Convergence Analysis for Anisotropic Monte Carlo Sampling Spectra

GURPRIT SINGH and WOJCIECH JAROSZ, Dartmouth College

Traditional Monte Carlo (MC) integration methods use point samples to numerically approximate the underlying integral. This approximation introduces variance in the integrated result, and this error can depend critically on the sampling patterns used during integration. Most of the well-known samplers used for MC integration in graphics—e.g., jittered, Latin-hypercube (N-rooks), multi-jittered—are anisotropic in nature. However, there are currently no tools available to analyze the impact of such anisotropic samplers on the variance convergence behavior of Monte Carlo integration. In this work, we develop a Fourier-domain mathematical tool to analyze the variance, and subsequently the convergence rate, of Monte Carlo integration using any arbitrary (anisotropic) sampling power spectrum. We also validate and leverage our theoretical analysis, demonstrating that judicious alignment of anisotropic sampling and integrand spectra can improve variance and convergence rates in MC rendering, and that similar improvements can apply to (anisotropic) deterministic samplers.

CCS Concepts: • Computing methodologies → Ray tracing;

Additional Key Words and Phrases: Monte Carlo, stochastic sampling, signal processing

ACM Reference format:
DOI: 10.1145/3072959.3073656

1 INTRODUCTION

Since being introduced to graphics by Cook et al. [1984], Monte Carlo (MC) integration has become the cornerstone of most modern rendering algorithms. Historically, MC integration in rendering has involved sampling a function at various stochastically placed points to approximate an integral, e.g., the radiance through a pixel. This estimation is error-prone, however, and many researchers [Durand et al. 2011; Öztireli 2016; Pilleboue et al. 2015; Subr and Kautz 2013; Subr et al. 2014] have therefore investigated how the properties of the integrand and sample points impact the error and convergence rate of this estimation. These analyses have provided important theoretical insights and have lead to tangible improvements in rendering. Many of these analyses have leveraged the Fourier domain to better understand the underlying characteristics of different sampling patterns. The *radially averaged* Fourier power spectrum [Ulichney 1987] has been perhaps the most widely used tool to analyze point samples, characterizing various stochastic sampling patterns ranging from white noise to blue noise, and more recently being used to derive variance convergence rates of various stochastic samplers [Pilleboue et al. 2015].

While *radial averaging* is appropriate for analyzing *isotropic* Fourier power spectra, many of the stochastic point sampling strategies used in rendering—such as N-rooks [Shirley 1991] or even jittered sampling [Cook 1986]—are in fact *anisotropic*. For anisotropic sampling power spectra, radial averaging can be less informative, or worse, misleading. For example, in Fig. 1, the 2D N-rooks sampling pattern has radial behavior of a jittered sampling power spectrum along the canonical axes, but a flat, white noise radial behavior in other directions. This information is lost in the *radially averaged* power spectrum shown at the top of the radial plots.

Most of the signals that we encounter in light transport are also anisotropic in nature, with their spectra having most of their energy confined to a wedge shape [Durand et al. 2005]. Existing sampling patterns, including quasi-random samples (e.g. Halton, Sobol), have not been able to exploit this knowledge despite having strong anisotropic properties in most projections. In this work, we establish a direct relation between the anisotropy of the sampler and the integrand under study, generalizing and extending the reach of prior analyses [Pilleboue et al. 2015] that relied on radial averaging.

We first study the anisotropic sampling processes in Monte Carlo integration and the impact they have on variance and convergence rate. Our primary contribution is a mathematical derivation of variance which enables Monte Carlo convergence rate analysis of arbitrary (isotropic or anisotropic) sampling power spectra. Based on our analysis, we propose a novel approach to transform samples so their power spectrum aligns with the high energy regions of the signal spectrum, resulting in improved variance and convergence rates. Our analysis also establishes a new set of design principles for sampling patterns which can be tailored according to the local light field interactions in a scene. Even though our theoretical framework is developed for stochastic samplers, we demonstrate
that similar improvements can be obtained for deterministic quasi-Monte Carlo (QMC) samplers. We validate all our theoretical results in a depth-of-field setup.

2 RELATED WORK

Point sampling, variance, & convergence. Since the introduction of Monte Carlo to graphics [Cook et al. 1984], researchers have noted that a careful arrangement of samples can impact the spectral distribution and dramatically reduce the overall magnitude of error in numerical integration [Cook 1986; Dippé and Wold 1985; Mitchell 1991]. This has lead to extensive work on generating sample patterns which are stochastic, yet still maintain a low discrepancy [Shirley 1986; Lagae and Dutré 2008]. Subr et al. [2016] summarized recent efforts establishing a firm mathematical connection between the spectral properties of the sampling pattern and the magnitude of Monte Carlo integration error, and Öztireli [2016] established connections to random point processes. Moreover, careful sample placement—which such as jittered [Cook 1986] and certain flavors of blue-noise sampling [Balzer et al. 2009; Heck et al. 2013]—have now been shown to actually lead to asymptotically faster convergence rates [Mitchell 1996; Pilleboue et al. 2015; Ramamoorthi et al. 2012; Subr and Kautz 2013; Subr et al. 2014]. We derive similar mathematical expressions governing variance and convergence rate, but for the case of stochastic placement and evaluation of point samples that might have an arbitrary anisotropic expected power spectrum.

Generation/analysis of anisotropic sampling patterns. Most sampling methods used in graphics are inherently anisotropic (e.g. jittered, N-rooks, multijittered), and anisotropic variants [Feng et al. 2008; Li et al. 2010; Wachtel et al. 2014] of popular isotropic sampling patterns also exist. Such approaches have been shown to be more suitable for certain geometry processing applications [Alliez et al. 2003; Lévy and Liu 2010], or for instancing anisotropic geometric primitives [Li et al. 2010]. Projective relaxation [Reinert et al. 2016] and low-discrepancy blue noise sampling [Ahmed et al. 2016] enforce blue noise properties under multiple planar projections of a point sampling pattern, resulting in anisotropic sampling spectra. Isotropic Fourier tools are ill-equipped to analyze the error arising from Monte Carlo integration using such sampling patterns.

A few analyses have been developed to handle specific forms of anisotropy outside the context of Monte Carlo variance analysis. For anisotropic distributions that arise from a global, invertible warp, it is possible to perform isotropic spectral analysis [Li et al. 2010] after warping the samples back to the uniform domain. Wei and Wang [2011] proposed a more general framework that operates on the inter-sample distances in order to analyze non-uniform/adaptive sampling. With knowledge of the warping function’s Jacobian, they can approximately eliminate differential anisotropy (i.e. local non-uniform scaling or sheering) of the samples for analysis. While these approaches were concerned with spectral analysis, they do not seek to establish how such anisotropy affects variance and convergence rate in Monte Carlo integration.

Light field reconstruction. Multi-dimensional light field reconstruction has been studied extensively in the past decade. Unlike traditional image-based adaptive sampling techniques [Whitted 1980], Hachisuka et al. [2008] proposed to adaptively sample edges in the original multi-dimensional integration domain, before subsequently projecting these samples onto image space for reconstruction. This work, combined with Durand et al.’s [2005] frequency analysis of light transport instigated much of the subsequent work [Zwicker et al. 2015] on sparse sampling and reconstruction. Some algorithms were developed independently for certain effects [Egan et al. 2009; Mehta et al. 2012; Soler et al. 2009; Vaidyanathan et al. 2015], whereas, others handle many distribution effects in a unified manner [Belcour et al. 2013; Lehtinen et al. 2011; Mehta et al. 2014]. While we leverage the insights from these prior analyses, we investigate the problem of integration, instead of reconstruction, which roughly corresponds to error introduced within a pixel vs. across pixels. We establish a direct relation between the anisotropy present in the integrand and the sampling power spectrum, and we demonstrate our results in a depth-of-field setup by leveraging insights from light field analysis [Mehta et al. 2014; Soler et al. 2009; Vaidyanathan et al. 2015] to warp samples prior to MC integration.

