Gradient-based Approximation of Nonuniform Low-Discrepancy Samples

XIANGYU LI, Brown University, USA
EGE CIKLABAKKAL, University of Waterloo, Canada
DANIEL RITCHIE, Brown University, USA
TOSHIYA HACHISUKA, University of Waterloo, Canada

Low-discrepancy points have proven to be a useful source of samples for numerical integration via quasi-Monte Carlo integration. Low-discrepancy points are typically uniformly distributed in the unit hypercube. It is thus common to transform the input uniform points into target non-uniform samples via mapping, just like how uniform random numbers are used for generating non-uniform samples. Unfortunately, this common approach does not guarantee to carry over the low-discrepancy nature of the input points into the output non-uniform samples, giving the resulting non-uniform samples suboptimal discrepancy. The existing methods for generating low-discrepancy points focus only on uniform points, thus generating non-uniform and low-discrepancy samples has been an open problem. We introduce an optimization framework to approximate non-uniform low-discrepancy samples by numerically minimizing the discrepancy metric for non-uniform samples. We propose a differentiable numerical estimator of the generalized non-uniform discrepancy to enable gradient-based optimization. Our differentiable discrepancy estimator turns a challenging problem of generating non-uniform low-discrepancy samples into a simple numerical optimization problem. Since optimization can generate only one set of samples, we further propose a neural network architecture to encode the sample generation process, allowing us to generate multiple sets of samples from conditional targets without runtime optimization. We experimentally validate that our method reduces discrepancy below prior methods at least up to a certain sample size and consequently achieves lower integration error on numerical integration problems.

CCS Concepts: • **Computing methodologies** → **Computer graphics**.

Additional Key Words and Phrases: Quasi-Monte Carlo methods, low-discrepancy samples, numerical integration, point set generation

ACM Reference Format:
Xiangyu Li, Ege Ciklabakkal, Daniel Ritchie, and Toshiya Hachisuka. 2026. Gradient-based Approximation of Nonuniform Low-Discrepancy Samples. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers (SIGGRAPH Conference Papers '26), July 19–23, 2026, Los Angeles, CA, USA*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3799902.3811140

Authors' Contact Information: Xiangyu Li, Brown University, Providence, USA, xiangyu_li@alumni.brown.edu; Ege Ciklabakkal, University of Waterloo, Waterloo, Canada, meciklab@uwaterloo.ca; Daniel Ritchie, Brown University, Providence, USA, daniel_ritchie@brown.edu; Toshiya Hachisuka, University of Waterloo, Waterloo, Canada, toshiya.hachisuka@uwaterloo.ca.

SIGGRAPH Conference Papers '26, Los Angeles, CA, USA

ACM ISBN 979-8-4007-2554-8/26/07
https://doi.org/10.1145/3799902.3811140

1 Introduction

Quasi-Monte Carlo (QMC) integration improves upon Monte Carlo (MC) integration by using low-discrepancy samples instead of random samples. The accuracy of QMC integration can be characterized by the Koksma–Hlawka (KH) inequality [Hlawka 1961; Koksma 1942] which provides an upper bound of the error as a product of the variation of the integrand and the discrepancy of samples. Since certain low-discrepancy samples have discrepancies that go to zero as $O((\log N)^d/N)$ in a d-dimensional space for N samples, the convergence rate of QMC integration is faster than that of MC integration $O(1/\sqrt{N})$. QMC integration has therefore been used in MC rendering to improve its efficiency under certain scenarios [Christensen et al. 2018; Heinrich and Keller 1994; Shirley 1991].

Importance sampling is another popular method to improve MC integration, aiming to reduce the variance of the integrand without increasing the number of samples N. One common approach for sampling from a given probability density function (PDF) p is to transform uniformly distributed random numbers in $[0,1]^d$ into non-uniformly distributed samples through a mapping, typically using its inverse cumulative distribution function (CDF).

Since both QMC and importance sampling are effective on their own for improving MC integration, it is compelling to combine them. However, this combination does not necessarily work as well as one would hope. Firstly, a low *variance* of f/p, where f is the integrand, does not necessarily imply a low *variation* of f/p. The latter is important for QMC but is not a direct objective in importance sampling. Secondly, QMC with importance sampling *warps* uniform low-discrepancy points into non-uniform samples which may no longer be low-discrepancy. As a result, while importance sampling may reduce variance, it can potentially increase variation and discrepancy, making the resulting estimator less effective than one might think. These issues have been empirically observed in certain scenarios for QMC rendering [Christensen et al. 2018; Heinrich and Keller 1994; Shirley 1991].

We propose a framework for approximating low-discrepancy non-uniform sample sets from any given distribution (e.g., an importance distribution). Such non-uniform generation has received little attention in prior work because the known constructions of low-discrepancy samples [Ahmed et al. 2016; Doignies et al. 2024; Halton 1964; Niederreiter 1992; Sobol 1967] focus only on a uniform distribution in a unit hypercube to make it mathematically tractable. Even prior works that numerically optimize discrepancy [Huffel et al. 2025; Rusch et al. 2024] have dealt only with uniformly distributed points so far. To this end, we employ a generalized KH inequality [Aistleitner and Dick 2015; Hlawka and Mück 1972a,b] and propose a *differentiable* numerical estimator for it. Our differentiable estimator leads to a gradient-based optimization problem

where the solution is a set of approximately low-discrepancy *and* non-uniformly distributed samples for a given target distribution. Since we directly minimize the upper bound of the generalized KH inequality, QMC integration with our samples can achieve lower error than uniform QMC followed by mapping to non-uniform samples. To support sampling from conditional distributions without runtime optimization, we also propose a neural network model that encodes the sample-generation process via our optimization. Once trained, our model can generate samples from an arbitrary conditional distribution without optimization at runtime. Beyond confirming that the discrepancy is reduced up to a certain size, we also demonstrate that the numerical integration error in QMC integration and rendering is also reduced in practice. To summarize, our contributions are threefold: a differentiable discrepancy estimator for non-uniform samples with respect to a given target distribution; a gradient-based optimization method to approximate low-discrepancy non-uniform samples; and a neural encoding of point set generation for a given conditional distribution.

2 Background

Low-discrepancy points are commonly uniformly distributed points in a unit hypercube with provably low-discrepancies [Halton 1964; Niederreiter 1992; Sobol 1967]. Since the input random numbers in MC integration are also uniformly distributed in the unit hypercube, QMC integration is often done by simply replacing them with low-discrepancy points. For the N input QMC points $u_1, u_2, ..., u_N$ in the unit hypercube and the CDF $P(x)$ for a given PDF $p(x)$, the N-sample (Q)MC estimator F_N with inverse CDF sampling is $F_N = \frac{1}{N}\sum_{i=1}^{N}\frac{f(P^{-1}(u_i))}{p(P^{-1}(u_i))} = \frac{1}{N}\sum_{i=1}^{N} g(u_i)$ where we have $x_i = P^{-1}(u_i)$ by the property of the inverse CDF $P^{-1}(u)$, and we wrote $g(u) = f(P^{-1}(u))/p(P^{-1}(u))$ to show that it can be interpreted as an integral of $g(u)$ with uniformly distributed samples u_i. Even for cases where the inverse CDF $P^{-1}(u)$ from a given PDF $p(x)$ is not available analytically, one could still use numerical solutions such as neural networks [Dinh et al. 2015; Müller et al. 2019] to approximate $P^{-1}(u)$. MC integration uses uniformly distributed random numbers for u_i, hence QMC is interpreted as replacing random numbers with low-discrepancy points. Note, however, that the low-discrepancy points u_i are *warped* into x_i by $x_i = P^{-1}(u_i)$ and x_i is generally not uniformly distributed in the sampling domain.

While this mapping of u_i to samples x_i is commonly done, the discrepancy of the resulting non-uniform samples x_i are known to be altered [Aistleitner and Dick 2015; Chelson 1976]. In other words, even when the inputs u_i are low-discrepancy points, it does not mean that the resulting samples x_i also have a low discrepancy to the same degree. One practical consequence is that the KH inequality for x_i thus does not actually warrant improvement with QMC over MC under this common approach. While a rigorous mathematical analysis of when and how this issue occurs is not our scope, this issue has been empirically and repeatedly observed in its application to rendering [Christensen et al. 2018; Heinrich and Keller 1994; Shirley 1991], where warping with $P^{-1}(u)$ is essential to generate path samples. There is currently no general approach that is capable of generating low-discrepancy non-uniform samples without mapping. We propose a first take on this lingering problem in QMC; our method indeed numerically minimizes the discrepancy of non-uniform samples x_i and approximately generates such x_i.

