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), multijittered—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:
Gurprit Singh and Wojciech Jarosz. 2017. Convergence Analysis for Anisotropic Monte Carlo Sampling Spectra. *ACM Trans. Graph.* 36, 4, Article 137 (July 2017), 14 pages.
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 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].

Author's addresses: Gurprit Singh and Wojciech Jarosz, Computer Science Department, Dartmouth College, Hanover, NH, USA.

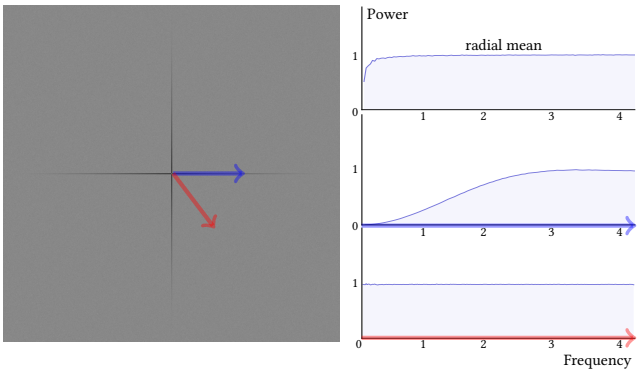


Fig. 1. The expected power spectrum of N-rooks with $N = 256$ samples is highly anisotropic, with drastically different radial behavior along different directions (blue vs. red arrows). Radial averaging (radial mean) masks the good anisotropic properties of the sampler along the canonical axes.

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 1991] or which exhibit so-called blue noise frequency spectra [Cook 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—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:

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

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 $\mathbf{s}_j$ 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:

$$I_N = \int_{\mathbb{D}} S(\mathbf{x}) f(\mathbf{x})\, \mathrm{d}\mathbf{x}, \text{ with } S(\mathbf{x}) = \frac{1}{N}\sum_{j=1}^{N} \delta(\|\mathbf{x} - \mathbf{s}_j\|). \tag{2}$$

In the frequency domain Φ, this integral takes the form:

$$I_N = \int_{\Phi} \mathcal{F}_S(\nu)\overline{\mathcal{F}_f(\nu)}\, \mathrm{d}\nu, \text{ with } \mathcal{F}_S(\nu) = \frac{1}{N}\sum_{j=1}^{N} e^{-2\pi i(\nu\cdot\mathbf{s}_j)}, \tag{3}$$

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, $\mathcal{P}(\nu) = \|\mathcal{F}(\nu)\|^2$, of the integrand and the expected power spectrum of the *homogenized*[1] sampling pattern:

$$Var(I_N) = \int_{\Theta} \langle \mathcal{P}_S(\nu) \rangle\, \mathcal{P}_f(\nu)\, \mathrm{d}\nu\,, \tag{4}$$

[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.

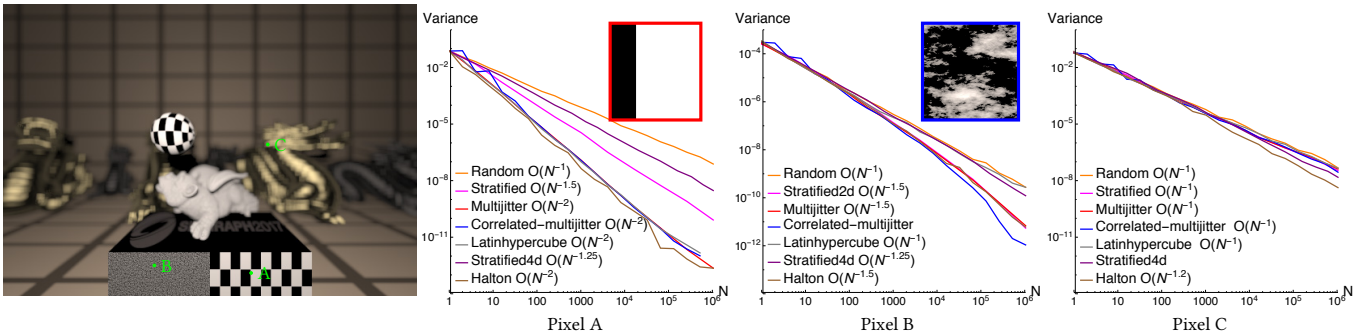


Fig. 2. Dragon scene rendered with strong defocus blur. The underlying integral is 4D (x, y, u, v) but the in-focus pixels (A & B) are still 2D, thereby converging much faster than out-of-focus pixel (C). Multijittered has $O\left(N^{-2}\right)$ convergence in Pixel A compared to $O\left(N^{-1.5}\right)$ for stratified2d, showing that if anisotropic samplers are aware of the anisotropic structures present in the integrand (Fig. 3), both variance and *convergence rate* improve substantially.