3 PRELIMINARIES

We are interested in computing integrals of the form:

\[ I = \int_{\mathbb{D}} f(x) \, dx, \]  

where \( \mathbb{D} \) is the \( d \)-dimensional Euclidean space.

Monte Carlo integration. Monte Carlo integration forms an approximation, \( I_N \), of \( I \) by evaluating the integrand \( f \) at \( N \) sample locations \( s_j \) uniformly distributed over the domain \( \mathbb{D} \). This sampling process can be expressed in continuous form by multiplying the original integrand \( f \) with a normalized sampling function \( S \) consisting of delta responses:

\[ I_N = \int_{\mathbb{D}} S(x) f(x) \, dx, \quad \text{with} \quad S(x) = \frac{1}{N} \sum_{j=1}^{N} \delta(\|x - s_j\|). \]  

In the frequency domain, this integral takes the form:

\[ I_N = \int_{\Phi} \mathcal{F}_S(v) \overline{\mathcal{F}_f(v)} \, dv, \quad \text{with} \quad \mathcal{F}_S(v) = \frac{1}{N} \sum_{j=1}^{N} e^{-2\pi i (v \cdot s_j)}, \]  

where \( \overline{\mathcal{F}_f} \) is the complex conjugate of the integrand’s Fourier spectrum, and \( \mathcal{F}_S \) is the spectrum of the normalized sampling function where each summand is the Fourier transform of a single point.

Variance formulation. Prior work [Durand 2011; Pilleboue et al. 2015] has shown that the variance of \( I_N \) depends on the power spectrum, \( \mathcal{P}(v) = |\mathcal{F}(v)|^2 \), of the integrand and the expected power spectrum of the homogenized\(^1\) sampling pattern:

\[ \text{Var}(I_N) = \int_{\Phi} \mathcal{P}_S(v) \mathcal{P}_f(v) \, dv, \]  

\(^1\)Homogenization of sampling patterns ensures that the sampling Fourier coefficients are uniformly distributed in the complex plane resulting in an unbiased estimator [Subr and Kautz 2013]. In the point processes literature [Illian et al. 2008], homogenization refers to stationary point processes for which the average number of points per some unit of extent such as length, area, or volume remains constant across the domain.
patterns, for instance [Gabrielli and Torquato 2004]:

\[
\langle \mathcal{P}_S(v) \rangle = \begin{cases} 
\frac{1}{N} \bigg( 1 - \prod_i \text{Sinc}(\pi v_i)^2 \bigg) & \text{for random, and} \\
\frac{1}{N} \bigg( 1 - \prod_i \text{Sinc}(\pi v_i)^2 \bigg) & \text{for jittered} 
\end{cases}
\]

where \(v_i\) is the \(i\)-th dimension of frequency vector \(v\).

To analyze the variance and convergence rate of specific sampling patterns, Pilleboue et al. [2015] further simplify Eq. (4) by going to polar coordinates and collapsing the integrand’s power spectrum \(\mathcal{P}_J\) and the expected sampling power spectrum \(\langle \mathcal{P}_S(v) \rangle\) under the assumption of isotropic sampling power spectra—into their radial averages \(\bar{\mathcal{P}}(\cdot)\), arriving at:

\[
\text{Var}(I_N) = \int_0^\infty \rho^{d-1} \bar{\mathcal{P}}(\rho) \bar{\mathcal{P}}_J(\rho) \, d\rho.
\]

With this simplification, their primary contribution was showing that if the radially averaged sampling power spectra can be expressed analytically, then the corresponding variance convergence rates can be derived for a given class of functions. To more easily apply this idea to complex radial power spectra, they showed that it is often sufficient to piecewise bound the radial mean power spectrum using a monomial in the low-frequency region and a constant for high frequencies, with the degree of the low-frequency monomial bound ultimately determining the convergence rate (refer to Sec. 1 of the supplemental for more details). Unfortunately, by relying on radially averaged power spectra, Pilleboue et al.’s analysis only truly applies to isotropic point sampling spectra. This also restricts the scope of their convergence tools since the radial mean would not take into account the anisotropy present within the integrand and therefore, cannot exploit it to improve convergence rates.

In this work, we generalize the variance formulation to anisotropic sampling power spectra while simultaneously coupling it more intimately to the anisotropic structures present in the integrand under study. Our analysis shows that, if the anisotropic structure of the integrand is known, then the anisotropy present within the sampling power spectrum can be exploited to improve convergence rates (Fig. 3, the dark cross region in the multijittered spectrum cancels out the step spectrum resulting in improved convergence for Pixel A in Fig. 2). We explore this idea in detail for the problem of depth-of-field rendering (Sec. 6 & Sec. 7) which has integrand spectra with anisotropic structures that are oriented (sheared) according to the

where \(\Theta\) includes all frequencies except DC, and \(\langle \mathcal{P}_S(v) \rangle\) is the expected power spectrum of the normalized\(^2\) sampling function. We visualize the integrand, sampling spectra, and their product in Fig. 3.

Prior work [Dippé and Wold 1985, 1992; Leneman 1966] has derived analytic expected power spectra for common point sampling

\(^2\)Pilleboue et al. [2015] use an unnormalized sampling function, so their expression for variance includes an additional \(1/N\) normalization factor. We instead fold this factor directly into the normalized sampling function.

ACM Transactions on Graphics, Vol. 36, No. 4, Article 137. Publication date: July 2017.
occluder depths within a pixel. To this end, we theoretically derive convergence tools that do not depend on radial averaging, allowing us to predict correct convergence rates for anisotropic samplers.

4 GENERALIZED VARIANCE FORMULATION

In order to obtain a variance formulation that works for anisotropic sampling spectra, we avoid relying on the radially averaged power spectra Eq. (6), and instead seek a formulation that allows analyzing the radial behavior along any direction. We first rewrite Eq. (4) in polar coordinates:

$$\text{Var}(I_N) = \int_0^\infty \rho^{d-1} \int_{S^{d-1}} <P_S(\rho n) \cdot P_f(\rho n) > \text{d}n \text{d}\rho$$

(7)

where $\rho$ represents the radial component and $n$ is a unit-length vector residing on the $(d-1)$-dimensional sphere $S^{d-1}$ representing the angular component of the frequency vector $\nu = \rho n$.

Generalization. By swapping the order of integration in Eq. (7):

$$\text{Var}(I_N) = \int_{S^{d-1}} \int_0^\infty \rho^{d-1} <P_S(\rho n) \cdot P_f(\rho n)> \text{d}\rho \text{d}n ,$$

(8)

variance becomes an integral over a hypersphere where the outer integral is over all directions and the inner integral is over all radial frequencies for a given direction $n$. We consider $f(\cdot)$ to be Lebesgue integrable, which renders the inner integral bounded. If we restrict our formulation to only well-behaved $P_f$ and $P_S$—having finite bounded discontinuities [Apostol 1974]—the variance integral from Eq. (8) can be written as a limit of sums which corresponds to subdividing the hypersphere into similar-sized (not necessarily equal-sized) cones of any base shape. These cones can be used to approximate the volume of this hypersphere by rewriting Eq. (8) in the following form (see intermediate steps in Sec. 2 of the supplemental):

$$\text{Var}(I_N) = \lim_{m \to \infty} \frac{1}{m} \sum_{k=1}^{m} \int_0^\infty \rho^{d-1} <P_S(\rho n_k) \cdot P_f(\rho n_k)> d\rho \Delta n_k$$

(9)

This is a valid representation in the context of true infinitesimal calculus [Keisler 2012]. In the above formulation, $\Delta n_k$ is the differential volume of the $k$-th cone. In the limit, we assume no angular variation within the $k$-th differential cone, which allows us to consider a single direction corresponding to each cone. Here, $\Delta n_k$ is a constant that becomes infinitesimally small as $m$ tends to infinity.