2.1 Discrepancy and QMC Integration

Let us briefly summarize some properties of low-discrepancy samples and their connection to integration error bounds. Let $u_1, ..., u_N$ now be a set of uniformly distributed (not necessarily low-discrepancy) points in the d-dimensional unit hypercube $[0,1]^d$. The discrepancy D_N of this set of points is defined as

$$D_N = \sup_{A\subset[0,1]^d}\left|\frac{1}{N}\sum_{i=1}^{N}\mathbb{1}_A(u_i) - \lambda(A)\right| \tag{1}$$

where A is an axis-aligned box, $\mathbb{1}_A(u_i)$ is the indicator function of A, and λ is the d-dimensional Lebesgue measure. When the supremum is over axis-parallel boxes which are anchored at the origin, D_N defines the commonly-used *star-discrepancy* D_N^*. When the dependence on N is clear from context, we simply write D^* for brevity. The term "low" in low-discrepancy samples may not have a universally agreed definition. In this paper, we measure numerically measure discrepancy and compare our samples against other commonly known low-discrepancy samples. This definition is commonly used in direct optimization of discrepancy [Clément et al. 2025; Huffel et al. 2025; Rusch et al. 2024]. We only claim our samples have smaller discrepancy up to the sizes we demonstrate, and they may not satisfy a certain rate like $O(\log N^d/N)$ for any size N.

KH Inequality. The KH inequality [Hlawka 1961; Koksma 1942] states that, for a set $u_1, \ldots, u_N$ and a function g which has bounded variation $\mathrm{Var}_{HK}(g)$ in the Hardy and Krause sense, we have

$$\left|\frac{1}{N}\sum_{i=1}^{N} g(u_i) - \int_{[0,1]^d} g(u)du\right| \le D^* \cdot \mathrm{Var}_{HK}(g). \tag{2}$$

This inequality serves as an error bound for QMC integration based on the star-discrepancy D^*. The existing low-discrepancy sequences [Halton 1964; Niederreiter 1992; Sobol 1967] are designed to reduce discrepancies D^* that also converge to zero at the rate of $O((\log N)^d/N)$ for N points. Using those sequences instead of random points generally improves the convergence rate from $O(1/\sqrt{N})$ of MC integration to $O((\log N)^d/N)$ in QMC integration, as supported by the KH inequality.

One important assumption made above is that points $u_1, ..., u_N$ are *uniformly* distributed, thus we needed to consider the integrand $g(u) = f(P^{-1}(u))/p(P^{-1}(u))$ instead of $f(x)$. Consequently, bounded variation on the original integrand $\mathrm{Var}_{HK}(f)$ does not necessarily imply bounded variation on this new integrand $\mathrm{Var}_{HK}(g)$. For example, when mapping $P^{-1}(u)$ is discontinuous, $g(u)$ can also become discontinuous and variation $\mathrm{Var}_{HK}(g)$ becomes *infinity* even when $\mathrm{Var}_{HK}(f)$ is finite. While such a case does not immediately invalidate QMC integration, the improvement is not guaranteed by the KH inequality [Owen 2023]. It is thus desirable to extend the concept of discrepancy and the KH inequality to general measures to directly handle f with non-uniformly distributed samples x_i.

Generalized KH Inequality. Fortunately, the KH inequality and discrepancy have already been generalized to non-uniform samples [Aistleitner and Dick 2015; Chelson 1976; Hlawka and Mück 1972a,b]. Let μ be a normalized *Borel* measure on $[0,1]^d$. The star-discrepancy with respect to μ of a point set $x_1, \ldots, x_N$ is

$$D^*(\mu) = \sup_{A \subset [0,1]^d} \left| \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_A(x_i) - \mu(A) \right| \tag{3}$$

and the generalized KH inequality for measure μ is

$$\left| \frac{1}{N} \sum_{i=1}^{N} \frac{f(x_i)}{q(x_i)} - \int_{[0,1]^d} f(x)dx \right| \leq D^*(\mu_q) \cdot \mathrm{Var}_{HK}\left(\frac{f}{q}\right) \tag{4}$$

where q is the density of the normalized Borel measure μ_q on $[0,1]^d$, which can be thought of as the importance sampling PDF in this paper. As in importance sampling for MC integration, the bound can be tightened by choosing q roughly proportional to f, and in this case this reduces the HK variation of f/q rather than the variance. This inequality directly shows how a non-uniform distribution q can be used to bound the error of numerical integration with $x_i \propto q$, which is akin to importance sampling with q. In the special case where q is constant, the inequality reduces to the standard KH inequality. Compared to Equation 2, Equation 4 is independent of the mapping $x = P^{-1}(u)$, thus avoiding any issues introduced by the mapping. The key question we tackle is to generate low-discrepancy non-uniform samples $x_i \sim q$ that directly minimize the generalized non-uniform discrepancy. Prior work on numerical optimization of discrepancy focus only on uniform samples (points) u_i [Huffel et al. 2025; Rusch et al. 2024] and we are the first to address non-uniform cases. For uniform cases, one successful approach is to optimize parametric generators of low-discrepancy samples [Ahmed et al. 2016; Doignies et al. 2024]. While this approach allows either empirical or theoretical guarantee of a low-discrepancy nature of samples regardless of the outcome of optimization, it is not applicable to non-uniform cases. We also aim to directly optimize discrepancy, not other indirect metrics like minimum distance among samples (e.g., Poisson disk samples) because prior work show that samples sets that optimize such a different metric do not necessarily show favorable scaling in numerical integration [Ahmed et al. 2016; Christensen et al. 2018; Pilleboue et al. 2015]. Finding q that effectively reduces the HK variation of f/q given f is not our focus.

Other Discrepancy Measures. While popular, the star discrepancy D^* is just one special form of the definition of discrepancy. Another common form, the L_2-discrepancy $D_2(\mu)$ [Chen and Song 2016; Hickernell 1998], is given by

$$D_2(\mu) = \left(\int_{[0,1]^d} \left(\frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_A(x_i) \; - \; \mu(A) \right)^2 \mathrm{d}a \right)^{\frac{1}{2}}, \tag{5}$$

where $A = [0, a)$ denotes the axis-aligned box anchored at the origin. L_2-discrepancy frequently appears in numerical optimization of discrepancy because it is computationally more feasible to evaluate compared to the star discrepancy [Rusch et al. 2024], which we also utilize in our method. The L_2-discrepancy can be connected back to an inequality similar to the KH inequality (Eqn. (4)) in the form of

$$\left| \frac{1}{N} \sum_{i=1}^{N} \frac{f(x_i)}{q(x_i)} - \int_{[0,1]^d} f(x)dx \right| \leq D_2(\mu_q) \cdot V_2\left(\frac{f}{q}\right) \tag{6}$$

where $V_2(f)$ uses a L_2 measure [Chen and Song 2016].

Beyond points in a unit hypercube, discrepancy can also be defined on the spherical domain, which is the natural setting for directions in 3D. The spherical analog of anchored boxes in $[0,1]^d$ is the *spherical cap*, defined as the intersection of the unit sphere $\mathbb{S}^2$ with a closed half-space [Schmidt 1969a,b]. We parametrize a cap by its center direction $\omega \in \mathbb{S}^2$ and angular radius $\theta \in [0, \pi]$, and write

$$C(\omega, \theta) = \{x \in \mathbb{S}^2 : \langle \omega, x \rangle \geq \cos\theta\}. \tag{7}$$

Similarly to our previous definitions for the hypercube, the spherical cap discrepancy with respect to a Borel measure μ is given by

$$D_{\mathrm{cap}}(\mu) = \sup_{\omega \in \mathbb{S}^2,\ \theta \in [0,\pi]} \left| \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{C(\omega,\theta)}(x_i) - \mu(C(\omega,\theta)) \right|, \tag{8}$$

where $\mathbb{1}_{C(\omega,\theta)}$ denotes the indicator function of the spherical cap. To express the L_2 spherical cap discrepancy, let $\mathcal{C}$ denote the family of all spherical caps on $\mathbb{S}^2$, $\mathcal{C} = \{\, C(\omega,\theta) : \omega \in \mathbb{S}^2,\ \theta \in [0,\pi] \,\}$, and let ν be the probability measure on $\mathcal{C}$ induced by sampling ω uniformly on $\mathbb{S}^2$ and θ uniformly in $[0,\pi]$. We then have

$$D_{\mathrm{cap},2}(\mu) = \left(\int_{\mathcal{C}} \left(\frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{C(\omega,\theta)}(\mathbf{x}_i) - \mu(C) \right)^2 d\nu(C) \right)^{\frac{1}{2}}, \tag{9}$$

where the integration is over all spherical caps $C \in \mathcal{C}$. Analytical generation of low-discrepancy 3D directions remains an open problem and the common approach is to parameterize the spherical domain as a 2D space and warp 2D points into 3D directions, with an expectation that it retains the low-discrepancy nature of 2D points. While not analytical, our method enables this generation by directly minimizing spherical discrepancy and we show it achieves lower spherical discrepancy than this common approach.