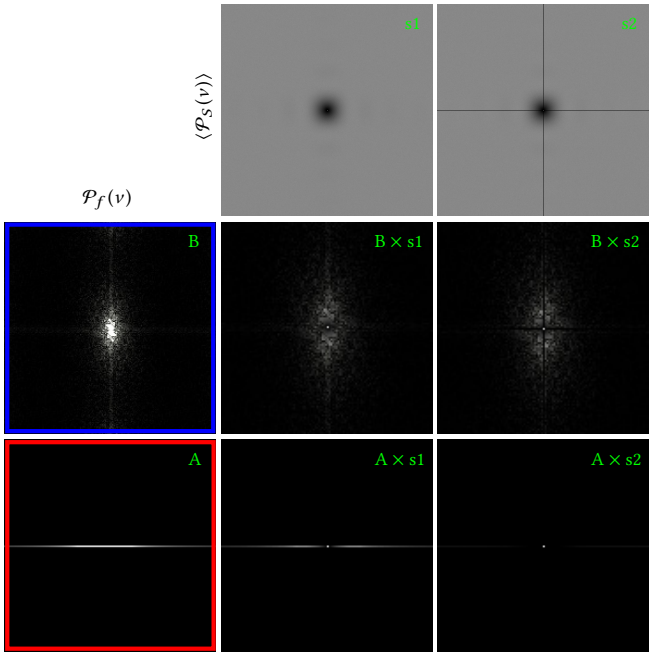


Fig. 3. Variance is the product integral of the integrand's power spectrum $\mathcal{P}_f(v)$ (left column) and expected sampling power spectrum $\langle \mathcal{P}_S(v) \rangle$ (top row), visualized here for the combination of jittered (s1) and multi-jittered (s2) sampling and the step (red) and Perlin noise (blue) integrands from Pixels A and B of Fig. 2, respectively.

where Θ 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 patterns, for instance [Gabrielli and Torquato 2004]:

$$\langle \mathcal{P}_S(v) \rangle = \begin{cases} \frac{1}{N} & \text{for random, and} \\ \frac{1}{N}\left(1 - \prod_i^d \mathrm{Sinc}(\pi v_i)^2\right) & \text{for jittered} \end{cases}, \quad (5)$$

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 $\mathcal{P}_f$ and the expected sampling power spectrum $\langle \mathcal{P}_S(\cdot) \rangle$—under the assumption of isotropic sampling power spectra—into their radial averages $\check{\mathcal{P}}(\cdot)$, arriving at:

$$Var(I_N) = \int_0^\infty \rho^{d-1}\, \check{\mathcal{P}}_S(\rho) \check{\mathcal{P}}_f(\rho)\, \mathrm{d}\rho. \quad (6)$$

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

[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.

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:

$$Var(I_N) = \int_0^\infty \rho^{d-1} \int_{\mathcal{S}^{d-1}} \langle \mathcal{P}_S(\rho\mathbf{n}) \rangle \, \mathcal{P}_f(\rho\mathbf{n}) \, \mathrm{d}\mathbf{n} \, \mathrm{d}\rho \qquad (7)$$

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

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

$$Var(I_N) = \int_{\mathcal{S}^{d-1}} \int_0^\infty \rho^{d-1} \langle \mathcal{P}_S(\rho\mathbf{n}) \rangle \, \mathcal{P}_f(\rho\mathbf{n}) \mathrm{d}\rho \, \mathrm{d}\mathbf{n} \,, \qquad (8)$$

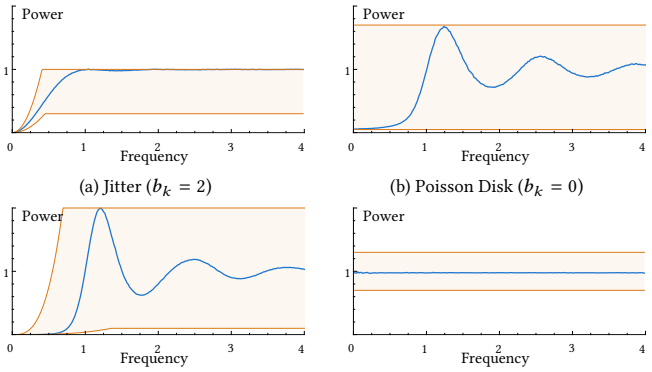
variance becomes an integral over a hypersphere where the outer integral is over all directions and the inner integral is over all radial frequencies for a given direction $\mathbf{n}$. We consider $f(\cdot)$ to be Lebesgue integrable, which renders the inner integral bounded. If we restrict our formulation to only well-behaved $\mathcal{P}_f$ and $\mathcal{P}_S$—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):