From Eq. (9), variance of Monte Carlo integration can be obtained by summing the radially integrated terms along each individual direction $n_k$. This implies that, irrespective of whether our expected sampling power spectrum $<P_S(\cdot)>$ is isotropic or not, we can analyze each direction $n_k$ independently to know the overall behavior of the underlying sampler. Doing this in practice is difficult, however, since 1) there are infinitely many possible directions, and 2) computing variance along any direction would require solving the integral analytically which is only possible if we know the analytical form of our sampling and integrand power spectra along each $n_k$.

Since we are only interested in asymptotic convergence rates, it is sufficient for us to analytically bound the radial sampling power spectrum for any direction $n_k$. Like prior work [Pilleboe et al. 2015], we bound the radial behavior, along each direction $n_k$, with a monomial of degree $b_k \geq 0$ in the low-frequency region $(0, \rho_0)$, up to a certain radial frequency $\rho_k$ (see Fig. 4), and a flat spectrum beyond. Instead of doing this for the radially averaged power spectrum, however, we consider this independently for each angular direction $n_k$. Assuming a highly anisotropic sampling power spectrum ($<P_S(\rho n_k)>$), we may require a monomial of different degree $b_k$ along each direction $n_k$ to bound the radial behavior.

In Sec. 5, we first develop our convergence tools for anisotropic sampling power spectra and show that in the limit, convergence rates can be deduced from a single particular direction of a sampling power spectrum. Later, we use our variance formulation to theoretically derive the convergence rates for some well-known samplers (multijittered, Latin-hypercube, randomly shuffled jitter) using the tools we develop along the way.

5 THEORETICAL CONVERGENCE ANALYSIS

Preliminaries. To perform the theoretical analysis, we first restrict our integrands to integrable functions of the form $f(x) \chi_\Omega(x)$ where $f(x)$ is defined in $\Omega$, a bounded domain with smooth boundary (where $\chi_\Omega(x)$ is the characteristic function of $\Omega$) [Brandolini et al. 2003]. This can, however, be extended to arbitrary bounded convex regions [Brandolini et al. 2003]. The worst case from this class of functions exhibits the power fall-off of order $O(\rho^{-(d+1)})$ where $\rho > 0$ is a radial frequency and $d$ is the dimensionality of the signal.

In contrast to Pilleboe et al. [2015], we consider this class of functions for each particular direction $n_k$. As a result, we will have a best and a worst case convergence rate along each direction $n_k$ of the sampling power spectrum. To obtain best and worst case convergence rates for various samplers, we first restrict our integrand power spectra $P_f(\rho n_k)$, along each direction $n_k$, to have the form:

$$P_f^W(\rho n_k) = \begin{cases} c_f \rho & \rho < \rho_0, \\ 0 & \rho \geq \rho_0, \end{cases}$$

$$P_f^B(\rho n_k) = \begin{cases} c_f \rho & \rho < \rho_0, \\ 0 & \rho \geq \rho_0. \end{cases}$$

(10)

where $P_f^B$ refers to the best and $P_f^W$ refers to the worst case integrands, $c_f > 0$ is a constant and $\rho_0 \in \mathbb{R}^+/0$ is finite. Note that, in
the rest of the paper we may drop the superscripts for brevity and refer implicitly to the specific case (best or worst) we are studying.

5.1 Anisotropic convergence tool
To derive a convergence tool that does not depend on radial averaging, we start from Eq. (9) and split the radial integral into separate regions \((0, \rho_k)\) and \((\rho_k, \infty)\):

\[
\mathrm{Var}(I_N) = \lim_{m \to \infty} \sum_{k=1}^{m} \left[ \int_{0}^{\rho_k} \rho^{d-1} \langle P_S(\rho n_k) \rangle P_f(\rho n_k) \, d\rho + \int_{\rho_k}^{\infty} \rho^{d-1} \langle P_S(\rho n_k) \rangle P_f(\rho n_k) \, d\rho \right] \Delta n_k ,
\]

where \(\rho_k = a_k N_k\). Here, \(a_k \in \mathbb{R}^+/0\) is used to quantify the low-frequency region with little or no energy along each \(k\)-th direction. \(N_k\) represents the effective number of samples along the direction \(n_k\). For samplers that rely on stratification (e.g., jittered, multijittered, N-rooks), \(N_k\) would correspond to the effective number of strata along the \(k\)-th direction.

We now characterize sampling power spectra that have monomial behavior (with degree \(b_k > 0\)) in the low-frequency region (Fig. 4), along each \(k\)-th direction in terms of this effective sample count \(N_k\):

\[
\langle P_S(\rho n_k) \rangle = \begin{cases} 
\frac{Y_k}{N} \left( \frac{\rho}{a_k N_k} \right)^{b_k} & \rho < a_k N_k \\
\frac{Y_k}{N} & \text{otherwise}
\end{cases}
\]

where \(Y_k > 0\). In Fig. 4, we illustrate some sampling patterns with their corresponding monomial bounds (proposed by Pilleboue et al. [2015]) in the low-frequency region. After plugging Eq. (12) back into Eq. (11), we obtain our anisotropic convergence tool. Note that, contrary to Pilleboue et al. [2015], our convergence tool establishes a direct link between the anisotropic structures present within the sampler and the integrand under study, due to its dependence on the radial behavior of their product along each direction \(n_k\). Isotropic sampling power spectra become a special case in our formulation where all directions behave exactly the same, thereby ignoring the anisotropic structures present within \(P_f\). As a result, to obtain convergence rates for stochastic samplers with isotropic power spectra, we can simply use \(N_k = \sqrt{N}\) in Eq. (12) and get the convergence rates shown by Pilleboue et al. [2015].

Since variance convergence rate directly depends on the degree of the monomial [Pilleboue et al. 2015] used to bound the radial curve, we can derive convergence rates along each direction \(n_k\) separately in terms of the order \(O()\) of the effective number of samples along \(n_k\). Asymptotically, the summation from Eq. (11) showing the worst convergence rate among all would dominate the overall convergence rate of the underlying anisotropic samples. This implies that, ultimately, the convergence rate would be dictated by the direction \(n_k\) that exhibits the worst convergence behavior. A mathematically more rigorous proof of this statement is given in Appendix A. This also establishes a direct relation between the anisotropic structures present within the sampler and the integrand under study.

5.2 Anisotropic point sampling power spectra
While our convergence tool is valid for arbitrary spectra, leveraging this in the abstract case would require immense knowledge about the anisotropy present within the sampling and integrand power spectra. Luckily, anisotropic samplers are typically designed while keeping in mind the target spectra. Even though existing techniques [Wachtel et al. 2014; Zhou et al. 2012] assume complicated targets, we later show that (Sec. 7), in practice for MC integration, it is beneficial to introduce anisotropy in the directions where the behavior of the integrand is known. In this section, we derive the convergence rates for some well-known stratified samplers (with anisotropic power spectra) as well as the samplers that combine these low-dimensional stratified samplers to higher dimensions.