3 Gradient-based Discrepancy Optimization

While recent work has proposed numerically optimizing the analytical discrepancy [Huffel et al. 2025; Rusch et al. 2024] using neural approaches to approximate low-discrepancy points, these methods are restricted to a uniform target distribution for a few reasons. First of all, the discrepancy is evaluated with respect to the Lebesgue measure (as opposed to the Borel measure μ_q in Equation 3) and therefore does not account for non-uniform densities. They also rely on the existence of the analytical solution of the integral for discrepancy, which is usually not available for non-uniform densities. Note that the inequality in Equation 3 alone does not dictate *how to generate* such points, thus one would still need to design an algorithm to generate points. We propose a numerical approach based on a differentiable discrepancy estimator and gradient descent to approximate low-discrepancy non-uniform samples. To this end, we introduce a differentiable modification of average squared-discrepancy [Clément et al. 2025] generalized for non-uniform distributions. Our estimator supports any non-uniform distribution as long as pointwise evaluation of the density is available.

3.1 Differentiable Discrepancy Integrand

One challenge in designing such an estimator is that the L_2 discrepancy (Eqn. (5)) is non-differentiable with respect to the point locations x_i due to its indicator function. Prior numerical optimization methods [Huffel et al. 2025; Rusch et al. 2024] do not face this issue since they use the analytical integrals that are readily differentiable. Among various potential approaches to handle similar cases, inspired by the success of a similar idea in differentiable rasterization [Kato et al. 2018], we found that simple mollification based on *signed distance* to boxes (or spherical caps) works well.

Let us denote $d_{(A,x_i)}$ as the signed distance from a point x_i to the box A. The distance is positive inside the box and negative outside. We can reinterpret the indicator function with the unit step function H as $\mathbb{1}_A(x_i) = H\left(d_{(A,x_i)}\right)$ which leads to an approximation based on a scaled sigmoid function as $\mathbb{1}_A(x_i) \approx \sigma\left(\alpha d_{(A,x_i)}\right) = \frac{1}{1+e^{-\alpha d_{(A,x_i)}}}$ where α is a parameter to scale the slope of the sigmoid. This approximation is differentiable and therefore allows us to optimize point sets by minimizing the corresponding approximate discrepancy for a given target Borel measure μ. During optimization, we use a surrogate-gradient method where the forward pass uses the non-differentiable indicator function and the backward pass uses the gradient of the sigmoid approximation. We found that starting from a small value and linearly increasing to a large value worked well across all of our experiments.

3.2 Monte Carlo Discrepancy Estimator

One might consider numerically estimating Equation 5 by simply averaging the local discrepancy (i.e., the integrand) in the unit hypercube over M randomly sampled boxes anchored at the origin in a manner similar to MC integration:

$$D_2(\mu) \approx \left(\frac{1}{M} \sum_{m=1}^{M} \left(\frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{A_m}(x_i) \; - \; \mu(A_m) \right)^2 \right)^{\frac{1}{2}}. \tag{10}$$

Based on numerical experiments, we found that the emphasis on the origin of anchored boxes has undesirable effect of giving more importance to points with small coordinate values [Clément et al. 2025; Clément et al. 2025] when combined with numerical optimization. Fig. 1a highlights this behavior for uniform density. The 2D point set is obtained by optimizing (our mollified variant of) Eqn. (10) starting from Sobol points. The optimization fails to achieve uniform coverage in the local regions near the corners other than the origin, in particular near $\mathbf{1} = (1, 1, \ldots, 1)$. This suboptimal placement of points near the corners, in turn, degrades the overall distribution of the point set.

We found that the *average squared-discrepancy* D_2^{asd} [Clément et al. 2025] is a better fit for numerical optimization based on Eqn. (10), which is a variant of the L_2 discrepancy that treats each vertex of $[0,1]^d$ as a box anchor and averages the corresponding D_2 values. Note that Clément et al. [2025] do not propose to estimate D_2^{asd} by sampling boxes as in our method, but rather derived its analytical integral for uniformly distributed points. To sample multi-corner boxes, we draw $u \in [0,1]^d$ and form the 2^d axis-aligned boxes spanned by u and each corner $c \in \{0,1\}^d$. Minimizing

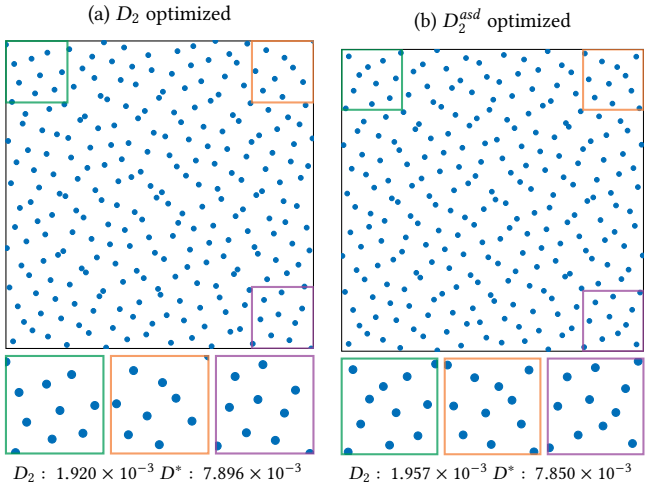

Fig. 1. Comparison of point sets optimized for uniform density by minimizing D_2 and D_2^{asd}. (a) Using origin-anchored boxes overemphasizes the origin, leading to less uniform coverage near the other corners of the domain. (b) Using all corners as anchors and averaging their D_2 discrepancies yields more uniform coverage across the entire domain.

an estimate of D_2^{asd} based on Eqn. (10) addresses the aforementioned artifacts while achieving similar L_2 and star discrepancy values (Fig. 1b).

3.3 Area under Non-uniform Measures

To evaluate Equation 10 for non-uniform cases, for each box A_m, we compute its area under the non-uniform measure, $\mu(A_m)$. A closed-form expression is usually unavailable, so we estimate it with another MC integration that samples K points $s_1, \ldots, s_K$ according to another PDF p and approximates $\mu(A_m)$ as

$$\mu(A_m) = \int_{A_m} d\mu(s) \; \approx \frac{1}{K} \sum_{i=1}^{K} \frac{\mathbb{1}_{A_m}(s_i)\, q(s_i)}{p(s_i)}. \tag{11}$$

where $q(s_i)$ is the density at the point s_i. For efficient optimization, the proposal density p is ideally chosen proportional to the target density q. In our numerical results, we used $p \propto q$ whenever possible. It is not strictly required to have $p \propto q$, and one can simply use a uniform distribution for p. To amortize the sampling cost across all boxes, we also generate a large set of samples that is shared among all boxes. This sharing of samples is again optional, and we found that individually sampling each box can sometimes be beneficial when sampled boxes are small.

3.4 Non-uniform Discrepancy Optimization

By replacing the indicator function with a sigmoid and computing the non-uniform measure $\mu(A_m)$ via yet another MC integration as in Equation 11, one can obtain a *differentiable* discrepancy estimator that can be used with gradient-based methods. While one may argue that we are not directly minimizing the star discrepancy, the star discrepancy is not easily differentiable due to the supremum operator. Even though no theoretical proof has been given so far, it has been empirically observed that minimizing the L_2 discrepancy

usually achieves a smaller star discrepancy than analytical constructions (e.g., Sobol) in the uniform case [Rusch et al. 2024]. As we will demonstrate later in our results, our generalized $D_2^{asd}(\mu)$ also exhibits a similar behavior even for *non-uniform cases*.