$$\boxed{Var(I_N) = \lim_{m\to\infty} \sum_{k=1}^{m} \int_0^\infty \rho^{d-1} \langle \mathcal{P}_S(\rho\mathbf{n}_k) \rangle \, \mathcal{P}_f(\rho\mathbf{n}_k) \mathrm{d}\rho \, \Delta\mathbf{n}_k} \qquad (9)$$

This is a valid representation in the context of true infinitesimal calculus [Keisler 2012]. In the above formulation, $\Delta\mathbf{n}_k$ 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, $\Delta\mathbf{n}_k$ 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 $\mathbf{n}_k$. This implies that, irrespective of whether our expected sampling power spectrum $\langle \mathcal{P}_S(\cdot) \rangle$ is isotropic or not, we can analyze each direction $\mathbf{n}_k$ 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 $\mathbf{n}_k$.

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 $\mathbf{n}_k$. Like prior work [Pilleboue et al. 2015], we bound the radial behavior, along each direction $\mathbf{n}_k$, with a *monomial* of degree $b_k \geq 0$ in the low-frequency region $(0, \rho_k]$,



(a) Jitter ($b_k = 2$) (b) Poisson Disk ($b_k = 0$)

(c) CCVT [Balzer et al. 2009] ($b_k = 3$) (d) Random ($b_k = 0$)

Fig. 4. *Radial* power spectra for common samplers can be approximated (bounded) by simple monomial profiles. Here, monomials of degree (a) $b = 2$ (quadratic), (b) $b = 0$ (constant), (c) $b = 3$ (cubic), and (d) $b = 0$ (constant) bound the radial power spectra.

up to a certain radial frequency ρ_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 $\mathbf{n}_k$. Assuming a highly anisotropic sampling power spectrum ($\langle \mathcal{P}_S(\rho\mathbf{n}_k) \rangle$), we may require a monomial of different degree b_k along each direction $\mathbf{n}_k$ 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 Ω, a bounded domain with smooth boundary (where $\chi_\Omega(x)$ is the characteristic function of Ω) [Brandolini et al. 2001]. 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 $O\left(\rho^{-(d+1)}\right)$ where $\rho > 0$ is a radial frequency and d is the dimensionality of the signal.

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

$$\mathcal{P}_f^W(\rho\mathbf{n}_k) = \begin{cases} c_f & \rho < \rho_0, \\ c_f \rho^{-d-1} & \rho \geq \rho_0 \end{cases} \text{ and } \mathcal{P}_f^B(\rho\mathbf{n}_k) = \begin{cases} c_f & \rho < \rho_0, \\ 0 & \rho \geq \rho_0 \end{cases}, \qquad (10)$$

where $\mathcal{P}_f^B$ refers to the best and $\mathcal{P}_f^W$ 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 (ρ_k, ∞):

$$Var(I_N) = \lim_{m\to\infty} \sum_{k=1}^{m} \left[\int_0^{\rho_k} \rho^{d-1} \langle \mathcal{P}_S(\rho \mathbf{n}_k) \rangle \mathcal{P}_f(\rho \mathbf{n}_k) \mathrm{d}\rho + \int_{\rho_k}^{\infty} \rho^{d-1} \langle \mathcal{P}_S(\rho \mathbf{n}_k) \rangle \mathcal{P}_f(\rho \mathbf{n}_k) \mathrm{d}\rho \right] \Delta \mathbf{n}_k \,, \tag{11}$$

where $\rho_k = \alpha_k N_k$. Here, $\alpha_k \in \mathbb{R}^+/0$ 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 $\mathbf{n}_k$. 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:

$$\langle \mathcal{P}_S(\rho \mathbf{n}_k) \rangle = \begin{cases} \dfrac{\gamma_k}{N} \left(\dfrac{\rho}{\alpha_k N_k} \right)^{b_k} & \rho < \alpha_k N_k \\ \dfrac{\gamma_k}{N} & \text{otherwise} \end{cases} \tag{12}$$

where $\gamma_k > 0$. 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 $\mathbf{n}_k$. *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 $\mathcal{P}_f$. As a result, to obtain convergence rates for stochastic samplers with *isotropic* power spectra, we can simply use $N_k = \sqrt[d]{N}$ 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 $\mathbf{n}_k$ separately in terms of the order $O(\cdot)$ of the effective number of samples along $\mathbf{n}_k$. Asymptotically, the summand 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 $\mathbf{n}_k$ 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.