5.2.1 Randomly shuffled stratification. Generating jittered samples in 2D (or 1D) subspaces and then randomly shuffling these subspaces to form a high dimensional sampler is a common technique in rendering [Pharr et al. 2016]. Cook [1986] first proposed this idea to create 5D samples for depth of field and motion blur (randomly coupling a 1D jittered point set with two 2D jittered sets).

We consider a \(d\)-dimensional randomly shuffled multijittered pattern with \(N\) samples, where we generate multijittered samples within 2D subspaces which are then randomly permuted to form a \(d\)-dimensional tuple. In Fig. 5, we show the expected power spectra of different 2D projections of a 4D multijittered sampler which is created by first generating two 2D multijittered point sets in the XY & UV subspaces. The samples are then randomly shuffled between the original subspaces XY & UV, which results in a power spectrum with a dark cross around the DC in the mixed projections (XY, UV).

Canonical axes. From Fig. 5, we observe that the original subspaces XY and UV have \(N\) strata along the 1D canonical axes due to the multijittered sampling pattern, which implies that the effective number of samples along the canonical axes is \(N_k = N\). In the mixed projections, the stratification is only preserved along the canonical axes after the random shuffling between XY & UV. Therefore, overall the canonical axes would behave exactly the same in all projections. To obtain the convergence rate along the canonical axes (with \(d = 1\)), we substitute Eqs. (10) and (12) into Eq. (11). As a
result, the worst case convergence rate can be derived as:

\[
\operatorname{Var}(I_N) < \int_0^{aN} \rho^d \frac{Y}{\alpha N} \left( \frac{\rho}{\alpha N} \right)^2 d\rho + \int_0^{\infty} \rho^{d-1} \frac{Y}{\alpha N} c_f \rho^{d-1} d\rho.
\]

\[
= \frac{c_f y}{2 N^3} \int_0^{aN} \rho^2 d\rho + \frac{c_f y}{N} \int_0^{\infty} \rho^{-2} d\rho = O(N^{-2}). \quad (13)
\]

and the best case convergence along the canonical axes is given by:

\[
\operatorname{Var}(I_N) < \int_0^{aN} \rho^d d\rho + \int_0^{\infty} \rho^{d-1} d\rho = O(N^{-3}). \quad (14)
\]

 Arbitrary directions in XY & UV. In all other arbitrary directions (except the canonical axes) in the original subspaces, the samples have 2D stratification. This implies that the effective number of samples along these directions would be \(N_k = \sqrt{N}\), for which the worst and best case convergence rates have been already derived as \(O(N^{-1.5})\) and \(O(N^{-2})\), respectively [Pilleboue et al. 2015].

 Arbitrary direction and mixed projections. In the mixed projections (e.g., XV, XU, YV, YU), or any directions not satisfying the above special cases, the power spectrum is flat. This is because after the random shuffling between the original (XY & UV) subspaces, only the stratification along the canonical axes is preserved. As a result, we would obtain a convergence rate of \(O(N^{-1})\) in all these directions for best or worst case integrands.

We can summarize the convergence rates for randomly shuffled multijittered sampling from all these regions in the following form:

\[
\operatorname{Var}(I_N) < \begin{cases} 
    w_{1D} N^{-2} + w_{2D} N^{-1.5} + w_{dD} N^{-1} & \text{worst-case} \\
    w_{1D} N^{-3} + w_{2D} N^{-2} + w_{dD} N^{-1} & \text{best-case},
\end{cases}
\]

where \(w_{1D}, w_{2D}\) and \(w_{dD}\) are weighting constants based on the amount of energy in the integrand spectrum in the 1D, 2D and \(d\)-dimensional subspaces discussed above. If the integrand varies only along the canonical axes (i.e. \(w_{2D} = 0, w_{dD} = 0\)) we will obtain 1D convergence rates due to \(N_k = N\) effective samples. If the integrand has only 2D variation and lies completely within one of the original 2D stratified subspaces (XY or UV), both \(w_{1D}\) and \(w_{dD}\) are 0 and we get 2D jittered convergence rates. If the integrand spectrum has energy varying in more than two dimensions (i.e. \(w_{dD} \neq 0\)), we should observe an overall convergence rate of \(O(N^{-1})\) in the limit. However, because Eq. (15) is a polynomial, the weighting constants will impact the behavior of variance at finite samples counts, allowing the variance to follow different slopes before settling on this asymptotic behavior. For instance, if \(w_{1D} \gg w_{dD}\) and/or \(w_{2D} \gg w_{dD}\), we may initially observe good variance reduction following the 1D or 2D rates for small \(N\), before ultimately deteriorating to \(O(N^{-1})\) in the limit. While the primary motivation of our formulation was on asymptotic convergence rates, it also provide (to our knowledge) the first principled explanation for the multi-slope convergence behaviors of certain sampler-integrand combinations.

 Latin-hypercubes. A Latin-hypercube sampler (also called \(N\)-rooks [Shirley 1991] in 2D), generates \(N\) jittered samples along each dimension which are then randomly shuffled to form a \(d\)-dimensional tuple. The number of strata along each canonical axis is \(N\), resulting in \(N_k = N\) effective samples. This resembles the behavior of Xu & YV (mixed projections) in Fig. 5 which results in the following convergence behaviour:

\[
\operatorname{Var}(I_N) < \begin{cases} 
    w_{1D} N^{-2} + w_{2D} N^{-1} & \text{worst-case} \\
    w_{dD} N^{-1} & \text{best-case}.
\end{cases}
\]

Randomly shuffled jitter. For randomly shuffled 2D jittered samples (aka uncorrelated jitter [Cook 1986]), the effective number of samples along all directions within the original subspaces is \(N_k = \sqrt{N}\). In the mixed projections it would also have \(N_k = \sqrt{N}\) effective samples along the canonical axes, which results in an overall convergence rate of:

\[
\operatorname{Var}(I_N) < \begin{cases} 
    w_{1D} N^{-2} + w_{2D} N^{-1.5} + w_{dD} N^{-1} & \text{worst-case} \\
    w_{1D} N^{-3} + w_{2D} N^{-2} + w_{dD} N^{-1} & \text{best-case}.
\end{cases}
\]

If we compare Eq. (17) with Eq. (15), we can see that for integrands with 1D variations along the canonical axes, randomly shuffled multijitter outperforms randomly shuffled jitter. This can be explained by the fact that the effective number of strata present along the canonical axes for jitter is \(N_k = \sqrt{N}\) (see supplemental Fig. 4). This indicates that it is better to use multijittered samples instead of jittered samples (which are common in rendering) to obtain randomly shuffled high-dimensional stratified samples.

Anisotropic blue noise samplers. There exist many blue noise samplers that allow generating anisotropic blue noise power spectra. Fig. 6 shows an example of an anisotropic blue noise [BNOT [de Goes et al. 2012]] sampling power spectrum generated using AA patterns [Ahmed et al. 2015]. Note that this kind of anisotropy only scales the radial behavior by a constant, and would therefore not affect the asymptotic convergence rate in any direction. This is different from what we have seen in the case of stratification methods (Latin-hypercube, multijitter and randomly shuffled jitter) where the anisotropy introduces different convergence rates along different directions. This difference stems from the fact that these stratification methods change the effective number of samples (strata) along some particular directions, whereas anisotropic blue noise samplers that are generated by applying a Jacobian to match some characteristics of the underlying density function [Li et al. 2010; Wei and Wang 2011] change \(N_k\) by at most a constant factor.
When light travels in free space it undergoes a shear in ray space. This knowledge has been leveraged to craft sheared reconstruction filters for different distribution effects like defocus [Soler et al. 2009]. Corresponding sampling strata are also sheared, as shown in Fig. 7.