Optimizing a point set is now a simple process starting from a randomly initialized point set (which themselves are optionally from another low-discrepancy sequence) and minimizing its estimated discrepancy via gradient descent. The other input and parameters of our algorithm are the number of sampled boxes M, the number of space samples K used for non-uniform area estimation, the density q of the target normalized Borel measure μ, and optimization hyperparameters such as the sigmoid scale α and the learning rate β, which can be chosen adaptively. Further details on the hyperparameter choices are provided in the supplementary material.

Our optimization gradually moves the points to follow the target density. However, due to the numerical nature of optimization, we observed that sometimes samples can get stuck on the edges or at the corners of the domain or fail to move away from zero density regions. To address these issues, one can employ optional mapping of the points in the unit hypercube to the target density $q(x)$ using its inverse-CDF transform (which can often be already available in MC integration) as an additional step in the optimization. In this parameterization, no point can land in a zero-density region, and since the optimization explicitly accounts for this mapping, we obtain approximately low-discrepancy point sets with respect to the non-uniform target density. We emphasize that this use of mapping is entirely optional in our optimization and discrepancy is estimated for samples *after* mapping in any case. In cases where an explicit inverse CDF is not available or is not sufficiently smooth, tweaking of training hyperparameters together with a toroidal warping of the domain can also substantially mitigate the issues. We provide an example illustrating this behavior in the supplementary material.

3.5 Low-discrepancy 3D Directions

Sampling 3D directions is commonly required in MC rendering. The optimization domain becomes a unit sphere, so that the discrepancy is estimated directly on the *spherical* domain. The optimization framework from the previous sections remains mostly the same, except that instead of minimizing the $D_2^{asd}(\mu)$ discrepancy using anchored boxes, we minimize the $D_{\text{cap},2}(\mu)$ using spherical caps. We follow the spherical cap discrepancy [Schmidt 1969a,b] without a fixed anchor point. We consider all spherical caps and estimate the spherical cap discrepancy using randomly sampled caps, parameterized by a center direction and an angular radius. To mollify the indicator function, we compute signed angular distances to spherical cap boundaries. For a cap $C(\omega, \theta)$, the signed distance of a point x_i is given by $d_{(C,x_i)} = \theta - \arccos(\omega \cdot x_i)$. We represent each direction on the sphere by a 2D point in the unit square $(u, v) \in [0, 1]^2$, which is subsequently mapped to a direction via an octahedral area-preserving mapping [Clarberg 2008]. We emphasize again that this mapping is not critical in our optimization framework. Regardless of mapping, our framework always strives to optimize discrepancy of directions directly defined in the spherical domain. The mapping is there to help keep the resulting samples on the spherical domain in a manner that is useful for optimization. Different mapping will result in different performance in optimization, but further study along this line is left for future work.

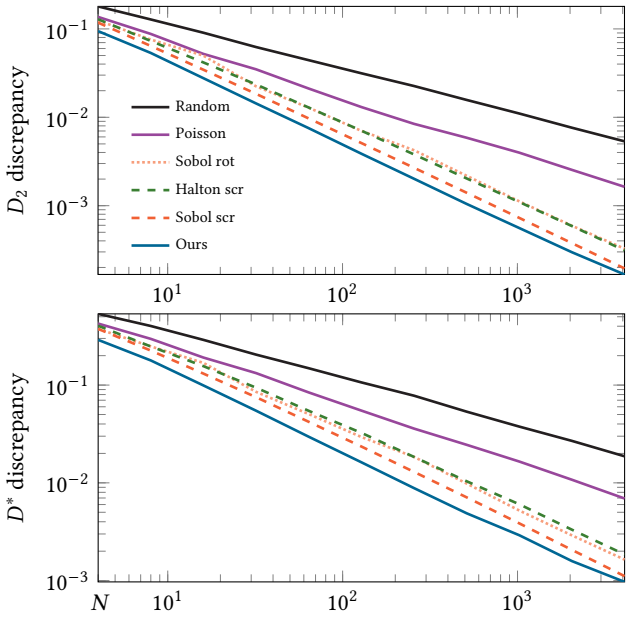

Fig. 2. L_2 and star discrepancy values (D_2 and D^*) for different sizes N of point set under a 2D uniform distribution. We compare against random, Poisson disk, scrambled Halton and rotated or scrambled Sobol. Our optimization starts from scrambled Sobol points and successfully reduces both discrepancy values beyond the most common baseline of scrambled Sobol despite optimizing only a differentiable variant of the L_2 discrepancy.

4 Low-discrepancy Samples from Conditional Targets

Any point set optimization approach, including ours, must run a new optimization whenever a new target distribution is given. While this approach is still practical if the target distribution is known in advance (e.g., sampling a given environment map), it becomes impractical when sampling from a *conditional* distribution whose conditions vary at runtime, so distributions are not fixed apriori. For example, importance-sampling PDFs for BRDFs are commonly conditional on the view direction. In this case, one would need to either rerun an optimization process for a given view direction at runtime or precompute multiple point sets for sufficiently many discretized view directions, both of which are impractical in terms of computational or storage cost. To support such a scenario, we train a network to encode the point generation process by minimizing on the same discrepancy loss over an entire distribution for multiple point sets dictated by conditional distributions, rather than optimizing a single set only. Our network takes parameters γ (e.g., the view direction) that encodes conditional variables as an input. The network additionally takes a random "offset" vector to capture the fact that each point set is just one instance among all the possible point sets with approximately the same discrepancy. Just like some related work on encoding low-discrepancy point sets [Leimkühler et al. 2019], the training process does not require running separate optimizations to generate the reference outputs for the training data

set (in contrast to supervised approaches [Doignies et al. 2023]). Instead, we train the network so that its output directly minimizes our discrepancy loss, which can be seen as a form of *weakly supervised* learning. The details are in the supplementary material.

5 Evaluation

Uniform Validation. We empirically validate that our optimization approximates low-discrepancy point sets for the uniform distribution. We compare against several common analytical low-discrepancy baselines. Fig. 2 and Fig. 3 report the D_2 and D^* discrepancies for point sets with sizes that are powers of two ($N = 4, 8, \ldots, 4096$) for both 2D points and 3D directions. As baselines, we include Cranley–Patterson rotated Sobol point sets as well as scrambled Halton and Sobol sequences [Owen 2003], and we report discrepancies averaged over 128 realizations. We additionally show the discrepancy of Poisson disk samples [Yuksel 2015]. We compute exact discrepancy values for the uniform 2D cases, and otherwise estimate them accurately. To generate 3D directions, we use the octahedral square-to-sphere mapping to transform 2D points into 3D directions. Our method starts from scrambled Sobol point sets, and we report the averaged result over 16 sets. Although our method optimizes a mollified variant of D_2^{asd}, we report D_2 together with D^* as these are more standard discrepancy measures. We did not use the analytical form of discrepancies (e.g., Warnock's L_2 formula) for even uniform points since it would not generalize to non-uniform cases. The results confirm that our optimization can indeed approximate low-discrepancy point sets, comparable to analytical baselines such as Halton and Sobol. Improvement over scrambled Sobol might be attributed to our method producing point sets rather than point sequences, thus we do not conclude that our method is better than the baselines. Being a direct optimization of discrepancy, it is challenging to optimize a large sample size as was demonstrated in parametric optimization of uniform sequences [Ahmed et al. 2016]. For instance, non-gradient approach reports optimization of 100 uniform samples in 11 hours [Clément et al. 2025]. Our method can optimize 2048 directions in 75 seconds in contrast to 8 directions in prior work [Rusch et al. 2024]. While theoretical analysis is left for future work, our optimization cost should scale at least $O(N^2)$, speculating from the uniform cases.

