solid-angle measure.

Additionally, Veach has introduced the **path integral** formulation for simulating surface-only light transport. Under this formulation, the response of a radiometric detector can be expressed as

$$I = \int_\Omega f(\bar{\boldsymbol{x}})\, \mathrm{d}\mu(\bar{\boldsymbol{x}}), \tag{2}$$

where $\Omega := \bigcup_{N=1}^{\infty} \mathcal{M}^{N+1}$ is the **path space** comprised of all light transport paths (with finite lengths), and $\mu$ is the area-product measure. Further, for any **light path** $\bar{\boldsymbol{x}} = (\boldsymbol{x}_0, \ldots, \boldsymbol{x}_N) \in \Omega$, $f$ is the **measurement contribution function** given by

$$f(\bar{\boldsymbol{x}}) = L_\mathrm{e}(\boldsymbol{x}_0 \to \boldsymbol{x}_1) \left[ \prod_{i=1}^{N-1} G(\boldsymbol{x}_{i-1} \leftrightarrow \boldsymbol{x}_i)\, f_\mathrm{s}(\boldsymbol{x}_{i-1} \to \boldsymbol{x}_i \to \boldsymbol{x}_{i+1}) \right] G(\boldsymbol{x}_{N-1} \leftrightarrow \boldsymbol{x}_N)\, W_\mathrm{e}(\boldsymbol{x}_{N-1} \to \boldsymbol{x}_N), \tag{3}$$

where $L_\mathrm{e}$ and $W_\mathrm{e}$ indicate, respectively, **source emission** and **detector importance**; and $G$ is the **geometric term** defined as

$$G(\boldsymbol{y} \leftrightarrow \boldsymbol{x}) := \mathbb{V}(\boldsymbol{y} \leftrightarrow \boldsymbol{x})\, \frac{|\boldsymbol{n}(\boldsymbol{x}) \cdot \overrightarrow{\boldsymbol{xy}}|\, |\boldsymbol{n}(\boldsymbol{y}) \cdot \overrightarrow{\boldsymbol{yx}}|}{\|\boldsymbol{x} - \boldsymbol{y}\|^2}, \tag{4}$$

with $\mathbb{V}$ being the **mutual visibility function**, and $\overrightarrow{\boldsymbol{xy}}$ denoting the unit vector pointing from $\boldsymbol{x}$ toward $\boldsymbol{y}$.

*Sampling for forward rendering.* Stochastic sampling of light transport paths has been a key ingredient for forward rendering. For *local sampling* methods (such as unidirectional path tracing) that construct light paths vertex by vertex, path sampling largely amounts to drawing incident directions given exitant ones at each vertex based on the local BSDFs. Previously, a large variety of BSDF models have been proposed (e.g., [Phong 1975; Cook and Torrance 1982; Ashikhmin and Shirley 2000; Oren and Nayar 1994]), and most of them can be efficiently importance sampled. Among these models, microfacet BSDFs (e.g., [Cook and Torrance 1982; Ward 1992; Schlick 1994; van Ginneken et al. 1998; Kelemen and Szirmay-Kalos 2001; Pont and Koenderink 2002; Walter et al. 2007; Heitz et al. 2016; Lee et al. 2018; Xie and Hanrahan 2018]) have been widely adopted. Importance sampling of these models, therefore, have been studied thoroughly and can be performed by sampling the underlying normal distributions [Walter et al. 2007; Heitz and d'Eon 2014] or via localized Monte Carlo processes [Heitz et al. 2016].

### 2.2 Differentiable Rendering

The main objective of differentiable rendering is to compute gradients of detector responses $I$ with respect to arbitrary scene parameters, which typically requires differentiating Eq. (1) or Eq. (2).

*Differential rendering equation.* Let

$$f_\mathrm{RE}(\boldsymbol{\omega}_\mathrm{i}; \boldsymbol{x}, \boldsymbol{\omega}_\mathrm{o}) := L_\mathrm{i}(\boldsymbol{x}, \boldsymbol{\omega}_\mathrm{i})\, f_\mathrm{s}(\boldsymbol{x}, \boldsymbol{\omega}_\mathrm{i}, \boldsymbol{\omega}_\mathrm{o})\, |\boldsymbol{n}(\boldsymbol{x}) \cdot \boldsymbol{\omega}_\mathrm{i}|, \tag{5}$$

be the integrand of Eq. (1). Then, differentiating this equation with respect to a scene parameter $\theta \in \mathbb{R}$ yields the **differential rendering equation** [Li et al. 2018; Zhang et al. 2019]:

$$\frac{\mathrm{d}L(\boldsymbol{x}, \boldsymbol{\omega}_\mathrm{o})}{\mathrm{d}\theta} = \overbrace{\boxed{\int_{\mathbb{S}^2} \frac{\mathrm{d}f_\mathrm{RE}(\boldsymbol{\omega}_\mathrm{i}; \boldsymbol{x}, \boldsymbol{\omega}_\mathrm{o})}{\mathrm{d}\theta}\, \mathrm{d}\sigma(\boldsymbol{\omega}_\mathrm{i})}}^{\text{interior}} + \overbrace{\boxed{\int_{\Delta\mathbb{S}^2} V_{\Delta\mathbb{S}^2}(\boldsymbol{\omega}_\mathrm{i}; \boldsymbol{x}, \boldsymbol{\omega}_\mathrm{o})\, \Delta f_\mathrm{RE}(\boldsymbol{\omega}_\mathrm{i}; \boldsymbol{x}, \boldsymbol{\omega}_\mathrm{o})\, \mathrm{d}\ell(\boldsymbol{\omega}_\mathrm{i})}}^{\text{boundary}}, \tag{6}$$

[1] We hyperlink keywords to their definitions.

where the *boundary* integral is over curves $\Delta\mathbb{S}^2 \subset \mathbb{S}^2$ comprised of jump discontinuity points of $F_{\mathrm{RE}}$, and $V_{\Delta\mathbb{S}^2}$ denotes the scalar normal velocity (i.e., change rate) of the discontinuity point $\boldsymbol{\omega}_{\mathrm{i}}$ with respect to $\theta$. Further, the derivatives on both sides of Eq. (6) are essentially *material derivatives* (aka. *total derivatives*) that take into consideration dependencies of $\boldsymbol{x}$ and $\boldsymbol{\omega}_{\mathrm{o}}$ on the scene parameter $\theta$. This is usually the case when $\theta$ controls scene geometry such as the global pose of an object or the position of a mesh vertex.

Previous studies have shown that the *boundary* integral in the differential rendering equation of Eq. (6) can be handled by Monte Carlo edge sampling [Li et al. 2018; Zhang et al. 2019] or avoided altogether by reparameterizing the rendering equation [Loubet et al. 2019; Bangaru et al. 2020].

*Differential path integral.* Zhang et al. [2020] has shown that derivatives of the path integral of Eq. (2) can be expressed as **differential path integrals** in a similar fashion. Their material-form formulation starts with rewriting the ordinary path integral using a change of variable so that the new integral domain, which is called the *material path space*, is independent of the scene parameter $\theta$. To this end, assume that the scene geometry $\mathcal{M}$—which can depend on $\theta$—can be parameterized globally using a differentiable mapping $\mathsf{X}$ such that $\mathsf{X}(\cdot,\theta) : \mathcal{B} \mapsto \mathcal{M}(\theta)$ is a bijection from some fixed **reference configuration** $\mathcal{B}$ to the scene geometry $\mathcal{M}(\theta)$ for any $\theta$. Then, the **material path space** $\hat{\Omega}$ is defined to be the set of all finite-length paths over the reference configuration (that is, $\hat{\Omega} := \cup_{N=1}^{\infty} \mathcal{B}^{N+1}$). Further, the global parameterization $\mathsf{X}$ induces another path-wise mapping $\bar{\mathsf{X}}$ that, for each $\theta$, $\bar{\mathsf{X}}(\cdot,\theta)$ transforms each **material path** $\bar{\boldsymbol{p}} = (\boldsymbol{p}_0, \ldots, \boldsymbol{p}_N) \in \hat{\Omega}$ to an ordinary path $\bar{\boldsymbol{x}} = \bar{\mathsf{X}}(\bar{\boldsymbol{p}},\theta) := (\mathsf{X}(\boldsymbol{p}_0,\theta), \ldots, \mathsf{X}(\boldsymbol{p}_N,\theta)) \in \Omega(\theta)$.

In practice, when estimating derivatives of detector responses $\mathrm{d}I/\mathrm{d}\theta$ at some $\theta = \theta_0$, the reference configuration is typically set to the scene geometry at $\theta_0$: That is, $\mathcal{B} := \mathcal{M}(\theta_0)$. In this way, at $\theta = \theta_0$, the material path space $\hat{\Omega}$ to coincide with the ordinary one $\Omega(\theta_0)$. Further, the mappings $\mathsf{X}(\cdot,\theta_0)$ and $\bar{\mathsf{X}}(\cdot,\theta_0)$ reduce to identity maps.

Applying to the path integral of Eq. (2) the change of variable given by $\bar{\mathsf{X}}(\cdot,\theta)$ yields the **material-form path integral**:

$$I = \int_{\hat{\Omega}} \hat{f}(\bar{\boldsymbol{p}})\, \mathrm{d}\mu(\bar{\boldsymbol{p}}), \tag{7}$$

where the **material measurement contribution** $\hat{f}$ is given by the original measurement contribution of Eq. (3) and a Jacobian determinant capturing this change of variable:

$$\hat{f}(\bar{\boldsymbol{p}}) := f(\bar{\boldsymbol{x}}) \left\| \frac{\mathrm{d}\mu(\bar{\boldsymbol{x}})}{\mathrm{d}\mu(\bar{\boldsymbol{p}})} \right\|, \tag{8}$$

where $\bar{\boldsymbol{x}} = \bar{\mathsf{X}}(\bar{\boldsymbol{p}},\theta)$.

Differentiating Eq. (7), whose integral domain becomes independent of the scene parameter $\theta$, yields the **differential path integral**:

$$\frac{\mathrm{d}I}{\mathrm{d}\theta} = \overbrace{\boxed{\int_{\hat{\Omega}} \frac{\mathrm{d}\hat{f}(\bar{\boldsymbol{p}})}{\mathrm{d}\theta}\, \mathrm{d}\mu(\bar{\boldsymbol{p}})}}^{\text{interior}} + \overbrace{\boxed{\int_{\partial\hat{\Omega}} \Delta\hat{f}_K(\bar{\boldsymbol{p}})\, V_{\Delta\mathcal{B}_K}(\boldsymbol{p}_K)\, \mathrm{d}\mu'(\bar{\boldsymbol{p}})}}^{\text{boundary}}. \tag{9}$$

In this equation, the *interior* term is over the same material path space $\hat{\Omega}$ as the material-form path integral of Eq. (7). We refer the readers to the work by Zhang et al. [2020] for more details, including a complete definition of the *boundary* term, which is orthogonal to our technique.