Energy regions of the integrand spectra. In the spatial domain, the alignment between the low energy regions of the sampling power spectrum with high energy regions of the integrand spectrum are maintained entirely along the horizontal axis and the other in arbitrary orientation. By simply shearing the multijittered samples, the corresponding sampling power spectrum also gets sheared, which turns the problem into anisotropic structures present in the integrand. Fig. 7 illustrates our setup, we see the aperture can be directly extended to 4D without loss of generality. We follow a standard setup and parameterize rays using two parallel planes: a virtual image plane (focal plane) with x-coordinate and an aperture along the u-coordinate, as shown in Fig. 8. We assume a Lambert object O positioned at a distance d which is visible from the entire aperture. As a result, the outgoing radiance from O would be the same when viewed from different locations on the aperture (e.g. the red ray (u0, x0) and the blue ray (u1, x1)). The radiance from this set of rays would be smeared over multiple pixels on the image/focal plane at distance F, rendering O out of focus. From the geometry of our setup, we see the x-axis is sheared w.r.t. u as follows:

\[ x_0 = x_1 + \frac{F - d}{d} u_1 = x_1 + c_F(d) u_1 \text{, on the focal plane .} \]  

Here, \( c_F(d) \) is the circle of confusion on the focal plane. The corresponding shear in the Fourier domain would be: \( \nu_u = \nu_u - c_F(d) \nu_x \) for pixels with no occluders (we handle occluders directly in Sec. 7).

If anisotropic blue noise samplers would somehow change the effective number of samples by more than a constant factor in a particular direction or region while keeping the underlying blue noise characteristics, the convergence rate along that set of directions would change. This is the case with the recent low-discrepancy blue noise [Ahmed et al. 2016] and projective blue noise [Reinert et al. 2016] sampling approaches which ensure that point samples maintain a denser stratification along 1-dimensional projections.

6 SHEARING SAMPLES FOR INTEGRAND SPECTRA

Our theoretical convergence analysis has shown that most well-known (randomly shuffled) stratification methods with anisotropic power spectra can have different convergence rates in different directions, depending on the behavior of the integrand along that direction. Exploiting this for variance reduction, however, requires alignment between the low energy regions of the sampling power spectrum and the high energy regions of the integrand spectrum (recall Fig. 3). This is unlikely to happen automatically because existing anisotropic samplers have anisotropy mainly along the canonical axes, whereas the integrands that we encounter in rendering can have power spectra with arbitrary orientation. To overcome this problem, we propose to shears the samples according to the anisotropic structures present in the integrand. Fig. 7 illustrates our approach using two different 2D integrands, one with a spectrum lying entirely along the horizontal axis and the other in an arbitrary orientation. By simply shearing the multijittered samples, the corresponding sampling power spectrum also gets sheared, which aligns the low energy regions of the sampling spectra with high energy regions of the integrand spectra. In the spatial domain, the corresponding sampling strata are also sheared, as shown in Fig. 7.

To extend our approach to real scenarios, we leverage prior light transport frequency analyses [Durand et al. 2005] which shows that when light travels in free space it undergoes a shear in ray space. This knowledge has been leveraged to craft sheared reconstruction filters for different distribution effects like defocus [Soler et al. 2009]. and motion blur [Egan et al. 2009]. We, however, propose to shears the samples prior to MC integration using the following algorithm:

1. Use frequency analysis to develop an oracle that can predict shear parameters of the integrand spectrum, per pixel.
2. Depending on the pixel frequency characteristics, choose an existing sampler or design a new sampler (potential future work) to obtain maximum benefits.
3. For each pixel, use oracle from step 1 to shear the samples.
4. Perform Monte Carlo integration with these sheared samples.

The proposed algorithm can be applied to any light transport effect for which Fourier analysis could provide the warping (shearing) parameters. In this paper, we demonstrate one realization of this algorithm for the case of depth-of-field rendering.

In the next section, we first briefly analyze depth of field in flatland, both in the spatial and Fourier domains and later show that it is possible to get 2D convergence rates for a 4D depth-of-field integral by simply shearing existing anisotropic samplers. However, for the cases where the anisotropy within the light field cannot be known before hand, it is safer to stick to samplers like 4D jitter which can give convergence rates of \( O(N^{-1.25}) \), which is better than Monte Carlo \( O(N^{-1}) \).

6.1 Defocus blur in flatland

To simplify exposition, we analyze depth of field in flatland, but this can be directly extended to 4D without loss of generality. We follow a standard setup and parameterize rays using two parallel planes: a virtual image plane (focal plane) with x-coordinate and an aperture along the u-coordinate, as shown in Fig. 8. We assume a Lambert object O positioned at a distance d which is visible from the entire aperture. As a result, the outgoing radiance from O would be the same when viewed from different locations on the aperture (e.g. the red ray (u0, x0) and the blue ray (u1, x1)). The radiance from this set of rays would be smeared over multiple pixels on the image/focal plane at distance F, rendering O out of focus. From the geometry of our setup, we see the x-axis is sheared w.r.t. u as follows:

\[ x_0 = x_1 + \frac{F - d}{d} u_1 = x_1 + c_F(d) u_1 \text{, on the focal plane .} \]  

The proposed algorithm can be applied to any light transport effect for which Fourier analysis could provide the warping (shearing) parameters. In this paper, we demonstrate one realization of this algorithm for the case of depth-of-field rendering.

In the next section, we first briefly analyze depth of field in flatland, both in the spatial and Fourier domains and later show that it is possible to get 2D convergence rates for a 4D depth-of-field integral by simply shearing existing anisotropic samplers. However, for the cases where the anisotropy within the light field cannot be known before hand, it is safer to stick to samplers like 4D jitter which can give convergence rates of \( O(N^{-1.25}) \), which is better than Monte Carlo \( O(N^{-1}) \).

6.1 Defocus blur in flatland

To simplify exposition, we analyze depth of field in flatland, but this can be directly extended to 4D without loss of generality. We follow a standard setup and parameterize rays using two parallel planes: a virtual image plane (focal plane) with x-coordinate and an aperture along the u-coordinate, as shown in Fig. 8. We assume a Lambert object O positioned at a distance d which is visible from the entire aperture. As a result, the outgoing radiance from O would be the same when viewed from different locations on the aperture (e.g. the red ray (u0, x0) and the blue ray (u1, x1)). The radiance from this set of rays would be smeared over multiple pixels on the image/focal plane at distance F, rendering O out of focus. From the geometry of our setup, we see the x-axis is sheared w.r.t. u as follows:

\[ x_0 = x_1 + \frac{F - d}{d} u_1 = x_1 + c_F(d) u_1 \text{, on the focal plane .} \]  

Here, \( c_F(d) \) is the circle of confusion on the focal plane. The corresponding shear in the Fourier domain would be: \( \nu_u = \nu_u - c_F(d) \nu_x \) for pixels with no occluders (we handle occluders directly in Sec. 7).

ACM Transactions on Graphics, Vol. 36, No. 4, Article 137. Publication date: July 2017.
This implies that, a horizontal shear in the primal domain would result in a vertical shear in the Fourier domain. This knowledge is sufficient for our purposes. For a more in-depth study of depth of field please refer to the work by Vaidyanathan et al. [2015].