Non-Uniform Distributions. The key strength of our approach is that the same optimization procedure can approximate low-discrepancy samples for any given non-uniform distribution, as long as the target density can be evaluated pointwise. Fig. 5 compares our optimization against the low-discrepancy baselines in 2D across several non-uniform target functions. The baselines here uses low-discrepancy points in the unit square and apply the inverse transform associated with the target density, whereas our optimization directly produces point sets that follow the target density. We use the standard marginal-then-conditional inverse CDF construction for most cases, except for the 2D image example in the rightmost column, where we apply hierarchical warping [Clarberg et al. 2005], which warps the domain coherent manner and thus likely better preserves the low-discrepancy structure of the input point set [Pharr 2019]. Fig. 6 shows the corresponding results on the 3D spherical domain with non-uniform directional samples. Across

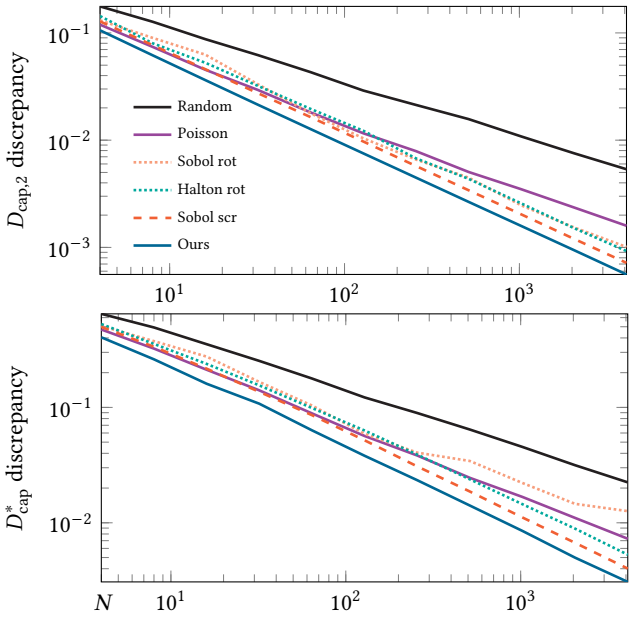

Fig. 3. L_2 and star discrepancy values ($D_{\text{cap},2}$ and D^*_{cap}) for different sizes N of point set under a uniform 3D directional distribution. The baseline methods need additional mapping from 2D points to 3D directions in this example. Our optimization starts from scambled Sobol and directly optimize the spherical discrepancy $D_{\text{cap},2}$.

all experiments, our optimized point sets outperform the baselines in terms of both L_2 and star discrepancies. The results affirm our claim that this common approach of warping QMC point sets into non-uniformly distributed samples leads to suboptimal non-uniform discrepancy metrics, leaving room for improvement. Additionally, we empirically confirmed that the discrepancy improvements indeed come from directly optimizing the non-uniform metric, as opposed to optimizing for uniform discrepancy and then mapping the resulting samples. In other words, improvements in Fig. 5 and Fig. 6 are *not* due to our lower uniform discrepancy values in Fig. 2 and Fig. 3. This additional result is in the supplementary material.

Numerical Integration Tests. While we confirmed that discrepancy values are reduced for both uniform and non-uniform cases, it is known that low discrepancy values do not necessarily transfer to more accurate numerical integration [Christensen et al. 2018], and thus conclusions drawn based only on discrepancy values can be misleading. We thus turn our attention to actual numerical integration error when point sets are used for numerical integration.

In addition to the baselines considered before, here we also compare against non-discrepancy-based methods, *blue-noise sampling* and *optimal transport*, which aim to achieve similar goals of generating well-distributed samples, albeit under completely different designs. Blue noise (BN) sampling produces a point set with a power spectrum with little energy in the low-frequency region, which is known to affect the variance and the convergence rate in integration [Singh et al. 2020]. Optimal transport optimizes a *transport map*, which can be connected to the Wasserstein distance and defines the error bound in integration [Paulin et al. 2020]. To this end, we compare against sliced optimal transport sampling (SOT) [Paulin

et al. 2020] and Gaussian blue noise (GBN) [Ahmed et al. 2022] as representatives of those two classes of methods. Our method, blue noise, and optimal transport are all based on optimization-based approaches and produce a sample set (not a sequence).

Fig. 7 summarizes the results for both uniform and non-uniform distributions for 2D point and 3D directional samples. For 2D point sets, we first evaluate the integration error of uniform samples on the binary quarter-disc integrand $f(x, y) = \mathbb{1}\left[x^2 + y^2 < \frac{2}{\pi}\right]$. Our optimized points are well suited for numerical integration and exhibit low-discrepancy behavior. Next, we consider non-uniform sampling tests in which the integrand is a product of a binary and a continuous function, a situation that occurs frequently in rendering. Specifically, we use the products of a step function $f(x, y) = \mathbb{1}\left[x < \frac{1}{\pi}\right]$ with a Gaussian $f(x, y) = e^{-x^2-y^2}$, and of the quarter-disc with a bilinear function $f(x, y) = xy$ [Christensen et al. 2018]. We choose the Gaussian and quarter-disc factors, respectively, as the target densities for importance sampling and optimize our samples accordingly. For the Gaussian target, our samples behave similarly to mapped low-discrepancy baselines due to the smoothness of the density. For the quarter-disc target, we observe lower error, with asymptotic convergence matching scrambled Sobol. For uniform 3D directions, we evaluate the symmetric cosine integrand $|z|$. For the non-uniform cases, we use a cosine-weighted BRDF [Burley 2012] and environment maps [Debevec 1998] as target densities and form product integrands by sampling one of the factors. Non-uniform SOT and GBN are generated over a projected density from sphere to square since their non-uniform spherical generalization is not known. While SOT and certain blue-noise samplers can generate non-uniform samples (as visualized in Fig. 4), we did not find empirical evidence that these non-uniform samples can reduce error in numerical integration beyond random samples, and our results also align with this observation. Although SOT sometimes performs better for small sample counts, our method generally outperforms all the baselines, even asymptotically. Further details are provided in the supplementary material, including the integration errors for the uniformly optimized SOT and GBN baselines.

Rendering Application with Conditional Distributions. As a practical scenario, we evaluate our method for rendering. For this experiment, we render an image by importance sampling the BRDF function or the environment map. The target distributions are conditional distributions, thus the optimization-based approaches like our direct optimization, GBN, and SOT are not practical. For GBN and SOT, we thus used uniform target distributions and apply mapping for importance sampling as is commonly done in (Q)MC rendering. For our method, we can use the neural approach by using view direction for BRDF or normal direction for environment map as conditional parameters. We use normalized cosine-weighted BRDF $f_r(\omega_o, \omega_i)|cos\theta_i|$ or bilinear filtered environment map values with incoming radiance cosine term as the target distribution to be sampled. Our method can still learn to sample from these product distributions directly which do not have analytical form. Fig. 8 shows the results of this experiment. We rendered the images with pbrt [Pharr et al. 2023] and compared RMSE. Our method outperforms the baselines and also converges faster in this practical scenario. The overhead of sample generation is negligible compared to other stages of MC rendering, thus we report this comparison as equal time as well as equal sample count. On an RTX 5090 laptop, generating 10,000 instances of 512 direction samples takes 0.035 seconds.

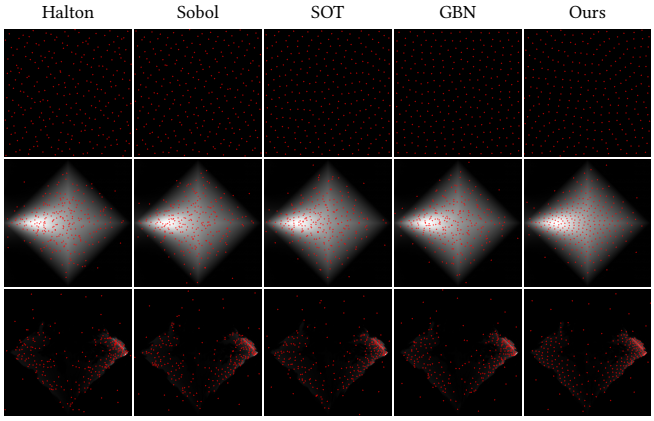

Fig. 4. Visualization of directional samples from Halton, Sobol, Sliced optimal transport, Gaussian blue noise, and our low-discrepancy samples for uniform sphere sampling, BRDF importance sample, and environment map importance sampling. The grayscale background represents the target distribution density. The samples are shown in the octahedral layout.

Optimization cost. In our 2D non-uniform experiments, we optimize for 1000 epochs across all the sample sizes on NVIDIA RTX A5000. Optimizing a 2K-sample set takes about 2.5 minutes for bilinear densities (analytic areas) and about 10.5 minutes for arbitrary densities (MC estimates). For directions, 5000 epochs for a 2K-sample set took 75 seconds on an RTX 5090 laptop. Training a 2K-sample network for a BRDF took 94 minutes on an RTX 3090 desktop. All plots in the paper are averaged over 16 independently optimized solutions.