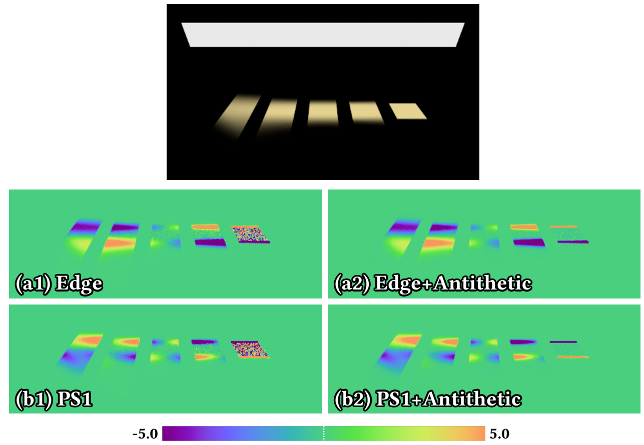



Fig. 2. **BSDF antithetic sampling:** This example contains a simple scene where a row of reflectors—whose roughnesses decrease from left to right—are lit by a large area light. When estimating derivatives (with respect to the rotation angle around the horizontal axis), the computational efficiency of conventional sampling methods declines when the surface roughnesses decrease, as shown in (a1) and (b1). We use "Edge" and "PS1" to indicate, respectively, differentiable path tracing with edge sampling [Li et al. 2018] and the unidirectional path-space method [Zhang et al. 2020]. Coupled with the same base methods, our BSDF antithetic sampling offers significant variance reduction in equal time, as shown in (a2) and (b2). We compute only the *interior* components of the derivatives given by Eq. (6) for (a) and Eq. (9) for (b), which differ numerically due to their varying parameterizations.

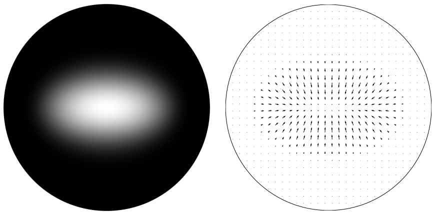
*Sampling for differentiable rendering.* When estimating the *interior* integrals in Eqs. (6) and (9) using Monte Carlo methods, previous works have relied on sampling techniques developed for forward rendering. Unfortunately, as we will demonstrate in §3, doing so can be highly inefficient when computing geometric derivatives (that is, those with respect to scene geometries).

## 2.3 Antithetic Sampling

*Antithetic sampling.* Being a classic variance reduction framework for Monte Carlo estimation, antithetic sampling [Hammersley and Mauldon 1956; Geweke 1988] has been studied in probabilistic inference and machine learning [Ren et al. 2019; Wu et al. 2019]. In computer graphics, this technique has been explored by several previous works in forward rendering [Subr et al. 2014; Öztireli 2016; Singh et al. 2019, 2020]. In Monte Carlo differentiable rendering, Bangaru et al. [2020] have applied antithetic sampling to efficiently handle discontinuity boundaries under the warped-area formulation.

*1D example.* The core idea of antithetic sampling is to use (negatively) correlated samples (instead of independent ones). We consider the problem of estimating

$$I := \int_{-\infty}^{\infty} F(x)\, \mathrm{d}x, \tag{10}$$



(a) **NDF** $D(\boldsymbol{\omega}_\mathrm{h})$ (b) **NDF deriv.** $\mathrm{d}D/\mathrm{d}\boldsymbol{\omega}_\mathrm{h}$

Fig. 3. Normal distributions $D(\boldsymbol{\omega}_\mathrm{h})$ usually exhibit point symmetry, causing their derivatives $\mathrm{d}D/\mathrm{d}\boldsymbol{\omega}_\mathrm{h}$—which is vector-valued—to possess point symmetry (with respect to the origin). We visualize an anisotropic normal distribution in (a) and its derivative in (b).

where the integrand $F$ is approximately an odd function with $F(x) \approx -F(-x)$. When $F$ contains high-magnitude positive and negative regions, estimating $I$ using Monte Carlo methods with independent samples can suffer from very slow convergence.

To address this problem, one can draw $x_1$ from some predetermined probability density $p$ and then set $x_2 := -x_1$, resulting in an antithetic estimator

$$\langle I \rangle_\mathrm{antithetic} := \frac{F(x_1) + F(x_2)}{p(x_1) + p(x_2)}. \tag{11}$$

Since $F(x_1) + F(x_2) \approx 0$, $\langle I \rangle_\mathrm{antithetic}$ can offer significantly lower variance.

In this paper, we introduce new antithetic estimators for Monte Carlo differentiable rendering of glossy and near-specular materials. Our key observation is that geometric derivatives of BSDFs, under certain parameterizations, are approximately odd functions.

## 3 ANTITHETIC SAMPLING OF GLOSSY BSDFS

The *interior* integrals in Eqs. (6) and (9) involve, respectively, derivatives of the integrand $F_\mathrm{RE}$ of the rendering equation (1) and the (material) measurement contribution function $\hat{f}$ with respect to the scene parameter $\theta$. In what follows, we address the problem of efficient Monte Carlo estimation of these *interior* integrals with the presence of highly glossy or near-specular BSDFs. We focus our derivations on Eq. (6) for simplicity, and the resulting algorithm applies to other formulations like Eq. (9).

Estimating the *interior* component of Eq. (6) using Monte Carlo integration requires stochastic sampling of the incident direction $\boldsymbol{\omega}_\mathrm{i}$. Previously, this has typically been achieved using standard BSDF sampling techniques developed for forward rendering [Li et al. 2018; Zhang et al. 2019; Loubet et al. 2019; Zhang et al. 2020]. Although this works adequately for rough BSDFs, it can lead to high variance for those that are glossy or near-specular.

To see why this is the case, we examine the integrand more closely. Specifically, according to the product rule, we have

$$\frac{\mathrm{d}F_\mathrm{RE}(\boldsymbol{\omega}_\mathrm{i}; \boldsymbol{x}, \boldsymbol{\omega}_\mathrm{o})}{\mathrm{d}\theta} = \frac{\mathrm{d}f_\mathrm{s}^\perp(\boldsymbol{x}, \boldsymbol{\omega}_\mathrm{i}, \boldsymbol{\omega}_\mathrm{o})}{\mathrm{d}\theta} L_\mathrm{i}(\boldsymbol{x}, \boldsymbol{\omega}_\mathrm{i}) + f_\mathrm{s}^\perp(\boldsymbol{x}, \boldsymbol{\omega}_\mathrm{i}, \boldsymbol{\omega}_\mathrm{o}) \frac{\mathrm{d}L_\mathrm{i}(\boldsymbol{x}, \boldsymbol{\omega}_\mathrm{i})}{\mathrm{d}\theta}. \tag{12}$$

where

$$f_\mathrm{s}^\perp(\boldsymbol{x}, \boldsymbol{\omega}_\mathrm{i}, \boldsymbol{\omega}_\mathrm{o}) := f_\mathrm{s}(\boldsymbol{x}, \boldsymbol{\omega}_\mathrm{i}, \boldsymbol{\omega}_\mathrm{o}) \, |\boldsymbol{n}(\boldsymbol{x}) \cdot \boldsymbol{\omega}_\mathrm{i}|. \tag{13}$$

**ALGORITHM 1:** Antithetic sampling of microfacet BSDFs

AntitheticBSDFSample($\boldsymbol{x}, \boldsymbol{\omega}_\mathrm{o}$)
**begin**
  Draw $\boldsymbol{\omega}_{\mathrm{h},1} = [x_\mathrm{h}, y_\mathrm{h}, z_\mathrm{h}] \sim p_\mathrm{h}$; // The ordinary sample
  Set $\boldsymbol{\omega}_{\mathrm{h},2} \leftarrow [-x_\mathrm{h}, -y_\mathrm{h}, z_\mathrm{h}]$; // The antithetic sample
  **for** $j \in \{1, 2\}$ **do**
    Compute incident direction $\boldsymbol{\omega}_{\mathrm{i},j}$ based on $\boldsymbol{\omega}_\mathrm{o}$ and $\boldsymbol{\omega}_{\mathrm{h},j}$;
    Compute $p_j := p(\boldsymbol{\omega}_{\mathrm{i},j})$ based on $p_\mathrm{h}(\boldsymbol{\omega}_{\mathrm{h},j})$;
  **end**
  **return** $(\boldsymbol{\omega}_{\mathrm{i},1},\ p_1,\ \boldsymbol{\omega}_{\mathrm{i},2},\ p_2)$;
**end**

Similar to Eq. (6), all derivatives (with respect to $\theta$) in Eq. (12) are material derivatives. Traditional BSDF sampling techniques typically draw $\boldsymbol{\omega}_\mathrm{i}$ with probability densities proportional to $f_\mathrm{s}^\perp$ (or $f_\mathrm{s}$). For differentiable rendering of glossy BSDFs, unfortunately, this is insufficient due to the vast difference between $\mathrm{d}f_\mathrm{s}^\perp/\mathrm{d}\theta$ and $f_\mathrm{s}^\perp$.

### 3.1 BSDF Antithetic Sampling

We have discussed in §2.3 that near-odd integrands with high-magnitude regions can lead to slow convergence of Monte Carlo integration. In the context of differentiable rendering, a common example of such functions are glossy or near-specular BSDFs (see Figure 2). To address this problem, we introduce an antithetic technique for BSDF sampling. We base our derivation on microfacet BSDFs that generally take the form

$$f_\mathrm{s}(\boldsymbol{\omega}_\mathrm{i}, \boldsymbol{\omega}_\mathrm{o}) = D(\boldsymbol{\omega}_\mathrm{h}) \, f_\mathrm{s}^{(0)}(\boldsymbol{\omega}_\mathrm{i}, \boldsymbol{\omega}_\mathrm{o}), \tag{14}$$

where $D$ is the **normal distribution function** (NDF) parameterized using the half-way vector $\boldsymbol{\omega}_\mathrm{h} := (\boldsymbol{\omega}_\mathrm{i} + \boldsymbol{\omega}_\mathrm{o}) / \|\boldsymbol{\omega}_\mathrm{i} + \boldsymbol{\omega}_\mathrm{o}\|$, and $f_\mathrm{s}^{(0)}$ captures other factors such as Fresnel reflection/transmission and shadowing/masking terms.