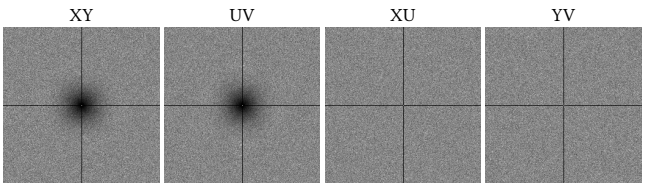


Fig. 5. Expected power spectra computed using $N = 4096$ samples for randomly shuffled multijitter in 4D. Power spectra for other stratification strategies is shown in Fig. 4 of the supplemental.

5.2 Anisotropic point sampling power spectra

While our convergence tool is valid for arbitrarily 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* (XU, YV).

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:

$$Var(I_N) < \int_0^{\alpha N} \rho^{d-1} \frac{\gamma}{N} \left(\frac{\rho}{\alpha N}\right)^2 c_f \, \mathrm{d}\rho + \int_{\alpha N}^{\infty} \rho^{d-1} \frac{\gamma}{N} c_f \rho^{-d-1} \, \mathrm{d}\rho,$$
$$= \frac{c_f \gamma}{\alpha^2 N^3} \int_0^{\alpha N} \rho^2 \mathrm{d}\rho + \frac{c_f \gamma}{N} \int_{\alpha N}^{\infty} \rho^{-2} \mathrm{d}\rho = O\left(N^{-2}\right), \quad (13)$$

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

$$Var(I_N) < \int_0^{\rho_0} \rho^{d-1} \frac{\gamma}{N} \left(\frac{\rho}{\alpha N}\right)^2 c_f \, \mathrm{d}\rho = O\left(N^{-3}\right). \quad (14)$$

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 $O\left(N^{-1.5}\right)$ and $O\left(N^{-2}\right)$, 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 $O\left(N^{-1}\right)$ 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:

$$Var(I_N) < \begin{cases} \mathrm{w}_{1D} N^{-2} + \mathrm{w}_{2D} N^{-1.5} + \mathrm{w}_{dD} N^{-1} & \text{worst-case} \\ \mathrm{w}_{1D} N^{-3} + \mathrm{w}_{2D} N^{-2} + \mathrm{w}_{dD} N^{-1} & \text{best-case}, \end{cases} \quad (15)$$

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. $\mathrm{w}_{2D} = 0, \mathrm{w}_{dD} = 0$) 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. $\mathrm{w}_{dD} \neq 0$), we should observe an overall convergence rate of $O\left(N^{-1}\right)$ 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 $\mathrm{w}_{1D} \gg \mathrm{w}_{dD}$ and/or $\mathrm{w}_{2D} \gg \mathrm{w}_{dD}$, we may initially observe good variance reduction following the 1D or 2D rates for small N, before ultimately deteriorating to $O\left(N^{-1}\right)$ 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:

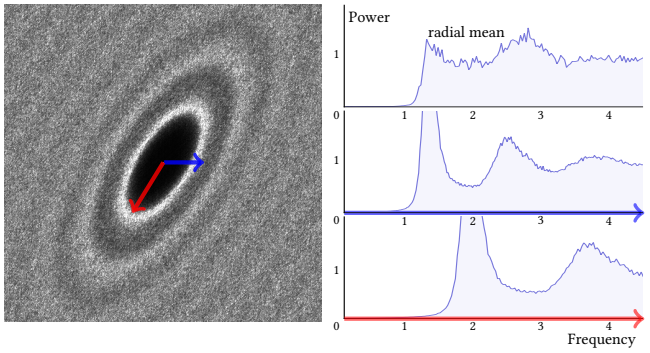


Fig. 6. Stretching blue noise (BNOT) [de Goes et al. 2012] samples produces an anisotropic sampling spectrum (left), but this only influences the effective number of samples in any given direction (right) by a constant factor.

$$Var(I_N) < \begin{cases} \mathrm{w}_{1D} N^{-2} + \mathrm{w}_{dD} N^{-1} & \text{worst-case} \\ \mathrm{w}_{1D} N^{-3} + \mathrm{w}_{dD} N^{-1} & \text{best-case}. \end{cases} \quad (16)$$

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:

$$Var(I_N) < \begin{cases} \mathrm{w}_{1D} N^{-1.5} + \mathrm{w}_{2D} N^{-1.5} + \mathrm{w}_{dD} N^{-1} & \text{worst-case} \\ \mathrm{w}_{1D} N^{-2} + \mathrm{w}_{2D} N^{-2} + \mathrm{w}_{dD} N^{-1} & \text{best-case}, \end{cases} \quad (17)$$

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.