6.1.1 Shearing the samples. From Eq. (18) we can see that a finite aperture shears the light field in the mixed projections (XU & YV shears in a similar fashion). We propose to shear the samples by the same amount in these mixed projections while toroidally wrapping them to keep them within the original integration domain. Consequently, the sheared samples would have a power spectrum that is sheared according to the light field. Note that this is equivalent to inverse shearing the light field to align the integrand spectrum with the original sampling spectrum (inverse sheared light fields shown in Fig. 10).

Shearing x w.r.t. u vs. u w.r.t. x. All well-known anisotropic samplers used in practice have anisotropy along the canonical axes (see Fig. 4 & 5 in the supplemental). Since the Fourier power spectrum is symmetric (Hermitian), it is possible to shear either the horizontal or the vertical axis to align the spectra (see Fig. 13, bottom row). Mathematically, these two shears can be summarized as follows:

\[ x' = x + c_T(d)u \]
\[ u' = u + \frac{1}{c_T(d)}x' \]

To analyze what these two shears correspond to in the flatland light field, we rearrange the depth-of-field setup as shown in Fig. 9, where we have now three objects \((O_1, O_2\) and \(O_3)\) placed around the focal plane. Depending on the distance of these objects from the focal plane, we might want to shear the samples (or inverse shear the light field) along the horizontal or the vertical axis. In Fig. 10, we demonstrate two XU slices that correspond to \(O_2\) (top row) and \(O_3\) (bottom row) object locations alongside their power spectra (second column) and the inversely sheared light fields (third and fourth columns, respectively). Since the XU slice for \(O_2\) location has power spectrum oriented at \(< 45^\circ\) from the \(x\)-axis, inverse shearing the light field along the \(x\)-axis appears to introduce only one discontinuity (due to the toroidal wrapping). If we inverse shear along the \(u\)-axis, however, the toroidal wrapping introduces many discontinuities. In the spatial domain, this can be explained by the distance of the object from the focal plane. If the object \((O_2)\) is within \(d_1 < d < d_2\), there will always be an overlapping region on \(O_2\) as seen from any two locations on the aperture (shown for \(a\) and \(b\) in Fig. 9). The inverse shearing along the \(x\)-axis doesn’t change the underlying slice content drastically as long as there is a non-zero overlapping region (top row, Fig. 10).

For objects \(O_1\) and \(O_3\) where \(d_1 > d > d_2\), there will instead be an overlapping region \((R_1, R_3)\) on each of them when they are seen from two different locations \(p\) and \(q\) on the focal plane (see Fig. 9). An XU slice corresponding to an object at \(O_3\) is shown in Fig. 10 (bottom row), which suggests that in this case, inverse shearing along the \(u\)-axis will work better.

We can easily derive this depth range, \(d_1 < d < d_2\), in terms of the radius \(R\) of the aperture (lens) and the width \(2W\) of the (pixel) focal plane which is given by: \(d_1 = \frac{RE}{2W}\), \(d_2 = \frac{RE}{2W}\). This implies that, depending on the distance from the focal plane, we might want to shear the aperture \(u\)-coordinate samples or the focal plane \(x\)-coordinate samples. In the next section, we will show the impact of shearing these two coordinates on the variance convergence rates.

7 EXPERIMENTS

We now validate our theory with a set of depth-of-field rendering experiments, analyzing variance when employing various anisotropic samplers commonly used in rendering. We also demonstrate that our approach of shearing samples is not only limited to stochastic samplers, but can lead to similar benefits with common deterministic low-discrepancy sequences such as Halton [1960] and Sobol [1967]. We perform our light field and variance analysis using PBRT-v3 [Pharr et al. 2016] and will release our PBRT plugins.

Implementation. We use a square (finite size) aperture for all the scenes rendered in this paper and a box pixel filter throughout our analysis to avoid any impact of filtering on the variance. To restrict the dimensionality of the integration to 4D, all the scenes contain one point light source. We use eight different anisotropic samplers in our experimental analysis. We include some standard samplers included in PBRT like the Latt hypercube sampler and stratified2d.
which is a form of randomly shuffled jitter that stratifies the XY and UV planes with jitted samples and then randomly permutes the samples in between the two subspaces to form a 4D sample. On top of that, we have added randomly shuffled multijitter, randomly shuffled correlated-multijitter [Kensler 2013], and a 4D stratified sampler (stratified4d) which directly stratifies in 4D. We include independent random samples for reference. We also include PBRT’s Halton (non-scrambled) and Sobol (non-scrambled) samplers in our analysis. All samplers are homogenized in the random number space. We show the power spectra of all these samplers (for different XY, XU, YV, UV projections) in the supplemental (Fig. 4 & 5).

We compute the power spectra of all the signals for validation and visualization purposes (shown in Fig. 11) by first sampling the underlying signal with \[ N = 16384 \] regular samples in the XU space and then applying the continuous Fourier transform on these samples (all in PBRT). For Pixels 3, 5, 6, 7, 8, a few parallel bright streaks are visible alongside the central replica. These are the aliases of the central replica due to the lower sampling count. Pixels 9 & 10 show a strong high energy cross (\( \text{Sinc}(\cdot) \)) around the DC. Any filter that goes to zero at the boundaries (e.g. truncated Gaussian or a triangle filter) can be used to reduce the impact of this \( \text{Sinc}(\cdot) \). In other pixels, this high energy cross is not visible due to no sudden change in the signal amplitude across the boundaries of the domain.

7.1 Convergence analysis of in-focus pixels

**Dragons.** We start with in-focus pixels (Pixels A & B) from Fig. 2. Even though we are solving a 4D \((x, y, u, v)\) integral here, the in-focus pixels (at the focal plane) have variations only within the 2D XY subspace. As a result, only the samples in the image plane (XY projection) would impact the convergence rate, which explains the 2D convergence rate of \( O(N^{-1.5}) \) with stratified2d for Pixels A & B.

For samplers like multijitter and Latin-hypercube, there is a strong hairline anisotropy present along the canonical axes. Since the frequency content of the step function in Pixel A lies completely along the horizontal axis (Fig. 3), the resultant spectral product drops down readily, resulting in an improved convergence rate of \( O(N^{-2}) \). For the Perlin noise texture (Pixel B, Fig. 2 and 3), since the integrand has energy spread in all the directions, we fall back to the 2D worst case convergence of \( O(N^{-1.3}) \) with multijitter. This also validates our theory that regardless of the good (low energy) directions present in the sampling spectra, which can give good convergence rates, the true convergence rate stems from the direction with worst behavior.

The correlated-multijittered power spectrum has wider anisotropy along the canonical axes (see supplemental Fig. 4) compared to Latin-hypercube and multijitter, which would indicate further improvement in variance and convergence for Pixel B (Fig. 2). However, due to the high energy streaks in the correlated-multijitter spectrum, the expected improvement is hindered.

7.2 Convergence analysis of out-of-focus pixels

**Cornell box setup.** In Fig. 11, we create a scene to analyze pixels that are affected by both XY and UV samples and show how shearing the samples can improve convergence rate. The scene contains some constant depth objects (disks and planes) that are textured with

ACM Transactions on Graphics, Vol. 36, No. 4, Article 137. Publication date: July 2017.
a very high frequency Perlin noise texture, except the walls and the floor which have an image texture on it (no texture on the ceiling). We visualize the underlying textures for each object in Fig. 9 of the supplementary material. We follow the steps proposed in Sec. 6 to analyze this setup. We first leverage the light transport frequency analysis from Sec. 6.1 that gives the shear parameters (focal distance & depth) for different pixels (step 1). To obtain these shear parameters, we use an oracle (a depth map), that uses $N = 64$ regular grid samples per pixel and returns the depth for a hit point having maximum luminance. We use this depth information to shear the samples (step 3). The frequency power spectra computed for each pixel (shown in Fig. 11) show how the shear in the spatial domain also shears the frequency content in the mixed dimensions (XU & YV subspaces). We study each pixel using several existing samplers, though it would also be possible to choose or design a sampler depending on the underlying pixel frequency footprint (step 2). Fig. 12 shows all the variance plots (please consult the supplemental for better visualizations of these plots).