6 Limitations and Future Work

While our method serves as the first practical optimization method capable of approximating low-discrepancy non-uniform point sets, it has some limitations that motivate future work. Extending our approach to sampling in higher-dimensional spaces might be computationally challenging, as sampling and region-distance computation become harder as dimensionality increases. However, just as optimal transport initially faced scalability issues but became much more efficient [Cuturi 2013], we expect discrepancy optimization to follow a similar trajectory, and our method may scale better in higher dimensions. Our neural network is trained on a fixed number of samples, so it outputs a *set* rather than a *sequence*. A promising result [Huffel et al. 2025] has been reported for approximate sequence generation for uniform samples, where it minimizes prefix discrepancies up to 10K samples. Combining our non-uniform discrepancy loss with prefix-discrepancy optimization could extend this approach to non-uniform sample sequences. A different neural architecture (e.g., graph-based [Rusch et al. 2024]) or encoding may further improve this approach. Likewise, it is possible to include material parameters as additional dimensions in such encoding for BRDF sampling. While we demonstrate low discrepancy for our

tested sample sizes, formal scaling guarantees remain for future study.

Despite limitations, our work opens a new possibility for directly generating non-uniform low-discrepancy samples and already demonstrates improvements in our numerical tests. Together with other optimization-based methods for blue-noise sampling and optimal transport, our work fills in a missing piece in the existing toolset for low-discrepancy samples. For instance, the existing multiple importance sampling (MIS) [Veach and Guibas 1995] heuristics are designed to minimize variance rather than discrepancy, and our work may enable the redesign of the MIS heuristics to minimize discrepancy directly. We believe our work facilitates a more thorough study of different classes of sample generation strategies, using different metrics, for both uniform and non-uniform distributions, to address an open question: finding an ideal metric for samples in numerical integration.

Acknowledgments

This research was supported by NSERC (RGPIN-2020-03918). We thank Paul Debevec for the environment mapping [Debevec 1998], and Yasutoshi Mori (Orb), Christian Schüller (Dragon), and Simon Wendsche (Cup) for the 3D models.

References

Abdalla GM Ahmed, Hélène Perrier, David Coeurjolly, Victor Ostromoukhov, Jianwei Guo, Dong-Ming Yan, Hui Huang, and Oliver Deussen. 2016. Low-discrepancy blue noise sampling. *ACM Transactions on Graphics (TOG)* 35, 6 (2016), 1–13.

Abdalla G. M. Ahmed, Jing Ren, and Peter Wonka. 2022. Gaussian Blue Noise. *ACM Trans. Graph.* 41, 6, Article 260 (Nov. 2022), 15 pages. doi:10.1145/3550454.3555519

Christoph Aistleitner and Josef Dick. 2015. Functions of bounded variation, signed measures, and a general Koksma-Hlawka inequality. *Acta Arithmetica* 167, 2 (2015), 143–171. http://eudml.org/doc/279219

Brent Burley. 2012. Physically-Based Shading at Disney. In *ACM SIGGRAPH 2012 Courses*. Association for Computing Machinery, New York, NY, USA, 1–7. doi:10.1145/2343483.2343493

Paul Otto Chelson. 1976. *Quasi-random Techniques for Monte Carlo Methods*. Ph. D. Dissertation. The Claremont Graduate School.

Jianbing Chen and Pengyan Song. 2016. A generalized L2-discrepancy for cubature and uncertainty quantification of nonlinear structures. *Science China Technological Sciences* 59 (2016), 941–952. https://api.semanticscholar.org/CorpusID:123643162

Per H. Christensen, Andrew Kensler, and Charlie Kilpatrick. 2018. Progressive Multi-Jittered Sample Sequences. *Computer Graphics Forum* 37, 4 (2018), 21–33. doi:10.1111/cgf.13472

Petrik Clarberg. 2008. Fast equal-area mapping of the (hemi) sphere using SIMD. *Journal of Graphics Tools* 13, 3 (2008), 53–68.

Petrik Clarberg, Wojciech Jarosz, Tomas Akenine-Möller, and Henrik Wann Jensen. 2005. Wavelet Importance Sampling: Efficiently Evaluating Products of Complex Functions. *ACM Transactions on Graphics* 24, 3 (2005), 1166–1175. doi:10.1145/1073204.1073328

François Clément, Carola Doerr, Kathrin Klamroth, and Luís Paquete. 2025. Constructing optimal star discrepancy sets. *Proceedings of the American Mathematical Society, Series B* 12, 07 (2025), 78–90.

François Clément, Nathan Kirk, Art B. Owen, and T. Konstantin Rusch. 2025. On the optimization of discrepancy measures. arXiv:2508.04926 [math.NA] https://arxiv.org/abs/2508.04926

Marco Cuturi. 2013. Sinkhorn Distances: Lightspeed Computation of Optimal Transport. In *Advances in Neural Information Processing Systems*, Vol. 26. Curran Associates, Inc., 2292–2300.

Paul E. Debevec. 1998. Rendering Synthetic Objects Into Real Scenes: Bridging Traditional and Image-Based Graphics With Global Illumination and High Dynamic Range Photography. In *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques*. Association for Computing Machinery, New York, NY, USA, 189–198. doi:10.1145/280814.280864

Laurent Dinh, David Krueger, and Yoshua Bengio. 2015. NICE: Non-linear Independent Components Estimation. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7–9, 2015, Workshop Track Proceedings*. San Diego, CA, USA. https://arxiv.org/abs/1410.8516

Bastien Doignies, Nicolas Bonneel, David Coeurjolly, Julie Digne, Loïs Paulin, Jean-Claude Iehl, and Victor Ostromoukhov. 2023. Example-Based Sampling with Diffusion Models. In *SIGGRAPH Asia 2023 Conference Papers*. Association for Computing Machinery, New York, NY, USA, 1–11. doi:10.1145/3610548.3618243

Bastien Doignies, David Coeurjolly, Nicolas Bonneel, Julie Digne, Jean-Claude Iehl, and Victor Ostromoukhov. 2024. Differentiable owen scrambling. *ACM Transactions on Graphics (TOG)* 43, 6 (2024), 1–12.

J. H. Halton. 1964. Algorithm 247: Radical-Inverse Quasi-Random Point Sequence. *Commun. ACM* 7, 12 (dec 1964), 701–702. doi:10.1145/355588.365104

Stefan Heinrich and Alexander Keller. 1994. Quasi-Monte Carlo Methods in Computer Graphics Part II: The Radiance Equation. *Interner Bericht* 243 (1994), 94.

Fred J. Hickernell. 1998. A generalized discrepancy and quadrature error bound. *Math. Comp.* 67, 221 (1998), 299–322.

Edmund Hlawka. 1961. Funktionen von beschränkter Variatiou in der Theorie der Gleichverteilung. *Annali di Matematica Pura ed Applicata* 54, 1 (1961), 325–333.

E. Hlawka and R. Mück. 1972a. A transformation of equidistributed sequences. In *Applications of Number Theory to Numerical Analysis*. Elsevier, 371–388.

E. Hlawka and R. Mück. 1972b. Über eine Transformation von gleichverteilten Folgen II. *Computing* 9, 2 (1972), 127–138.

Michael Etienne Van Huffel, Nathan Kirk, Makram Chahine, Daniela Rus, and T. Konstantin Rusch. 2025. Neural Low-Discrepancy Sequences. arXiv:2510.03745 [cs.LG] https://arxiv.org/abs/2510.03745

Hiroharu Kato, Yoshitaka Ushiku, and Tatsuya Harada. 2018. Neural 3D Mesh Renderer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, Los Alamitos, CA, USA, 3907–3916. doi:10.1109/CVPR.2018.00411

Jurjen Ferdinand Koksma. 1942. A general theorem from the theory of uniform distribution modulo 1. *Mathematica, Zutphen. B* 11 (1942), 7–11.

Thomas Leimkühler, Gurprit Singh, Karol Myszkowski, Hans-Peter Seidel, and Tobias Ritschel. 2019. Deep Point Correlation Design. *ACM Trans. Graph.* 38, 6, Article 226 (nov 2019), 17 pages. doi:10.1145/3355089.3356562

Thomas Müller, Brian Mcwilliams, Fabrice Rousselle, Markus Gross, and Jan Novák. 2019. Neural Importance Sampling. *ACM Trans. Graph.* 38, 5, Article 145 (oct 2019), 19 pages. doi:10.1145/3341156

Harald Niederreiter. 1992. *Random Number Generation and Quasi-Monte Carlo Methods*. SIAM.