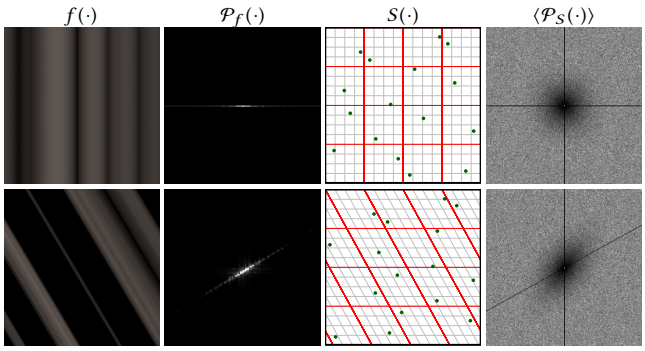


Fig. 7. **Top row** illustrates an integrand with spectrum energy only along the horizontal direction. Existing anisotropic samplers can be directly used to integrate this function. However, for an integrand with an arbitrarily oriented power spectrum (**bottom row**), we propose to shear the samples to match the low energy region of the sampling power specrum with the high energy regions of the integrand spectrum.

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 shear 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]

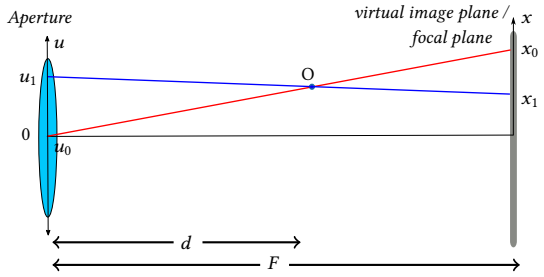


Fig. 8. The depth of field setup used in our implementation is shown here with an object O placed at a depth d from the lens. The shear due to the finite size of the aperture can be computed at the focal plane.

and motion blur [Egan et al. 2009]. We, however, propose to shear 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 $O\left(N^{-1.25}\right)$, which is better than Monte Carlo $O\left(N^{-1}\right)$.

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 Lambertian 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 (u_0, x_0) and the blue ray (u_1, x_1)). 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:

$$x_0 = x_1 + \frac{F-d}{d} u_1 = x_1 + c_F(d) u_1 \ , \text{ on the focal plane} . \qquad (18)$$

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).

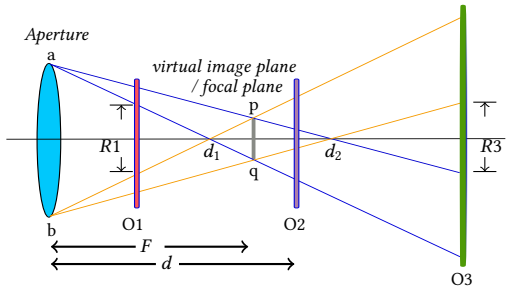


Fig. 9. The setup from Fig. 8 is modified to illustrate the distance range ($d_1 < d < d_2$) where an object (O1, O2, O3) can have a significant overlap in the ray space when seen either from the aperture (a and b) or the focal plane (p and q). This could help decide whether to shear the samples along the x-axis or the u-axis.

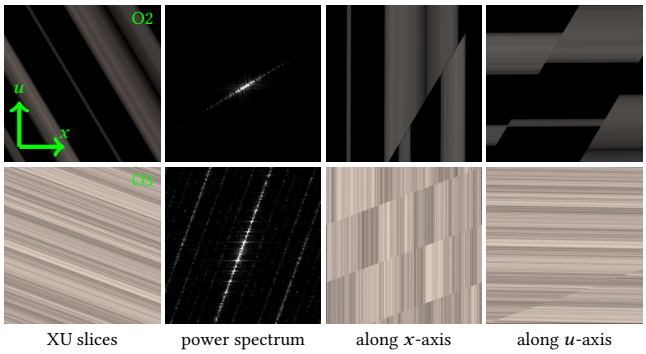


Fig. 10. Inverse shearing the light field (XU slices) could drastically change the original slice content if the shearing is not performed along the *proper* axis. In the **top row**, an XU slice corresponding to object location O2 maintains most of the original content with only one discontinuity when sheared along x-axis, whereas, shear along the u-axis seems to work well for O3 (**bottom row**, more details in the text).