### 7.2.1 Single depth pixels
In Fig. 11, pixels marked from 1 to 7 are constant depth pixels (with no occluder), with the depth increasing in the given order. Let us start by looking at the variance plots for Pixels 2 & 7 in Fig. 12. Pixel 2 shows a $10 \times$ reduction in variance from shearing Halton and Sobol samples at high sample counts, whereas Pixel 7, which has stronger defocus blur, shows remarkable improvement in convergence: multijitter and stratified2d give...
As expected, shearing stratified4d does not show improvement in convergence and we obtain the 4D convergence rate of $O(N^{-1.25})$, whereas, Pixel 7 which is far from the focal plane shows improvement in convergence only for the $uv$ samples.

2D convergence of $O(N^{-1.5})$ whereas correlated-multijitter gives $O(N^{-1.6})$ convergence for this 4D ($x, y, u, v$) integral. Halton and Sobol also show similar improvements in convergence for Pixel 7. This happens because there is no strong anisotropy present in any of the projections in the power spectrum of stratified4d (Fig. 4 in the supplemental).

**Shearing $xy$ w.r.t. $uv$ vs. shearing $uv$ w.r.t. $xy$.** We now analyze the Fourier content of Pixels 2 & 7. From Fig. 11, we observe that Pixel 2 has a spectrum with orientation of $< 45^\circ$ from $v_x$, whereas Pixel 7’s spectrum is oriented at an angle $> 45^\circ$ from $v_x$. To analyze the impact of these two orientations we study the variance plots for these two pixels separately in Fig. 13. Pixel 2 shows better variance reduction when $xy$ sample coordinates are sheared w.r.t. $uv$, whereas Pixel 7 shows 2D convergence when $uv$-coordinates are sheared w.r.t. $xy$ (see Fig. 13) for the underlying 4D integral $^3$.

To understand this change in convergence, we take a closer look at the power spectra of stratified2d and correlated-multijitter for Pixel 7 in Fig. 13 (bottom row). After shearing $xy$-coordinates w.r.t. $uv$, the only dominant low energy region left is the vertical axis (first sub-column in the bottom row), whereas, the sheared horizontal axis has almost lost its low energy properties (black dots can be seen instead of a dark line). However, shearing $uv$ w.r.t. $xy$ maintains the low energy profile vertically even after shearing (second sub-column in the bottom row), which results in better convergence. In Sec. 6.1.1, we analyzed these two shears in the primal domain.

---

$^3$Note that, since $x$ is parallel to $u$ and $y$ is parallel to $v$ in our setup, shearing $xy$ w.r.t. $uv$ simplifies to shearing $x$ w.r.t. to $u$ and $y$ w.r.t. $v$, and likewise for $uv$ w.r.t. $xy$. 

Fig. 13. **Pixels 2 & 7:** Variance convergence due to shearing the samples depends on the angle of the integrand power spectrum from the horizontal $v_x$ axis, which is directly correlated to the depth of the objects underneath each pixel. Pixel 2 from Fig. 11, which is very close to the focal plane directly benefits from shearing $xy$ samples, whereas, Pixel 7 which is far from the focal plane shows improvement in convergence only for the $uv$ samples.
where we obtained the depth range\((d_1 < d < d_2)\) that is suitable for shearing the \(x\)-coordinates \(w.r.t\) \(u\). We observe similar behavior for Halton and Sobol samplers. In Fig. 12, all the variance plots are computed by shearing the most optimal axis.

Pixel 3 & Pixel 5 show nominal improvements in the convergence. This is because the energy content of these relatively bright pixels is spread across higher frequencies compared to other pixels and therefore, the hairline anisotropic structures of the sampling spectra are not significantly improving the convergence even after \(10^6\) samples. In Sec. 8, we discuss future samplers with wider little or no energy anisotropic structures that could be more beneficial in these scenarios. Pixel 4 gets remarkable improvement in convergence rate of \(O(N^{-1.7})\) with Halton and Sobol after shearing. Pixel 6 shows convergence rates of \(O(N^{-1.25})\) with multijitter, \(O(N^{-1.35})\) with correlated-multijitter and \(O(N^{-1.5})\) with Halton and Sobol after shearing for this 4D integral, as shown in Fig. 12.

### 7.2.2 Two and three depth pixels

Pixels with one or more occluders in Fig. 11 have multiple depths. We start with Pixel 9 which has one occluder (two depths) and therefore, its power spectrum in Fig. 11 has a bright streak passing through the DC (at the center) for each depth (ignore axis-aligned streaks which represents the Sinc function). The front occluder orients the spectrum at an angle \(< 45^\circ\) from \(v_x\) whereas, the object behind orients at an angle \(> 45^\circ\). Following the analysis for Pixels 2 & 7 (Fig. 13), we know that the front occluder requires shearing \(xy\) \(w.r.t\) \(uv\) and vice-versa for the object behind. In Fig. 14, however, we analyze the impact of shearing the samples \(w.r.t\) to either depth separately \((uv \ w.r.t. \ xy \ \text{Fig. 14 (a)} \text{and} \ xy \ w.r.t. \ uv \ \text{Fig. 14 (b)})\). Since Halton and Sobol have \textit{wider} anisotropic structure in the mixed dimensions compared to stochastic samplers (see supplemental Fig. 4 & 5), we observe a slight improvement in convergence while shearing \(uv \ w.r.t.\) to \(xy\) (Fig. 14 (a)).

Pixel 10 in Fig. 11 has two occluders in front, one of which is brighter than the others. The spectrum of this pixel is very similar to Pixel 9, where one central line in the spectrum is due to the depth/shear of the brighter surface and the second is due to the sudden change in intensity while going from the dark occluder to the bright one. For variance analysis, we align the sampling spectra with the frequency content of the occluder which is brighter (our oracle chooses depth based on the luminance value to automatically handle these scenarios). Pixels with two depths can be better handled by performing a 2D shear which could align both the hairline anisotropic structures of the sampler to the integrand spectrum (future work).

#### 7.2.3 Pixels with varying depth

One of the challenging aspect of pixels with continuously varying depth is the accurate estimation of depth. Pixel 11 shows no major improvement despite having a very narrow wedge. Note that, in real scenarios it is common to find pixels with a wedge shaped spectrum due to multiple occluders or continuously varying depth (Pixel 13). Existing samplers cannot show any noticeable improvement for these pixels due to the hairline anisotropy in the mixed projections \((xU, yV)\). Our analysis suggests that new sampling patterns with little or no energy in a wedge of directions might be able to handle such common scenarios more robustly. Pixels 12 and 14 in Fig. 11 also represent regions with continuously varying depth. However, since these pixels are far from the focal plane and have an image texture underneath, the power spectra of these pixels have their frequency content fully packed within a very slim wedge (that looks like a line). If we look at the variance convergence plots in Fig. 12 for these pixels, all samplers (except stratified4d and LHS) show \textit{10x} improvement in variance for Pixel 12 at higher sample counts, whereas for Pixel 14 not only Sobol, multijitter, correlated-multijitter but also stratified2d show 10x improvement in variance (see supplemental material for better visualization of these variance plots). Halton’s convergence improves from \(O(N^{-1.35})\) to \(O(N^{-1.7})\) due to shearing in Pixel 14.