Art B. Owen. 2003. Variance with alternative scramblings of digital nets. *ACM Trans. Model. Comput. Simul.* 13, 4 (Oct. 2003), 363–378. doi:10.1145/945511.945518

Art B. Owen. 2023. *Practical Quasi-Monte Carlo Integration*. https://artowen.su.domains/mc/practicalqmc.pdf.

Lois Paulin, Nicolas Bonneel, David Coeurjolly, Jean-Claude Iehl, Antoine Webanck, Mathieu Desbrun, and Victor Ostromoukhov. 2020. Sliced optimal transport sampling. *ACM Trans. Graph.* 39, 4 (2020), 99.

Matt Pharr. 2019. Visualizing Warping Strategies for Sampling Environment Map Lights. https://pharr.org/matt/blog/2019/06/05/visualizing-env-light-warpings.

Matt Pharr, Wenzel Jakob, and Greg Humphreys. 2023. *Physically Based Rendering: From Theory to Implementation* (4th ed.). The MIT Press, San Francisco, CA, USA. 1312 pages.

Adrien Pilleboue, Gurprit Singh, David Coeurjolly, Michael Kazhdan, and Victor Ostromoukhov. 2015. Variance analysis for Monte Carlo integration. *ACM Transactions on Graphics (TOG)* 34, 4 (2015), 1–14.

T. Konstantin Rusch, Nathan Kirk, Michael M. Bronstein, Christiane Lemieux, and Daniela Rus. 2024. Message-Passing Monte Carlo: Generating low-discrepancy point sets via graph neural networks. *Proceedings of the National Academy of Sciences* 121, 40 (2024), e2409913121. doi:10.1073/pnas.2409913121

Wolfgang M. Schmidt. 1969a. Irregularities of distribution. III. *Pacific J. Math.* 29, 1 (1969), 225–234. doi:10.2140/pjm.1969.29.225

Wolfgang M. Schmidt. 1969b. Irregularities of distribution. IV. *Inventiones Mathematicae* 7 (1969), 55–82. doi:10.1007/BF01418774

Peter Shirley. 1991. Discrepancy as a Quality Measure for Sample Distributions. In *EG 1991 Technical Papers*. Eurographics Association, 183–194. doi:10.2312/egtp.19911013

Gurprit Singh, Kartic Subr, David Coeurjolly, Victor Ostromoukhov, and Wojciech Jarosz. 2020. Fourier Analysis of Correlated Monte Carlo Importance Sampling. *Computer Graphics Forum* (2020). doi:10.1111/cgf.13613

Ilya M. Sobol. 1967. On the distribution of points in a cube and the approximate evaluation of integrals. *U. S. S. R. Comput. Math. and Math. Phys.* 7 (1967), 86–112.

Eric Veach and Leonidas J. Guibas. 1995. Optimally combining sampling techniques for Monte Carlo rendering. In *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '95)*. Association for Computing Machinery, New York, NY, USA, 419–428. doi:10.1145/218380.218498

Cem Yuksel. 2015. Sample Elimination for Generating Poisson Disk Sample Sets. *Computer Graphics Forum (Proceedings of EUROGRAPHICS 2015)* 34, 2 (2015), 25–32. doi:10.1111/cgf.12538

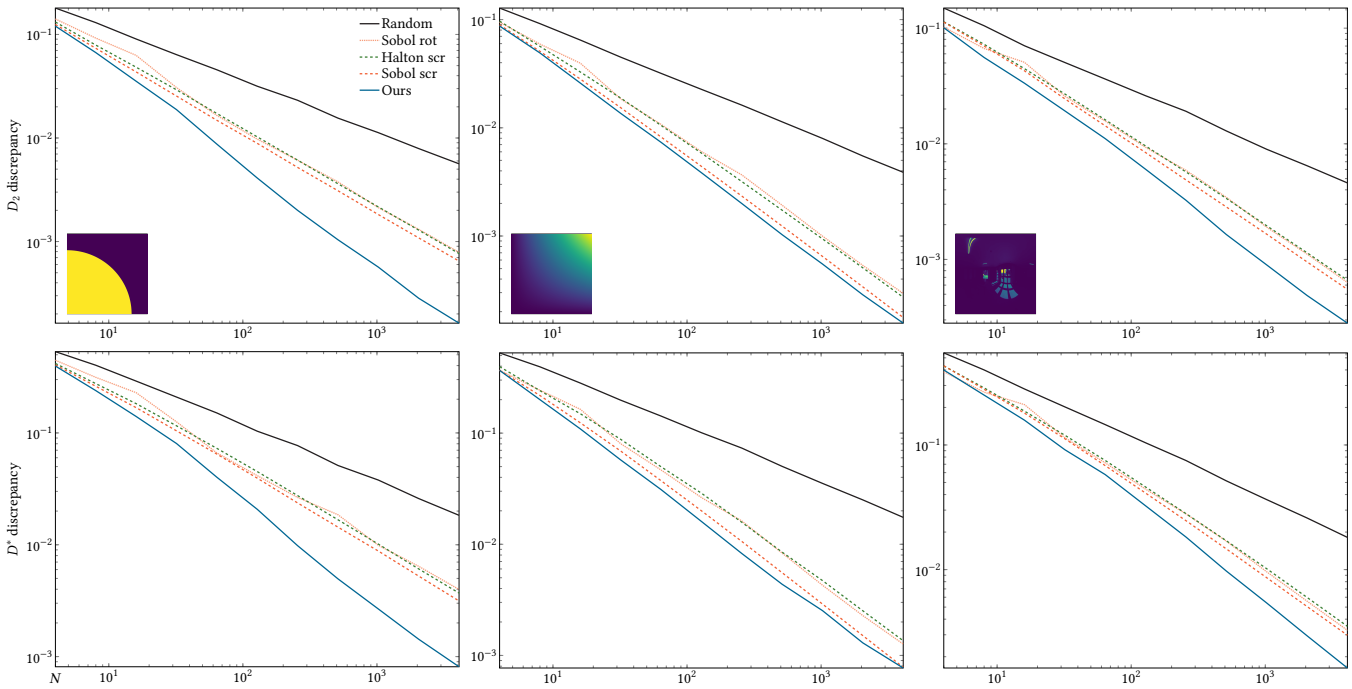

Fig. 5. 2D point samples: L_2 and star discrepancy values (D_2 and D^*) for different numbers N of samples under non-uniform distributions.