Differentiating Eq. (14) with respect to some scene parameter $\theta$ yields

$$\frac{\mathrm{d}f_\mathrm{s}(\boldsymbol{\omega}_\mathrm{i}, \boldsymbol{\omega}_\mathrm{o})}{\mathrm{d}\theta} = \frac{\mathrm{d}D(\boldsymbol{\omega}_\mathrm{h})}{\mathrm{d}\theta} f_\mathrm{s}^{(0)}(\boldsymbol{\omega}_\mathrm{i}, \boldsymbol{\omega}_\mathrm{o}) + D(\boldsymbol{\omega}_\mathrm{h}) \frac{\mathrm{d}f_\mathrm{s}^{(0)}(\boldsymbol{\omega}_\mathrm{i}, \boldsymbol{\omega}_\mathrm{o})}{\mathrm{d}\theta}, \tag{15}$$

where the NDF derivative $\mathrm{d}D/\mathrm{d}\theta$, according to the chain rule, equals

$$\frac{\mathrm{d}D}{\mathrm{d}\theta} = \frac{\mathrm{d}D}{\mathrm{d}\boldsymbol{\omega}_\mathrm{h}} \frac{\mathrm{d}\boldsymbol{\omega}_\mathrm{h}}{\mathrm{d}\theta}, \tag{16}$$

where $\mathrm{d}D/\mathrm{d}\boldsymbol{\omega}_\mathrm{h}$ can be obtained by analytically differentiating the NDF (for parametric BSDF models), and the exact form of $\mathrm{d}\boldsymbol{\omega}_\mathrm{h}/\mathrm{d}\theta$ depends on the differentiable-rendering formulation.

In Eqs. (15) and (16), $f_\mathrm{s}^{(0)}$ and $\mathrm{d}f_\mathrm{s}^{(0)}/\mathrm{d}\theta$ typically change slowly, while the normal distribution function $D$ and its derivative $\mathrm{d}D/\mathrm{d}\boldsymbol{\omega}_\mathrm{h}$ can vary rapidly for glossy materials.

*Symmetry in NDF derivatives.* Most, if not all, commonly used normal distributions, including the Beckmann and the GGX models, are symmetric. Specifically, under a local coordinate system with the surface normal aligned with the $z$-axis, it holds that

$$D([x_\mathrm{h}, y_\mathrm{h}, z_\mathrm{h}]) = D([-x_\mathrm{h}, -y_\mathrm{h}, z_\mathrm{h}]), \tag{17}$$

**ALGORITHM 2:** Differentiable path tracing with antithetic sampling

1 `diffPT`$(\boldsymbol{x}, \boldsymbol{\omega}_\mathrm{o})$
  **Input:** Surface position $\boldsymbol{x}$ and direction $\boldsymbol{\omega}_\mathrm{o}$
  **Output:** Radiance $L(\boldsymbol{x}, \boldsymbol{\omega}_\mathrm{o})$ and its gradient
2 **begin**
3   $L \leftarrow L_\mathrm{e}(\boldsymbol{x}, \boldsymbol{\omega}_\mathrm{o})$;
4   **if** *$\boldsymbol{x}$ lies on a glossy surface* **then**   `// Antithetic sampling`
5     $(\boldsymbol{\omega}_{\mathrm{i},1}, p_1, \boldsymbol{\omega}_{\mathrm{i},2}, p_2) \leftarrow$ `AntitheticBSDFSample`$(\boldsymbol{x}, \boldsymbol{\omega}_\mathrm{o})$;
6     $\boldsymbol{y}_1 \leftarrow$ `rayTrace`$(\boldsymbol{x}, \boldsymbol{\omega}_{\mathrm{i},1})$;
7     $\boldsymbol{y}_2 \leftarrow$ `rayTrace`$(\boldsymbol{x}, \boldsymbol{\omega}_{\mathrm{i},2})$;
8     $L \leftarrow [$`diffPT`$(\boldsymbol{y}_1, -\boldsymbol{\omega}_{\mathrm{i},1})\, f_\mathrm{s}^\perp(\boldsymbol{x}, \boldsymbol{\omega}_{\mathrm{i},1}, \boldsymbol{\omega}_\mathrm{o})\, +$
     `diffPT`$(\boldsymbol{y}_2, -\boldsymbol{\omega}_{\mathrm{i},2})\, f_\mathrm{s}^\perp(\boldsymbol{x}, \boldsymbol{\omega}_{\mathrm{i},2}, \boldsymbol{\omega}_\mathrm{o})]/(p_1 + p_2)$;
9   **else**   `// Standard sampling`
10     $(\boldsymbol{\omega}_\mathrm{i}, p) \leftarrow$ `BSDFSample`$(\boldsymbol{x}, \boldsymbol{\omega}_\mathrm{o})$;
11     $\boldsymbol{y} \leftarrow$ `rayTrace`$(\boldsymbol{x}, \boldsymbol{\omega}_\mathrm{i})$;
12     $L \leftarrow$ `diffPT`$(\boldsymbol{y}, -\boldsymbol{\omega}_\mathrm{i})\, f_\mathrm{s}^\perp(\boldsymbol{x}, \boldsymbol{\omega}_\mathrm{i}, \boldsymbol{\omega}_\mathrm{o})/p$;
13   **end**
  `/* (Optional) handle the boundary integral */`
14   **return** $L$;
15 **end**

for all $x_\mathrm{h}^2 + y_\mathrm{h}^2 + z_\mathrm{h}^2 = 1$ and $z_\mathrm{h} > 0$. We note that this is the case even for anisotropic normal distributions. This point symmetry (with respect to the origin) causes the derivative $\mathrm{d}D/\mathrm{d}\boldsymbol{\omega}_\mathrm{h}$ to be odd symmetric:

$$\frac{\mathrm{d}D}{\mathrm{d}\boldsymbol{\omega}_\mathrm{h}}([x_\mathrm{h}, y_\mathrm{h}, z_\mathrm{h}]) = -\frac{\mathrm{d}D}{\mathrm{d}\boldsymbol{\omega}_\mathrm{h}}([-x_\mathrm{h}, -y_\mathrm{h}, z_\mathrm{h}]). \tag{18}$$

Figure 3 visualizes the NDF and its derivative.

*BSDF antithetic sampling.* Utilizing the antithetic sampling framework presented in §2.3, we introduce **antithetic sampling of BSDFs** that exploits the symmetry of NDFs. As shown in Algorithm 1, the process starts with drawing a half-way vector $\boldsymbol{\omega}_{\mathrm{h},1}$ (Line 3) the same way as in forward rendering based on the NDF [Walter et al. 2007] or visible NDF [Heitz and d'Eon 2014]. Then, we take the antithetic sample $\boldsymbol{\omega}_{\mathrm{h},2} = [-x_\mathrm{h}, -y_\mathrm{h}, z_\mathrm{h}]$ assuming $\boldsymbol{\omega}_{\mathrm{h},1} = [x_\mathrm{h}, y_\mathrm{h}, z_\mathrm{h}]$ under a local coordinate system where the surface normal is aligned with the $z$-axis (Line 4).

With the half-way directions $\boldsymbol{\omega}_{\mathrm{h},1}$ and $\boldsymbol{\omega}_{\mathrm{h},2}$ generated, we calculate the corresponding incident directions $\boldsymbol{\omega}_{\mathrm{i},1}$ and $\boldsymbol{\omega}_{\mathrm{i},2}$ (Line 6) as well as the probability densities $p_1$ and $p_2$ (Line 7). We note that $p_1$ and $p_2$ are computed solely based on the probability density $p_\mathrm{h}$ (from which the ordinary sample $\boldsymbol{\omega}_{\mathrm{h},1}$ is drawn). To be precise, when sampling microfacet BRDFs (e.g., rough conductors) using $p_\mathrm{h} = D$, we have, for $j = 1, 2$:

$$p_j = \frac{D(\boldsymbol{\omega}_{\mathrm{h},j})}{4(\boldsymbol{\omega}_\mathrm{o} \cdot \boldsymbol{\omega}_{\mathrm{h},j})}. \tag{19}$$

In summary, our BSDF antithetic sampling offers several practical benefits:

- It can provide significant variance reduction for estimating geometric gradients when the scene is glossy.
- It is very easy to implement (that is, using a few lines of code), as demonstrated in Algorithm 1.

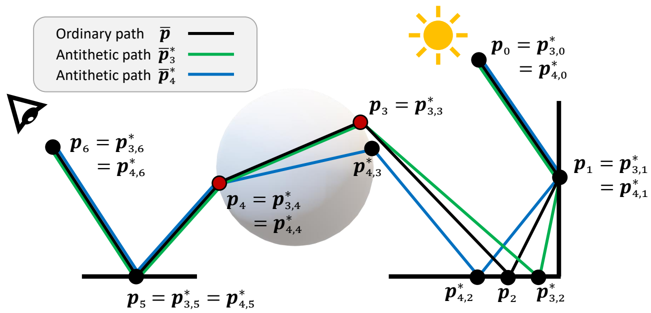



Fig. 4. **Unidirectional construction of antithetic paths** (starting from the detector): In this example, we construct antithetic paths $\bar{\boldsymbol{p}}_i^* = (\boldsymbol{p}_{i,0}^*, \boldsymbol{p}_{i,1}^*, \ldots, \boldsymbol{p}_{i,6}^*)$ for $i = 3, 4$ based on an *ordinary* one $\bar{\boldsymbol{p}} = (\boldsymbol{p}_0, \boldsymbol{p}_1, \ldots, \boldsymbol{p}_6)$ that contains two vertices $\boldsymbol{p}_3$ and $\boldsymbol{p}_4$ (shown in red) with glossy BSDFs. To obtain the first antithetic path $\bar{\boldsymbol{p}}_3^*$, we apply BSDF antithetic sampling to $\boldsymbol{p}_3$ (by taking the antithetic sample), resulting in a new incident direction that in turn yields a new vertex $\boldsymbol{p}_{3,2}^*$. To obtain the second antithetic path $\bar{\boldsymbol{p}}_4^*$, we take the antithetic BSDF sample at $\boldsymbol{p}_4$, leading to a new vertex $\boldsymbol{p}_{4,3}^*$. Since this vertex has a glossy BSDF, we continue tracing (using standard BSDF sampling) and obtain $\boldsymbol{p}_{4,2}^*$ before mering back with the ordinary path.