In Fig. 15, we visualize the variance (for \(N = 100489\) correlated-multijittered samples) across different pixels with (right) and without (left) shearing the samples (bright pixels correspond to high variance). We generate 2D correlated-multijittered samples in \(XY\) coordinates w.r.t. \(d\). 

![Image](image_url)
We proposed a generalized variance formulation that works for any
domain (e.g. the background wall) show dramatic improvements whereas
depth discontinuities have modest improvement in variance.

8 CONCLUSIONS AND FUTURE WORK

We proposed a generalized variance formulation that works for any
stochastic sampling power spectrum (isotropic or anisotropic) and
developed convergence tools that directly take into account the
anisotropy present within the sampler and the integrand spectra.
We show that, in the limit, the worst convergence rate of a sampling
pattern can be deduced from a single direction (within its spectrum)
along which the integrand has worst case radial profile. Based on our
theoretical analysis, we proposed to align the low energy regions of
the sampling spectrum with the high energy regions of the integrand
spectrum to reduce variance.

To demonstrate the applicability of our theory, we presented a
comprehensive analysis of existing anisotropic samplers for depth
of field, showing that the variance convergence rate of existing
samplers can be improved during Monte Carlo integration if we have
prior knowledge about the anisotropic structures present within
the integrand. Since a finite aperture causes a shear in ray space
(mixed projections XU, YV) for objects that are not at the focal
plane, we propose to shear samples by the same amount to achieve
convergence improvements for depth-of-field rendering. This can
equivalently be viewed as inverse shearing the light field, to restrict
its variation to axes-aligned projections in random number space.

Our analysis reveals that we should not expect major improve-
ments in variance reduction at low sample counts with current
samplers (even after shearing), and any variance improvement may
be sensitive to the robustness/accuracy of the oracle. This makes
our approach less practical compared to existing reconstruction
algorithms [Mehta et al. 2014; Vaidyanathan et al. 2015]. However,
our approach is unbiased which could eventually converge to the
correct result despite the inaccuracies present in the oracle (in pre-
dicting the spectrum or the depth). By shearing the samples, we
intuitively orient the sampling strata without changing the density
of the samples (see Fig. 7) which results in improved convergence
rate. This is in contrast to methods like MDAS [Hachisuka et al.
2008], which adaptively increases the sampling density across edges
(without increasing the effective stratification). Unlike these recon-
struction approaches and the recent work by Georgiev and Fajardo
[2016]—which exploits inter-pixel correlation to improve the spec-
tral distribution of noise without reducing total error—our analysis
does not take into account neighboring pixels.

Future work. It is common practice to randomly shuffle 2D or 1D
stratified samplers [Pharr et al. 2016] to create a high-dimensional
sampler for Monte Carlo rendering. However, this random shuf-
fling introduces a hairline anisotropy in the mixed dimensions
(projections) for most existing stochastic samplers. These hairline
anisotropic structures are not very helpful in reducing variance for
pixels with wider or wedge-shaped anisotropic structures, which are
common in many distribution effects [Durand et al. 2005; Egan et al.
2011, 2009; Mehta et al. 2012]. Our analysis suggests new design
principles for anisotropic samplers that can have wider or wedge
shaped anisotropies in their spectra in most of their projections.
Halton and Sobol samplers, which already have good anisotropic
structures, could, perhaps, be optimized to obtain wider low energy
regions in different projections. This could also help in light field
camera sample designs [Wei et al. 2015]. Lastly, our design principles
are inline with Pilleboue et al. [2015], which have emphasized on
obtaining isotropic zero energy region in the low frequencies (e.g.
step noise). Our analysis, however, proposes to obtain zero energy
regions along the directions/projections that matter the most, since
a large range of directions would not have any impact on variance
reduction during Monte Carlo integration.

9 ACKNOWLEDGEMENTS

We would like to thank the anonymous reviewers, whose feedback
dramatically improved the final version of this paper, and Emilio
Cobanera and Raghu Mahajan for discussions regarding sphere
discretization. This work was partially supported by NSF grant
CNS-1205521 and a generous gift from Activision.

A CONVERGENCE RATE FROM SINGLE DIRECTION

For brevity, we show only the crucial steps of the proof here. For
a detailed version, please refer to the supplemental Sec. 4. We first
represent the radial integral from Eq. (9) in the following form:

$$\mathcal{R}_n^b(n_k) = \int_a^b \rho^{d-1} \langle P_S(m_k) \rangle / P_f(m_k) dp.$$  \hfill (19)

For $m \to \infty$, we can rewrite Eq. (9) by separating one direction with
the rest of the directions as follows:

$$\text{Var}(I_N) < \mathcal{R}_n^{\infty}(n_1) + \sum_{k=0}^m \mathcal{R}_n^k(n_k).$$ \hfill (20)

Let's assume that we have a constant radial profile along direction
$n_1$ and a monomial behaviour of degree $b_k > 0$ along all other
directions. This would allow us to approximate $\mathcal{R}_n^k(n_k)$ as $O(N^{-1})$
irrespective of the integrand. The second part $\mathcal{R}_n^k(n_k)$ can be ex-
spanded similar to Eq. (11) as follows:

$$\text{Var}(I_N) < O(N^{-1}) + \lim_{m \to \infty} \sum_{k=0}^m (\mathcal{R}_n^k(n_k) + \mathcal{R}_n^\infty(n_k))$$ \hfill (21)

For a worse case, the radial integrals can be solved with a monomial
sampling profile, resulting in:

$$\text{Var}(I_N) < \sum_{k=0}^m O\left(\frac{1}{N^{k}}\right) + O\left(\frac{1}{N^{k}}\right) = O\left(\frac{1}{N^{k}}\right).$$ \hfill (22)

In a similar fashion, we can obtain the variance convergence rates for the best case, in all $n_k / k = 1$ directions:

$$\text{Var}(I_N) < \sum_{k=0}^m O\left(\frac{1}{N^{k}}\right) + O\left(\frac{1}{N^{k}}\right).$$ \hfill (23)

Since, the sum of $O(\cdot)$ notations is asymptotically dominated by
the worst $O(\cdot)$ behaviour (proof in Sec. 3 of the supplemental),
the overall convergence rate (for both the worst (Eq. (22)) and the
best (Eq. (23)) case) would be: $\text{Var}(I_N) < O\left(\frac{1}{N^{k}}\right)$. 

ACM Transactions on Graphics, Vol. 36, No. 4, Article 137. Publication date: July 2017.
However, from Eq. (23), if none of the directions has a constant profile behaviour, the overall convergence rate would be dominated by the k-th direction having the minimum value of $b_k$, and can be written in the following form:

$$\text{Var}(I_N) < O\left(\frac{1}{N N_k b_k}\right)$$

(24)

REFERENCES

Abdalla G. M. Ahmed, Hui Huang, and Oliver Deussen. 2015. AA Patterns for Point Sets with Controlled Spectral Properties. ACM Trans. Graph. (Proc. SIGGRAPH Asia) 34, 6, Article 212 (Oct. 2015), 8 pages.


Laurent Belcour, Cyril Soler, Kartic Suh, Nicolai Holzschoch, and Fredo Durand. 2013. 5D Covariance Tracing for Efficient Defocus and Motion Blur. ACM Transactions on Graphics 32, 3, Article 31 (July 2013), 18 pages.