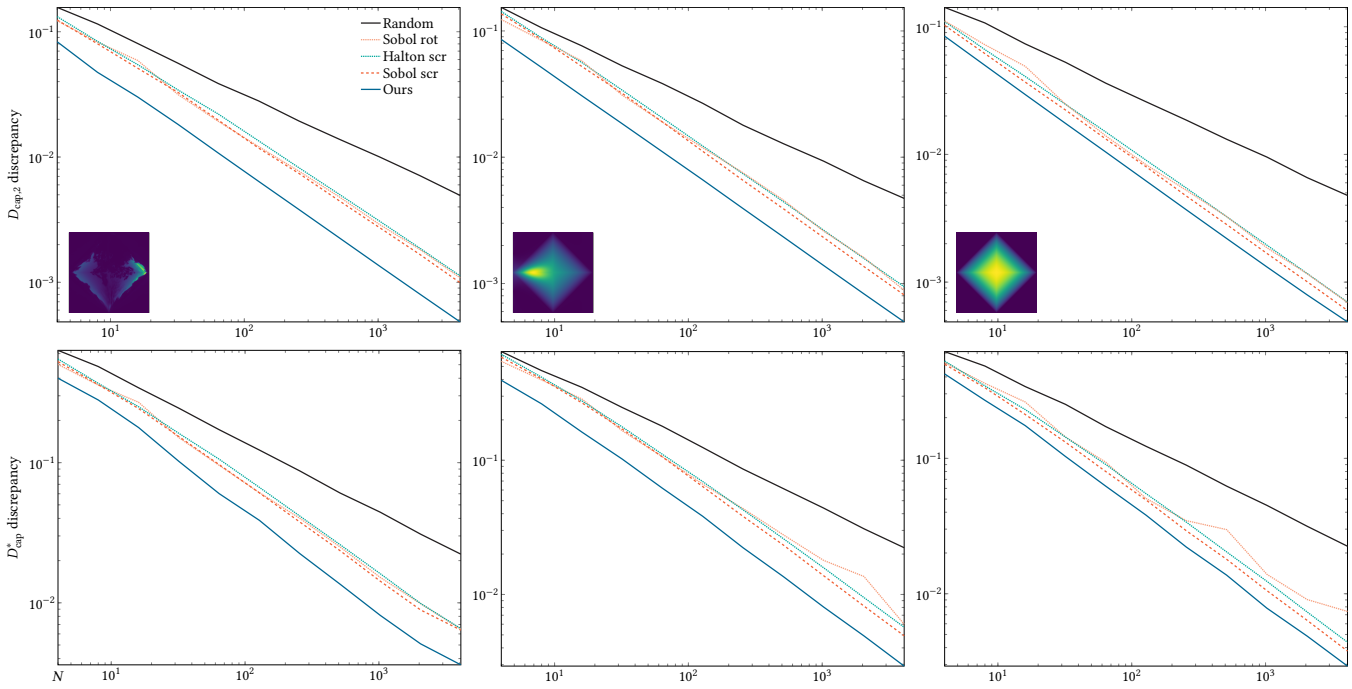

Fig. 6. 3D direction samples: L_2 and star discrepancy values ($D_{\text{cap},2}$ and D^*_{cap}) for different numbers N of directional samples set under non-uniform distributions.

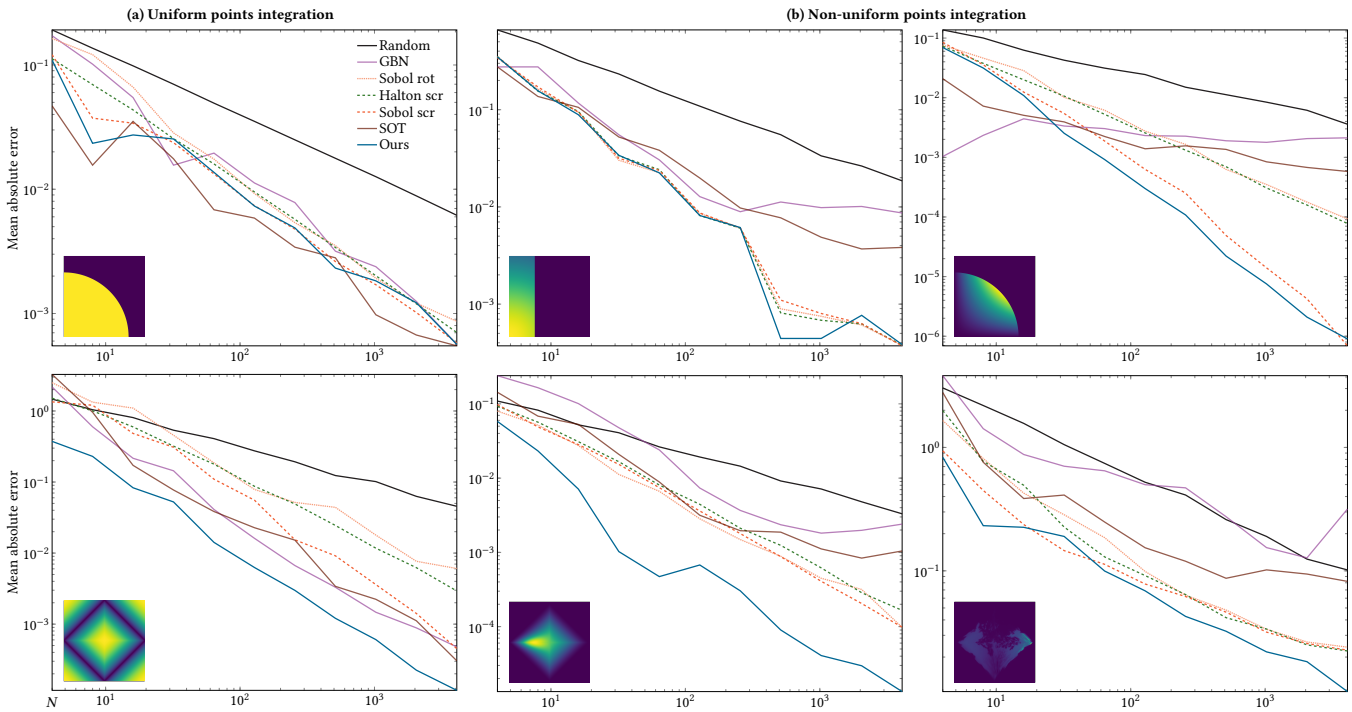

Fig. 7. Numerical integration errors with uniform (a) and non-uniform (b) 2D points (first row) and 3D directions (second row). The insets in this figure visualize the integrands, not the distributions. For 2D points, the integrands are the product of a binary and a continuous function, and for directions, they are the product of a cosine term and a non-uniform distribution.

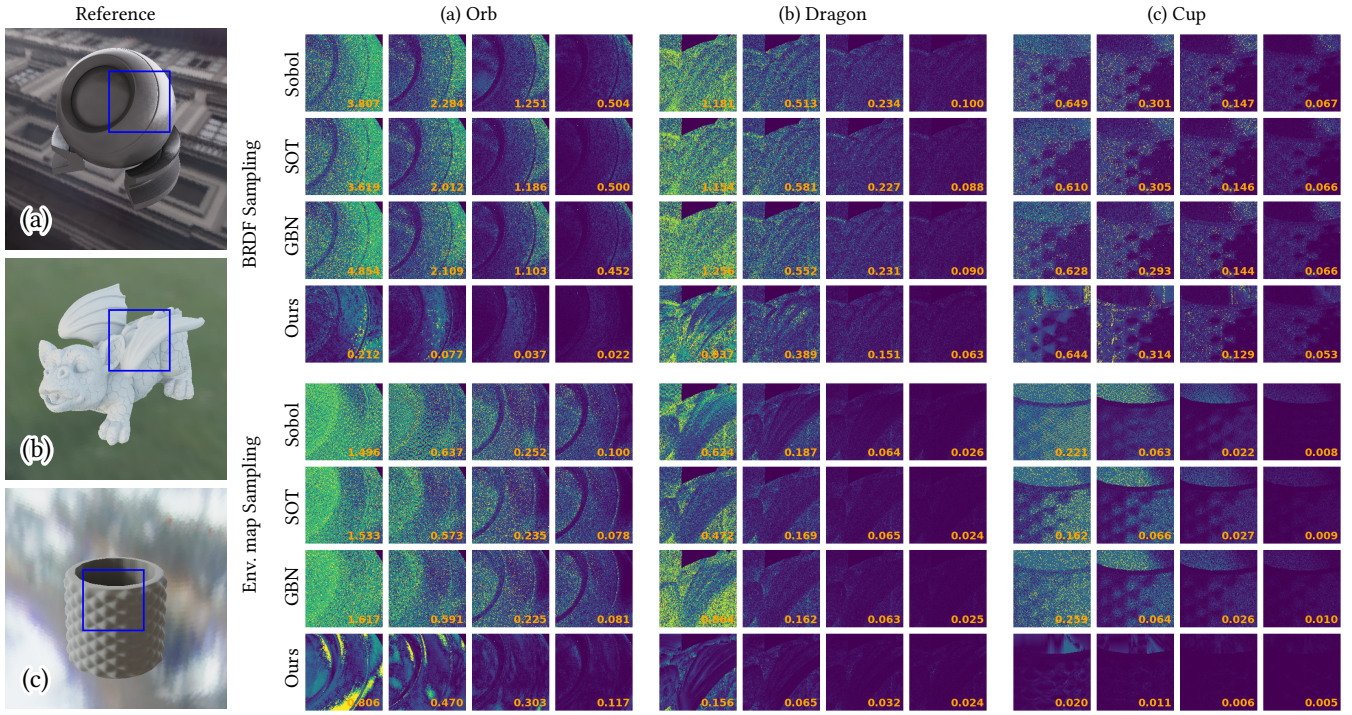

Fig. 8. Image-based lighting with 4, 16, 64, and 256 samples. The number shows the RMSE measured over the whole image, not the inset. The thumbnail shows the local error distributions. Our samples lead to lower error than Sobol, SOT, and GBN for both BRDF importance sampling and environment map importance sampling in most cases. Our method benefits from both direct learning of density and approximate low-discrepancy.