- Since we draw $\boldsymbol{\omega}_{\mathrm{h},1}$ the same way as in forward rendering, the resulting sampling pattern is well suited for forward rendering.
- Our BSDF antithetic sampling is not limited to microfacet BSDFs: The same algorithm can be applied to any BSDF and provide variance reduction as long as the BSDF derivative is similarly point symmetric.

## 3.2 Differentiable Rendering with BSDF Antithetic Sampling

Our BSDF antithetic sampling can be integrated into existing differentiable rendering algorithms to estimate the *interior* integrals in Eqs. (6) and (9). The *boundary* contribution can be handled by previous techniques [Li et al. 2018; Zhang et al. 2019, 2020].

In case of differentiable path tracing, as outlined in Algorithm 2, our technique draws two incident directions at each surface reflection/refraction, resulting in branching light paths.

*Next-event estimation.* The standard-sampling branch (Lines 10–12) of Algorithm 2 can be easily extended to utilize next-event estimation (NEE). Although this is also possible for our antithetic sampling (by generating $\boldsymbol{\omega}_{\mathrm{i},1}$ based on a position sample on a light source and $\boldsymbol{\omega}_{\mathrm{i},2}$ following the same symmetry), we find it unnecessary in practice since antithetic sampling is only applied when the surface is sufficiently glossy.

*Varying parameterizations.* It has been demonstrated previously that the parameterization of rendering integrals has a profound impact on the resulting differentiable rendering algorithms. Specifically, several prior works [Li et al. 2018; Zhang et al. 2019; Loubet et al. 2019; Bangaru et al. 2020] rely on the formulation of the differential rendering equation of Eq. (6). In contrast, Zhang et al. [2020] introduced the formulation of differential path integrals of Eq. (9).

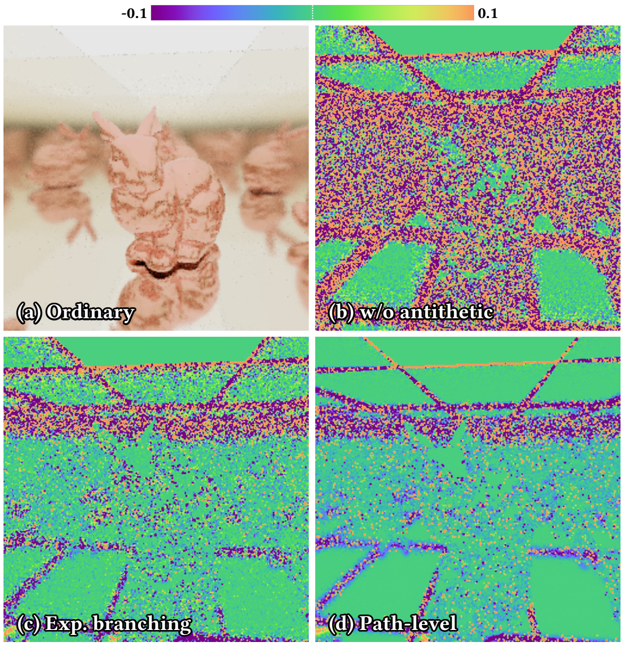



Fig. 5. Our **path-level antithetic sampling** avoids exponential branching of light paths. In this example, we show a glossy scene with a bunny inside a box lit by an area light from the above, leading to many reflections of the bunny (a). Estimating derivatives with respect to the vertical position of the light using the unidirectional path-space method [Zhang et al. 2020] without antithetic sampling leads to very high variance (b). Using our BSDF sampling (§3) with the same base algorithm, much cleaner derivative estimates can be obtained (c). However, since each light path contains many vertices that require BSDF antithetic sampling, naïvely branching at each vertex has suboptimal performance. Our unidirectional path-level antithetic sampling (§4) addresses this problem and produces even lower variance (d). The derivative estimates in (b—d) are computed in equal time.

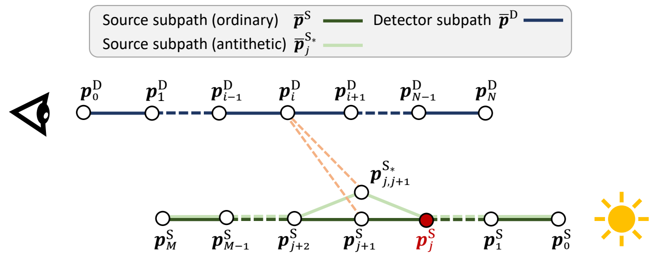



Fig. 6. **Bidirectional construction of antithetic paths:** Let $\bar{\boldsymbol{p}}^{\mathrm{S}} = (\boldsymbol{p}_0^{\mathrm{S}}, \ldots, \boldsymbol{p}_M^{\mathrm{S}})$ be a pre-generated ordinary source subpath associated with antithesis $\bar{\boldsymbol{p}}_j^{\mathrm{S}*}$, and $\bar{\boldsymbol{p}}^{\mathrm{D}} = (\boldsymbol{p}_0^{\mathrm{D}}, \ldots, \boldsymbol{p}_N^{\mathrm{D}})$ be an ordinary detector subpath. Then, for any $0 \le i \le N$ and $j < j' \le M$, connecting the $i$-th vertex $\boldsymbol{p}_i^{\mathrm{D}}$ on the ordinary detector subpath to, respectively, the $j'$-th vertex $\boldsymbol{p}_{j'}^{\mathrm{S}}$ on the ordinary source subpath and $\boldsymbol{p}_{j,j'}^{\mathrm{S}*}$ on its antithesis yields complete ordinary and antithetic paths $(\boldsymbol{p}_0^{\mathrm{S}}, \ldots, \boldsymbol{p}_{j'}^{\mathrm{S}}, \boldsymbol{p}_i^{\mathrm{D}}, \ldots, \boldsymbol{p}_0^{\mathrm{D}})$ and $(\boldsymbol{p}_{j,0}^{\mathrm{S}*}, \ldots, \boldsymbol{p}_{j,j'}^{\mathrm{S}*}, \boldsymbol{p}_i^{\mathrm{D}}, \ldots, \boldsymbol{p}_0^{\mathrm{D}})$. In this example, we have $j' = j+1$. Similarly, we can connect a vertex $\boldsymbol{p}_j^{\mathrm{S}}$ from the ordinary source subpath to a pair of vertices $\boldsymbol{p}_{i'}^{\mathrm{D}}$ and $\boldsymbol{p}_{i,i'}^{\mathrm{D}*}$ from the detector subpaths (with $i < i' \le N$) to form full ordinary-antithetic light paths.

Both formulations can be used to derive unidirectional path tracing algorithms, but with distinctive performance characteristics.

Algorithm 2 is applicable to both formulations since our BSDF antithetic sampling technique is largely independent of such parameterizations. To be precise, using different parameterizations would only affect (i) how gradients of individual variables (such as $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{\omega}_{\mathrm{i}}$) are calculated; and (ii) how the *boundary* integral is handled.

*Correlating subpaths.* By utilizing pairs of correlated samples, our BSDF antithetic sampling makes a light transport path to branch into two (Line 8 of Algorithm 2) that start with $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$, respectively. To ensure that the contributions of these two subpaths mostly cancel out when computing the gradient of $L$, we use correlated random samples to generate them. Conceptually, this is similar to computing finite differences using Monte Carlo methods.

*Path branching.* When only a small fraction of the scene is (highly) glossy, branching the light path at each vertex where BSDF antithetic sampling is performed has little impact on rendering performance. On the other hand, for mostly glossy scenes, frequent antithetic sampling can yield exponential branching of light paths and lowered performance. We will introduce a solution to this problem in §4.

*Relation to reparameterize-then-differentiate.* Another possibility to reduce the variance caused by BSDF derivatives is to reparameterize the rendering equation of Eq. (1) [Nimier-David et al. 2019; Zhao et al. 2020]. We consider our method to be largely complementary to this reparameterize-then-differentiate technique. Please see Appendix A for more discussions.

## 4 ANTITHETIC SAMPLING OF LIGHT PATHS

Based on the differential path integral formulation of Eq. (9), we further introduce a new path-level antithetic sampling technique that enjoys (i) having no exponential branching even for mostly glossy scenes; and (ii) supporting both unidirectional and bidirectional path sampling methods.

Our basic idea is to decompose derivatives of the measurement contribution—by applying the product rule—as the sum of multiple terms each of which involves one BSDF derivative. In this way, we can apply BSDF antithetic sampling once per term, avoiding exponential branching. We describe how our technique works in the following and provide detailed derivations in Appendix B.

*Unidirectional sampling.* The *unidirectional* variant of our technique starts with constructing an ordinary path $\bar{\boldsymbol{p}} = (\boldsymbol{p}_0, \boldsymbol{p}_1, \ldots, \boldsymbol{p}_N)$ using unidirectional path tracing. Assume that $\mathcal{I} \subseteq \{1, 2, \ldots, N-1\}$ denotes the indices of path vertices where BSDF antithetic sampling is needed. For each $i \in \mathcal{I}$, we accompany the same ordinary path $\bar{\boldsymbol{p}}$ with an antithesis $\bar{\boldsymbol{p}}_i^*$ generated by taking the antithetic incident direction sampled at the $i$-th vertex of the ordinary path $\bar{\boldsymbol{p}}$. To maximize the consistency between $\bar{\boldsymbol{p}}$ and $\bar{\boldsymbol{p}}_i^*$, we adapt the gradient-domain path tracing (GDPT) [Kettunen et al. 2015],

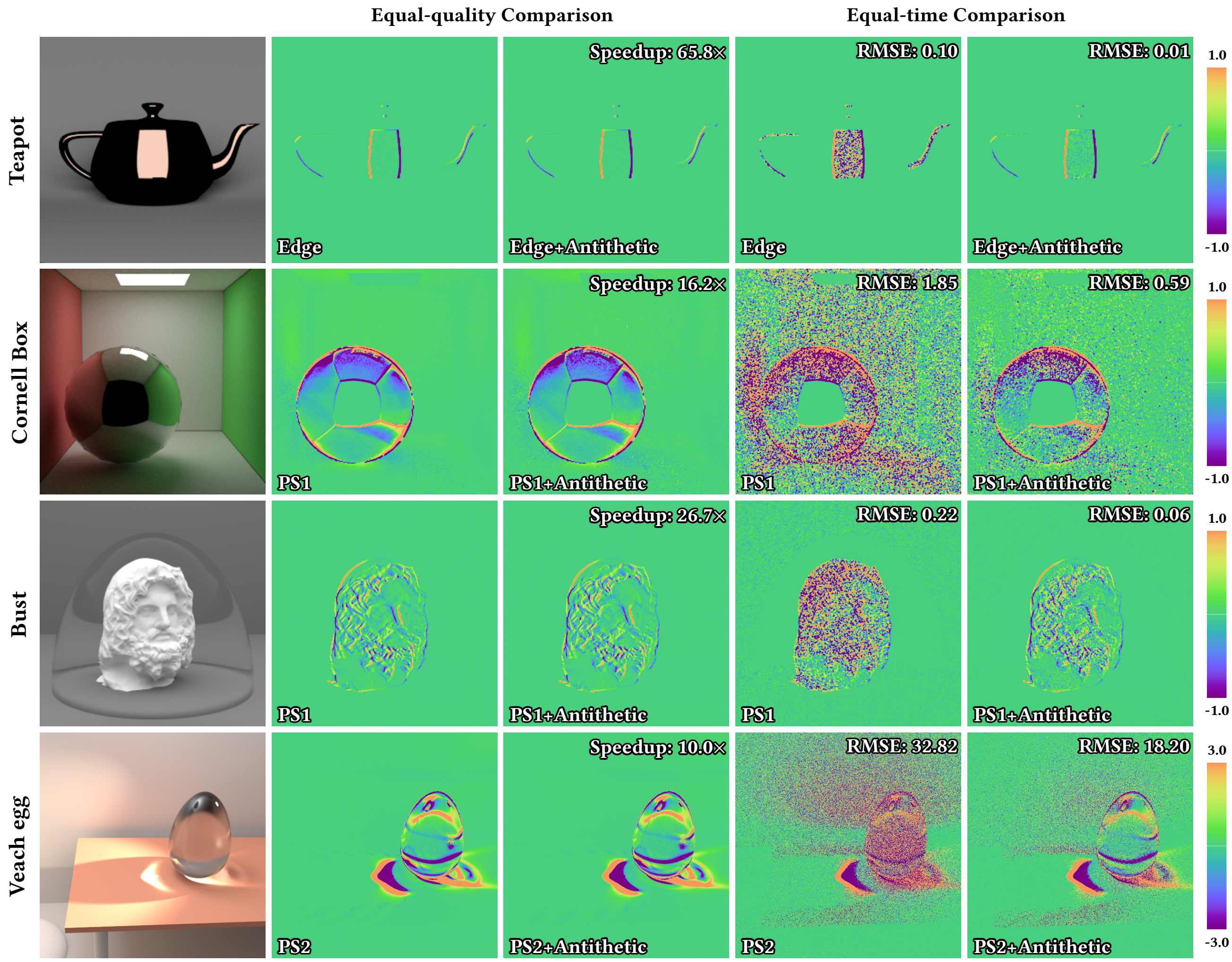



Fig. 7. **Differentiable rendering of isotropic BSDFs:** We show (the *interior* components of) derivatives estimated with and without our antithetic sampling technique using three base differentiable rendering methods: unidirectional path tracing with edge sampling (Edge) [Li et al. 2018], unidirectional (PS1) and bidirectional (PS2) path-space methods [Zhang et al. 2020]. Our technique allows significantly faster convergence for all base methods.

a forward-rendering technique. We note that the term "gradient-domain" in GDPT refers to image-space gradients that differ fundamentally from the *scene derivatives* with which differentiable rendering is concerned.

Specifically, given the ordinary path $\bar{\boldsymbol{p}} = (\boldsymbol{p}_0, \ldots, \boldsymbol{p}_N)$, our technique builds the antithetic path $\bar{\boldsymbol{p}}_i^* = (\boldsymbol{p}_{i,0}^*, \ldots, \boldsymbol{p}_{i,N}^*)$ as follows. The first $(i+1)$ vertices of the antithetic path coincide with those of the ordinary (that is, $\boldsymbol{p}_{i,j}^* = \boldsymbol{p}_j$ for all $0 \le j \le i$). The vertex $\boldsymbol{p}_{i,i+1}^*$ is obtained by tracing a ray from $\boldsymbol{p}_{i,i}^* = \boldsymbol{p}_i$ in the antithetic incident direction (given by our BSDF sampling). Then, starting from $\boldsymbol{p}_{i,i+1}^*$, we perform unidirectional path tracing with standard BSDF sampling until reaching a vertex $\boldsymbol{p}_{i,i'}^*$ with a non-glossy BSDF for some $i' \ge i+1$. Lastly, we merge the antithetic path $\bar{\boldsymbol{p}}_i^*$ back to the ordinary after $\boldsymbol{p}_{i,i'}^*$ by setting $\boldsymbol{p}_{i,k}^* = \boldsymbol{p}_k$ for all $k > i'$.

Further, for all $0 < j < N$, the vertex $\boldsymbol{p}_{i,j}^*$ of the antithetic path $\hat{\Omega}_i^*$ and the vertex $\boldsymbol{p}_j$ of the ordinary $\bar{\boldsymbol{p}}$ must be either both glossy or both rough. If this requirement is not satisfied, the antithetic path is rejected and considered to have zero contribution.

We illustrate this process in Figure 4 and demonstrate its effectiveness in Figure 5.

*Bidirectional sampling.* Our ordinary and antithetic paths can also be generated in a *bidirectional* fashion. Specifically, we build two ordinary subpaths $\bar{\boldsymbol{p}}^{\mathrm{S}} = (\boldsymbol{p}_0^{\mathrm{S}}, \ldots, \boldsymbol{p}_N^{\mathrm{S}})$ and $\bar{\boldsymbol{p}}^{\mathrm{D}} = (\boldsymbol{p}_0^{\mathrm{D}}, \ldots, \boldsymbol{p}_M^{\mathrm{D}})$ originated at the source and the detector, respectively. Assume that BSDF antithetic sampling is needed at vertices with indices $\mathcal{I}^{\mathrm{S}}$ in the source subpath and $\mathcal{I}^{\mathrm{D}}$ in the detector subpath. Then, using the aforementioned unidirectional method, we build an antithetic source

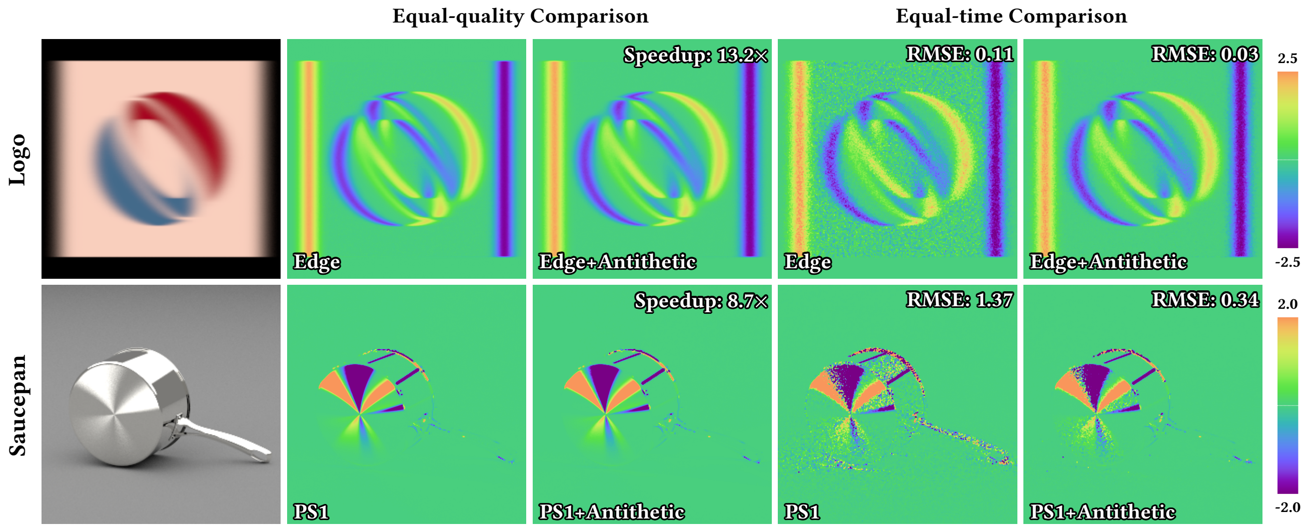



Fig. 8. **Differentiable rendering of anisotropic BRDFs:** We show (the *interior* components of) derivatives obtained with and without our antithetic sampling technique using three base differentiable rendering methods: unidirectional path tracing with edge sampling (Edge) [Li et al. 2018], unidirectional path-space method (PS1) [Zhang et al. 2020]. Our technique allows significantly faster convergence for both base methods.

subpath $\bar{\boldsymbol{p}}_i^{\mathrm{S}*}$ for each $i \in \mathcal{I}^{\mathrm{S}}$ and an antithetic detector subpath $\bar{\boldsymbol{p}}_j^{\mathrm{D}*}$ for each $j \in \mathcal{I}^{\mathrm{D}}$.

With all ordinary and antithetic subpaths constructed, we then make bidirectional connections between the source and the detector subpaths as follows. That is, for each $i \notin \mathcal{I}^{\mathrm{S}}$ and $j \notin \mathcal{I}^{\mathrm{D}}$, we connect $\boldsymbol{p}_i^{\mathrm{S}}$ in the ordinary source subpath and $\boldsymbol{p}_j^{\mathrm{D}}$ in the ordinary detector subpath, resulting in a full ordinary path $\bar{\boldsymbol{p}}_{i,j}$. To obtain the antitheses of this ordinary path, we reuse the precomputed antithetic subpaths. Please refer to Figure 6 for an illustration of this process.

## 5 RESULTS

We develop C++ implementations of the algorithms depicted in §3.2 on the CPU. In what follows, we evaluate the differentiable rendering results generated by our technique in §5.1. Additionally, we demonstrate the effectiveness of our technique by comparing inverse-rendering performance in §5.2.

### 5.1 Differentiable-Rendering Comparisons

We compare derivatives estimated with and without antithetic sampling using two base differentiable rendering algorithms: unidirectional path tracing with edge sampling [Li et al. 2018] (indicated as "Edge") and path-space differentiable rendering [Zhang et al. 2020] (with "PS1" indicating the unidirectional algorithm and "PS2" the bidirectional one). When applying antithetic sampling, we use our BSDF-level variant (discussed in §3) with "Edge" and the path-level one (presented in §4) with "PS1" and "PS2".

We only show the *interior* components of the derivatives emerging from the *interior* terms of Eqs. (6) and (9) since the estimation of *boundary* integrals is orthogonal to our work.

*Isotropic BSDFs.* In Figure 7, we show a few scenes with glossy objects depicted with isotropic microfacet BSDFs.

The **teapot** scene contains a glossy teapot lit by an area light. The derivatives are computed with respect to the rotation angle of the teapot (about its vertical axis). Using differentiable path tracing (Edge) [Li et al. 2018], our BSDF antithetic sampling offers a speedup of over 60× to produce derivative estimates with approximately the same quality. We conduct the equal-quality comparisons by: (i) generating a reference image with low noise; and (ii) computing derivative images with and without antithetic sampling progressively until the differences between the rendered results and the reference drops below a predetermined threshold. At equal time, standard BSDF sampling produces high variance in specular highlights on the teapot. When using our antithetic BSDF sampling, on the other hand, much cleaner derivative estimates can be obtained.

The rest of the examples in Figure 7 are rendered using the path-space method [Zhang et al. 2020]. The **Cornell box** scene contains a glossy sphere, and the derivatives are computed using the unidirectional algorithm (PS1) with respect to the vertical translation of the sphere. At equal quality, our path-level antithetic sampling (§4) offers a 16.2× speedup. At equal time, the estimated derivatives contain high variance without antithetic sampling. We note that even the non-glossy regions (such as the diffuse walls) suffer from high noise due to interreflections. With our technique, in contrast, the variance is greatly reduced.

The **bust** scene consists of a diffuse bust (whose 3D model is from McGuire's computer graphics archive [2017]) inside a glossy glass dome, and we compute derivatives with respect to the rotation of the bust about its vertical axis. Our antithetic sampling achieves a speedup of 26.7× to generate equal-quality derivative estimates and produces significantly lower variance at equal time.

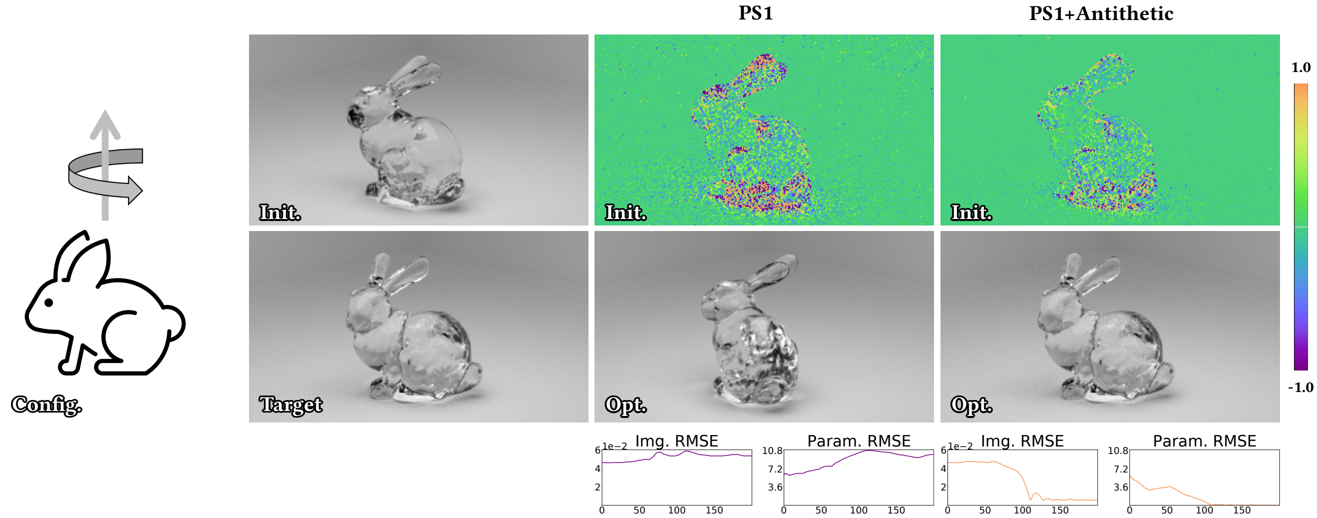



Fig. 9. **Inverse-rendering comparison (bunny):** We search for the rotation angle of the glass bunny to match the target image by minimizing the image RMSE (the plotted parameter RMSE information is used for evaluation only). Without antithetic sampling, derivatives estimated using the unidirectional path-space method (PS1) [Zhang et al. 2020] are too noisy for the optimization to converge. By adding antithetic sampling to the same base algorithm (PS1+Antithetic), much cleaner derivatives can be obtained, allowing the optimization to converge smoothly. The visualized derivatives involve both *interior* and *boundary* contributions with the latter estimated using the base method for both results.

Table 1. Performance statistics for the inverse-rendering comparisons in Figures 9, 10, and 11. The "*time*" numbers indicate average computation time per iteration for all methods—we adjust the sample count for each method so that one iteration takes approximately equal time. The experiments are conducted on a workstation with 8-core Intel® i7-7820X CPU.

| Scene | Bunny | Mug | Einstein |
|---|---|---|---|
| **# param.** | 1 | 3 | 100 |
| **# iter.** | 200 | 120 | 300 |
| **time** | 125s | 6.78s | 1.87s |

Lastly, the **Veach egg** scene shows a glass egg lit by a small spot light, creating caustics on the table. The derivatives are computed with respect to the vertical translation of the egg. Due to the complexity of light transport in this example, we estimate the derivatives using the bidirectional path-space algorithm (PS2). Our path-level antithetic sampling provides a 10× speedup and, similar to the previous examples, offers much cleaner results at equal time.

*Anisotropic BRDFs.* Our antithetic sampling technique also applies to anisotropic BSDFs, which we demonstrate in Figure 8. Similar to Figure 7, we only show contributions of the *interior* terms.

The **logo** scene shows the virtual image of a SIGGRAPH logo on an anisotropic reflector with derivatives computed with respect to the rotation angle of the logo around its horizontal axis. Our antithetic sampling technique achieves a speedup of 13.2× at equal-quality and provides considerably more accurate results in equal time.

The **saucepan** scene contains a glossy saucepan made of brushed metal lit by a small area light, resulting in characteristic anisotropic highlights on the bottom. When computing derivatives with respect to the vertical rotation of the saucepan, our technique offers a speedup of 8.7× at equal quality and much lower image RMSE in equal time.

## 5.2 Inverse-Rendering Comparisons

To further demonstrate the practical usefulness of our antithetic sampling, we compare the inverse-rendering performance using derivatives estimated with and without antithetic sampling. We use the unidirectional and bidirectional path-space algorithms [Zhang et al. 2020] (i.e., "PS1" and "PS2") as the base methods. For each example, we use the image root-mean-square error (RMSE) as the loss function and the Adam method [Kingma and Ba 2014] with identical initial configurations and learning rates to solve the inverse-rendering optimizations. We note that the parameter RMSE information is used only for evaluation (and not for optimization).

Figure 9 uses a **bunny** scene where a glass bunny model is rotated around the vertical axis (as illustrated in "Config."). Using the unidirectional algorithm as the base method, the derivative images (including both *interior* and *boundary* contributions) corresponding to the initial configuration are shown in the top row. These images are generated in equal time, and the one using antithetic sampling (i.e., PS+Antithetic) contains much lower noise. This reduced variance makes a significant difference in inverse rendering performance by allowing the inverse-rendering optimization to converge nicely. Without antithetic sampling, on the other hand, the optimization fails to converge.

In Figure 10, we show a **mug** scene modeled after the "mug" result from Zhang et al.'s work [2020]. As illustrated in "Config.", this example consists of a small area light inside a near-specular glass

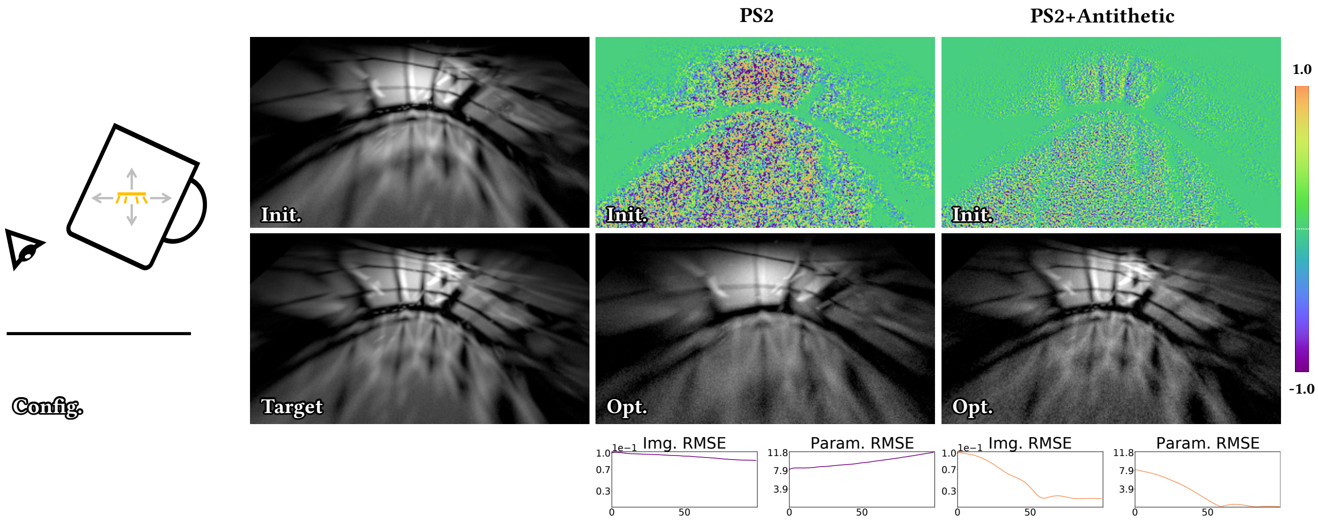



Fig. 10. **Inverse-rendering comparison (mug):** We search for the 3D location of the area light inside the glass mug to match the caustics patterns on the surface below. Without antithetic sampling, the estimated gradients are too noisy for the optimization to converge, even using the bidirectional path-space method (PS2) [Zhang et al. 2020]. Using the same base algorithm, our antithetic sampling technique (PS2+Antithetic) yields rapid convergence. Similar to Figure 9, the visualized derivatives involve both *interior* and *boundary* contributions.

mug, creating complex caustics on the surface below. Given a target image with the desired caustics pattern, we solve for the position (depicted with three parameters) of the area light. We use the bidirectional path-space algorithm as the base method for this example. Without antithetic sampling, even with bidirectional path sampling, the derivative image remains very noisy, causing the optimization to have difficulties in converging. With antithetic sampling, on the other hand, the derivative estimates become significantly cleaner, leading to much easier convergence.

Lastly, we show in Figure 11 an **Einstein** scene that contains an area light with spatially varying emission displaying a distorted photo of Einstein [Turner 1947]. The emitted light is then reflected by a glossy surface before reaching the camera. Given a target reflection that is non-distorted, we solve for the shape of this surface (parameterized using 100 variables). Without antithetic sampling, the unidirectional path-space algorithm fails to converge within 300 iterations. On the contrary, with antithetic sampling, the optimization successfully recovers the target geometry (as illustrated using the height maps).

## 6 CONCLUSION

*Limitations and future work.* Virtual scenes with strong glossy-to-glossy interactions are known to require advanced sampling methods in forward rendering. Combining these techniques with our method for physics-based differentiable rendering can allow efficient handling of challenging glossy scenes. Additionally, as our technique focuses on estimating the *interior* integrals, improving the efficiency of *boundary*-integral estimation for glossy materials is an interesting future topic.

*Conclusion.* In this paper, we introduced antithetic sampling—a classic variance reduction technique—to Monte Carlo differentiable rendering. Specifically, we develop new antithetic sampling algorithms for individual BSDFs and full light transport paths, allowing efficient estimation of geometric derivatives of glossy surfaces.

We evaluated the effectiveness of our technique by coupling it with a few recent differentiable rendering algorithms and comparing their performance with and without our antithetic sampling enabled. Additionally, we used a few inverse-rendering examples to demonstrate the benefit of reduced variance offered by our technique.

## ACKNOWLEDGMENTS

We thank the anonymous reviewers for their constructive comments. Cheng and Shuang are partially supported by NSF grant 1900927. Michael is partially supported by ELLIIT and the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

## REFERENCES

Michael Ashikhmin and Peter Shirley. 2000. An anisotropic phong BRDF model. *Journal of graphics tools* 5, 2 (2000), 25–32.
Sai Bangaru, Tzu-Mao Li, and Frédo Durand. 2020. Unbiased warped-area sampling for differentiable rendering. *ACM Trans. Graph.* 39, 6 (2020), 245:1–245:18.
Robert L Cook and Kenneth E. Torrance. 1982. A reflectance model for computer graphics. *ACM Transactions on Graphics (ToG)* 1, 1 (1982), 7–24.
John Geweke. 1988. Antithetic acceleration of Monte Carlo integration in Bayesian inference. *Journal of Econometrics* 38, 1-2 (1988), 73–89.
John Michael Hammersley and JG Mauldon. 1956. General principles of antithetic variates. In *Mathematical proceedings of the Cambridge philosophical society*, Vol. 52. Cambridge University Press, 476–481.
Eric Heitz and Eugene d'Eon. 2014. Importance sampling microfacet-based BSDFs using the distribution of visible normals. In *Computer Graphics Forum*, Vol. 33. Wiley Online Library, 103–112.
Eric Heitz, Johannes Hanika, Eugene d'Eon, and Carsten Dachsbacher. 2016. Multiple-scattering microfacet BSDFs with the Smith model. *ACM Trans. Graph.* 35, 4 (2016), 58:1–58:14.

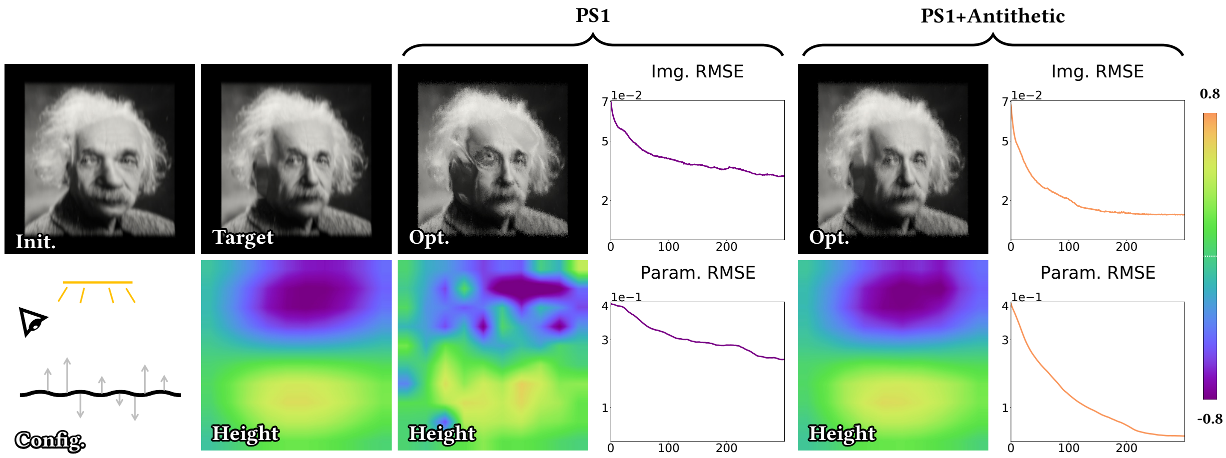



Fig. 11. **Inverse-rendering comparison (Einstein):** We search for the shape of the glossy reflector (parameterized using 100 variables) so that the reflection matches the target photo of Einstein [Turner 1947]. We visualize the target and optimized reflector geometries as height maps on the bottom row. Using the unidirectional path-space algorithm (PS1) [Zhang et al. 2020] as the base method, the gradient estimates are too noisy for the optimization to converge in 300 iterations without antithetic sampling. In contrast, gradients estimated with antithetic sampling (PS1+Antithetic) allow the optimization with otherwise identical configurations to converge nicely.


Csaba Kelemen and Laszlo Szirmay-Kalos. 2001. A microfacet based coupled specular-matte BRDF model with importance sampling. In *Eurographics short presentations*, Vol. 2. 4.

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 pages.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

Joo Ho Lee, Adrian Jarabo, Daniel S. Jeon, Diego Gutierrez, and Min H. Kim. 2018. Practical multiple scattering for rough surfaces. *ACM Trans. Graph.* 37, 6 (2018), 275:1–275:12.

Tzu-Mao Li, Miika Aittala, Frédo Durand, and Jaakko Lehtinen. 2018. Differentiable Monte Carlo ray tracing through edge sampling. *ACM Trans. Graph.* 37, 6 (2018), 222:1–222:11.

Guillaume Loubet, Nicolas Holzschuch, and Wenzel Jakob. 2019. Reparameterizing discontinuous integrands for differentiable rendering. *ACM Transactions on Graphics (TOG)* 38, 6 (2019), 1–14.

Morgan McGuire. 2017. Computer graphics archive. https://casual-effects.com/data

Merlin Nimier-David, Delio Vicini, Tizian Zeltner, and Wenzel Jakob. 2019. Mitsuba 2: a retargetable forward and inverse renderer. *ACM Transactions on Graphics (TOG)* 38, 6 (2019), 203.

Michael Oren and Shree K Nayar. 1994. Generalization of Lambert's reflectance model. In *Proceedings of the 21st annual conference on Computer graphics and interactive techniques.* 239–246.

A. Cengiz Öztireli. 2016. Integration with stochastic point processes. *ACM Trans. Graph.* 35, 5 (2016), 160:1–160:16.

Bui Tuong Phong. 1975. Illumination for computer generated pictures. *Commun. ACM* 18, 6 (1975), 311–317.

Sylvia C Pont and Jan J Koenderink. 2002. Bidirectional reflectance distribution function of specular surfaces with hemispherical pits. *JOSA A* 19, 12 (2002), 2456–2466.

Hongyu Ren, Shengjia Zhao, and Stefano Ermon. 2019. Adaptive antithetic sampling for variance reduction. In *International Conference on Machine Learning*. PMLR, 5420–5428.

Christophe Schlick. 1994. An inexpensive BRDF model for physically-based rendering. In *Computer graphics forum*, Vol. 13. Wiley Online Library, 233–246.

Gurprit Singh, Kartic Subr, David Coeurjolly, Victor Ostromoukhov, and Wojciech Jarosz. 2020. Fourier analysis of correlated Monte Carlo importance sampling. *Computer Graphics Forum* 39, 1 (2020), 7–19.

Gurprit Singh, Cengiz Öztireli, Abdalla G.M. Ahmed, David Coeurjolly, Kartic Subr, Oliver Deussen, Victor Ostromoukhov, Ravi Ramamoorthi, and Wojciech Jarosz. 2019. Analysis of sample correlations for Monte Carlo rendering. *Computer Graphics Forum* 38, 2 (2019), 473–491.

Kartic Subr, Derek Nowrouzezahrai, Wojciech Jarosz, Jan Kautz, and Kenny Mitchell. 2014. Error analysis of estimators that use combinations of stochastic sampling strategies for direct illumination. *Computer Graphics Forum* 33, 4 (2014), 93–102.

Orren Jack Turner. 1947. Einstein in 1947. https://commons.wikimedia.org/wiki/File:Albert_Einstein_Head.jpg

Bram van Ginneken, Marigo Stavridi, and Jan J Koenderink. 1998. Diffuse and specular reflectance from rough surfaces. *Applied optics* 37, 1 (1998), 130–139.

Bruce Walter, Stephen R Marschner, Hongsong Li, and Kenneth E Torrance. 2007. Microfacet models for refraction through rough surfaces. *Rendering techniques* 2007 (2007), 18th.

Gregory J Ward. 1992. Measuring and modeling anisotropic reflection. In *Proceedings of the 19th annual conference on Computer graphics and interactive techniques.* 265–272.

Mike Wu, Noah Goodman, and Stefano Ermon. 2019. Differentiable antithetic sampling for variance reduction in stochastic variational inference. In *The 22nd International Conference on Artificial Intelligence and Statistics*. PMLR, 2877–2886.

Feng Xie and Pat Hanrahan. 2018. Multiple scattering from distributions of specular V-grooves. *ACM Trans. Graph.* 37, 6 (2018), 276:1–2767:14.

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.

Cheng Zhang, Lifan Wu, Changxi Zheng, Ioannis Gkioulekas, Ravi Ramamoorthi, and Shaung Zhao. 2019. A differential theory of radiative transfer. *ACM Trans. Graph.* 38, 6 (2019), 227:1–227:16.

Shuang Zhao, Wenzel Jakob, and Tzu-Mao Li. 2020. Physics-based differentiable rendering: a comprehensive introduction. In *ACM SIGGRAPH 2020 Courses.* 14:1–14:30.


## A REPARAMETERIZE-THEN-DIFFERENTIATE

As mentioned in some prior works [Nimier-David et al. 2019; Zhao et al. 2020], reducing the variance caused by BSDF derivatives can also be achieved by reparameterizing the rendering equation of Eq. (1) as follows. Applying a change of variable from $\boldsymbol{\omega}_\mathrm{i}$ to some $\boldsymbol{u}$, we have

$$L(\boldsymbol{x}, \boldsymbol{\omega}_\mathrm{o}) = \int c(\boldsymbol{x}, \boldsymbol{\omega}_\mathrm{i}(\boldsymbol{u}, \boldsymbol{\omega}_\mathrm{o}), \boldsymbol{\omega}_\mathrm{o})\, L_\mathrm{i}(\boldsymbol{x}, \boldsymbol{\omega}_\mathrm{i}(\boldsymbol{u}, \boldsymbol{\omega}_\mathrm{o}))\, \mathrm{d}\boldsymbol{u}, \tag{20}$$

where

$$c(\boldsymbol{x}, \boldsymbol{\omega}_\mathrm{i}, \boldsymbol{\omega}_\mathrm{o}) := f_\mathrm{s}^\perp(\boldsymbol{x}, \boldsymbol{\omega}_\mathrm{i}, \boldsymbol{\omega}_\mathrm{o}) \left\| \frac{\mathrm{d}\boldsymbol{\omega}_\mathrm{i}}{\mathrm{d}\boldsymbol{u}} \right\|, \tag{21}$$

with $\|\mathrm{d}\boldsymbol{\omega}_\mathrm{i}/\mathrm{d}\boldsymbol{u}\|$ being the corresponding Jacobian determinant.

Assuming that the BSDF $f_\mathrm{s}^\perp$ can be important sampled with probability density $p(\boldsymbol{\omega}_\mathrm{i})$, the sampling process induces a mapping from the random numbers $\boldsymbol{u} \in [0,1)^2$ to $\boldsymbol{\omega}_\mathrm{i} \in \mathbb{S}^2$ with $\|\mathrm{d}\boldsymbol{\omega}_\mathrm{i}/\mathrm{d}\boldsymbol{u}\| = p(\boldsymbol{\omega}_\mathrm{i})^{-1}$. With efficient BSDF importance sampling—that is, when

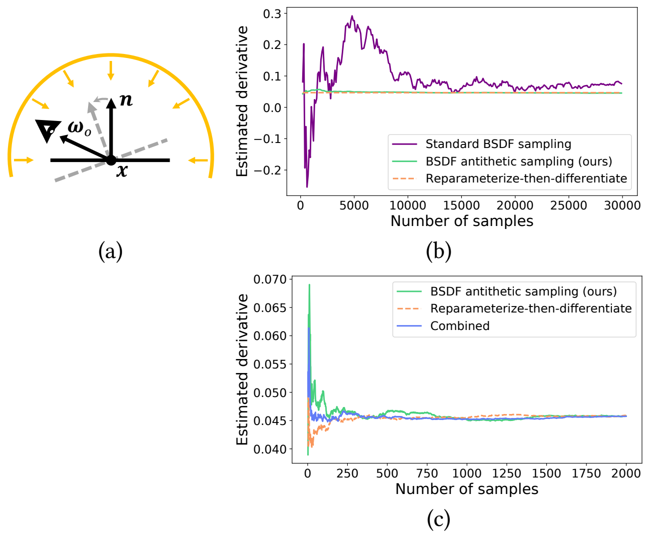



Fig. 12. **Comparison with reparameterize-then-differentiate:** This example involves a glossy plane under environmental lighting (expressed using a von Mises-Fisher function). As illustrated on the left, we differentiate the reflected radiance $L(\boldsymbol{x}, \boldsymbol{\omega}_\mathrm{o})$ with respect to the rotation angle of the surface normal $\boldsymbol{n}$ (using the formulation of differential rendering equation). Without antithetic sampling or reparameterization, the estimated derivative suffers from very high variance (as shown by the purple curve on the right). Our BSDF antithetic sampling (the green curve) and the reparameterization-then-differentiate method both (the orange dashed curve) provide significant variance reduction with the latter offering slightly lower variance (b). When using a different BSDF with less efficient importance sampling, our technique can be combined with reparameterization-then-differentiate (by exploiting the point symmetry of the function $c$) to offer even better performance (c).

$p(\boldsymbol{\omega}_\mathrm{i})$ is roughly proportional to $f_\mathrm{s}^\perp(\boldsymbol{x}, \boldsymbol{\omega}_\mathrm{i}, \boldsymbol{\omega}_\mathrm{o})$—the function $c$ becomes approximately constant (with respect to $\boldsymbol{\omega}_\mathrm{i}$). In this case, by differentiating Eq. (20), we can estimate derivatives of $L$ efficiently even if the BSDF is glossy.

*Discussion.* The reparameterize-then-differentiate technique outlined above can offer slightly better performance than our technique (see Figure 12-b) but is specialized to the forward-rendering formulation of Eq. (1). Whether it can be generalized to, for example, the path-space formulation of Eq. (2) remains an open problem. Additionally, when the mapping $\boldsymbol{\omega}_\mathrm{i}(\boldsymbol{u}, \boldsymbol{\omega}_\mathrm{o})$ is discontinuous with respect to $\boldsymbol{u}$, the *boundary* integral (and the algorithm estimating this term) may need to be modified to capture these discontinuities.

In contrast, our technique is largely parameterization-agnostic and applicable to most, if not all, differentiable-rendering formulations (and requires little changes to the base algorithms). Further, our method could be combined with the reparameterize-then-differentiate technique when, for instance, the BSDF cannot be efficiently importance sampled (as demonstrated in Figure 12-c).

We consider in-depth comparisons and advanced combinations of these methods an interesting topic for future research.

## B DERIVATIONS OF PATH-LEVEL ANTITHETIC SAMPLING

We now derive our path-level antithetic sampling technique (§4).

Let $\bar{\boldsymbol{p}} = (\boldsymbol{p}_0, \ldots, \boldsymbol{p}_N) \in \hat{\Omega}$ be some material light path and $\mathsf{X}(\bar{\boldsymbol{p}}, \theta) = (\boldsymbol{x}_0, \ldots, \boldsymbol{x}_N) \in \Omega(\theta)$ be the corresponding ordinary path with $\boldsymbol{x}_i = \mathsf{X}(\boldsymbol{p}_i, \theta)$ for $i = 0, 1, \ldots, N$. Given $\mathcal{I} \subseteq \{1, 2, \ldots, N-1\}$ consisting of vertex indices such that BSDF antithetic sampling is needed at $\boldsymbol{p}_i$ for each $i \in \mathcal{I}$, we factor out BSDF terms at these vertices in the material measurement contribution of Eq. (8), yielding:

$$\hat{f}(\bar{\boldsymbol{p}}) = \hat{f}_0(\bar{\boldsymbol{p}}) \prod_{i \in \mathcal{I}} f_{\mathrm{s},i}(\bar{\boldsymbol{p}}), \tag{22}$$

where $f_{\mathrm{s},i}(\bar{\boldsymbol{p}}) := f_\mathrm{s}(\boldsymbol{x}_{i-1} \to \boldsymbol{x}_i \to \boldsymbol{x}_{i+1})$; and $\hat{f}_0$ consists of all the other terms in $\hat{f}$ including rough BSDFs that do not need to be antithetically sampled, the geometric terms, and the Jacobian determinant in Eq. (8). Then, according to the product rule, differentiating Eq. (22) gives:

$$\frac{\mathrm{d}\hat{f}(\bar{\boldsymbol{p}})}{\mathrm{d}\theta} = \frac{\mathrm{d}\hat{f}_0(\bar{\boldsymbol{p}})}{\mathrm{d}\theta} \prod_{i \in \mathcal{I}} f_{\mathrm{s},i}(\bar{\boldsymbol{p}}) + \hat{f}_0(\bar{\boldsymbol{p}}) \sum_{i \in \mathcal{I}} \frac{\mathrm{d}f_{\mathrm{s},i}(\bar{\boldsymbol{p}})}{\mathrm{d}\theta} \prod_{j \in \mathcal{I} \setminus \{i\}} f_{\mathrm{s},j}(\bar{\boldsymbol{p}}). \tag{23}$$

It follows that the *interior* term of Eq. (9) can be rewritten as

$$\int_{\hat{\Omega}} \frac{\mathrm{d}\hat{f}(\bar{\boldsymbol{p}})}{\mathrm{d}\theta} \, \mathrm{d}\mu(\bar{\boldsymbol{p}}) = \int_{\hat{\Omega}} \frac{\mathrm{d}\hat{f}_0(\bar{\boldsymbol{p}})}{\mathrm{d}\theta} \left( \prod_{i \in \mathcal{I}} f_{\mathrm{s},i}(\bar{\boldsymbol{p}}) \right) \mathrm{d}\mu(\bar{\boldsymbol{p}}) + \sum_{i \in \mathcal{I}} \left[ \int_{\hat{\Omega}} \hat{f}_0(\bar{\boldsymbol{p}}) \frac{\mathrm{d}f_{\mathrm{s},i}(\bar{\boldsymbol{p}})}{\mathrm{d}\theta} \left( \prod_{j \in \mathcal{I} \setminus \{i\}} f_{\mathrm{s},j}(\bar{\boldsymbol{p}}) \right) \mathrm{d}\mu(\bar{\boldsymbol{p}}) \right]. \tag{24}$$

We note that the right-hand side of Eq. (24) involves multiple path integrals where the first one does not involve derivatives of glossy BSDFs and can be handled using an ordinary path $\bar{\boldsymbol{p}}$ generated with standard unidirectional or bidirectional method.

Each remaining path integral, on the other hand, involves exactly one derivative of the form $\mathrm{d}f_{\mathrm{s},i}/\mathrm{d}\theta$. We estimate this integral using $\bar{\boldsymbol{p}}$ and its *antithesis* $\bar{\boldsymbol{p}}_i^*$, as discussed in §4.