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 (inversely 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: $x = x + c_F(d)u$ and $u = u + \frac{1}{c_F(d)}x$.

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 ($O1$, $O2$ and $O3$) 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 $O2$ (top row) and $O3$ (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 O2 location has power spectrum oriented at $< 45°$ 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 ($O2$) is within $d_1 < d < d_2$, there will always be an overlapping region on $O2$ 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 $O1$ and $O3$ where $d_1 > d > d_2$, there will instead be an overlapping region ($R1$, $R3$) 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 $O3$ 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: $d_1 = \frac{RF}{R+W}$, $d_2 = \frac{RF}{R-W}$. 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 Latin hypercube sampler and stratified2d,

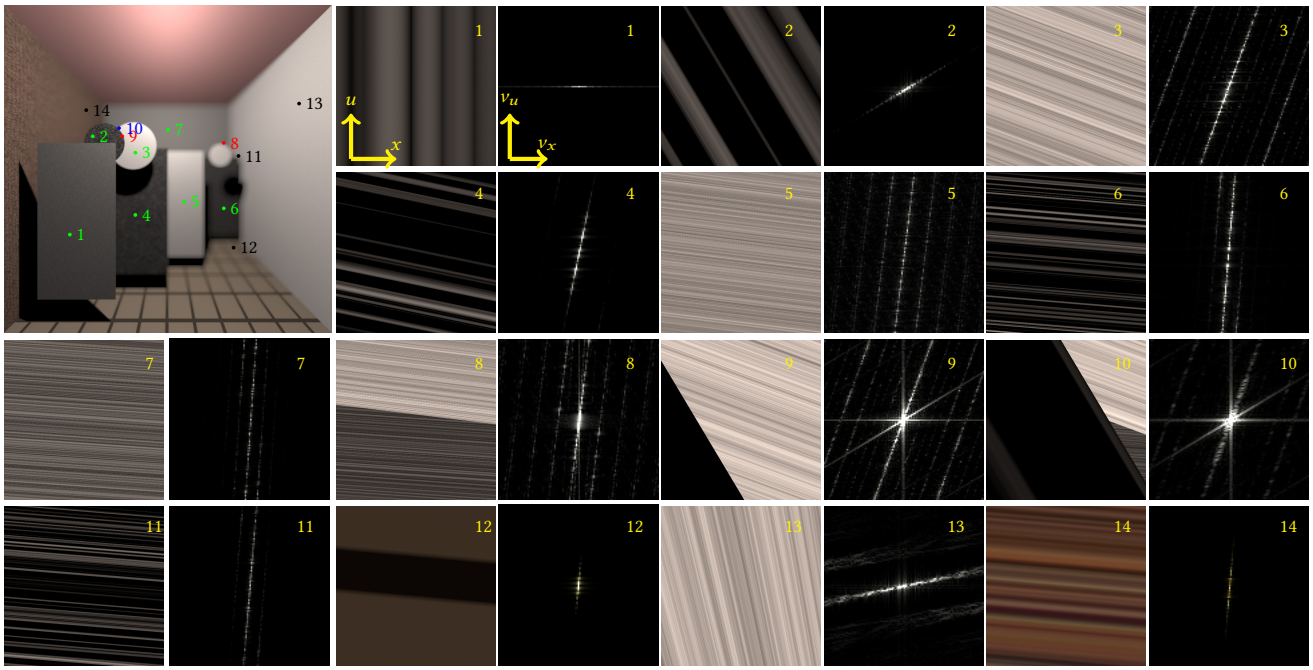


Fig. 11. Fourier content of different pixels is analyzed for this Cornell Box scene which is rendered with a defocus blur. The XU light field slice corresponding to each pixel is shown with the corresponding expected power spectrum of that slice. On the Cornell Box image, the *green dots* refer to **single depth** pixels, *red dots* refer to **two depth** pixels, *blue dot* refer to a pixel with **three depths** whereas the *black dots* refer to the pixels with continously varying depth.

which is a form of randomly shuffled *jitter* that stratifies the XY and UV planes with jittered 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 (Sinc(·)) 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 Sinc(·). 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 $O\left(N^{-1.5}\right)$ 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 $O\left(N^{-2}\right)$. 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 $O\left(N^{-1.5}\right)$ 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

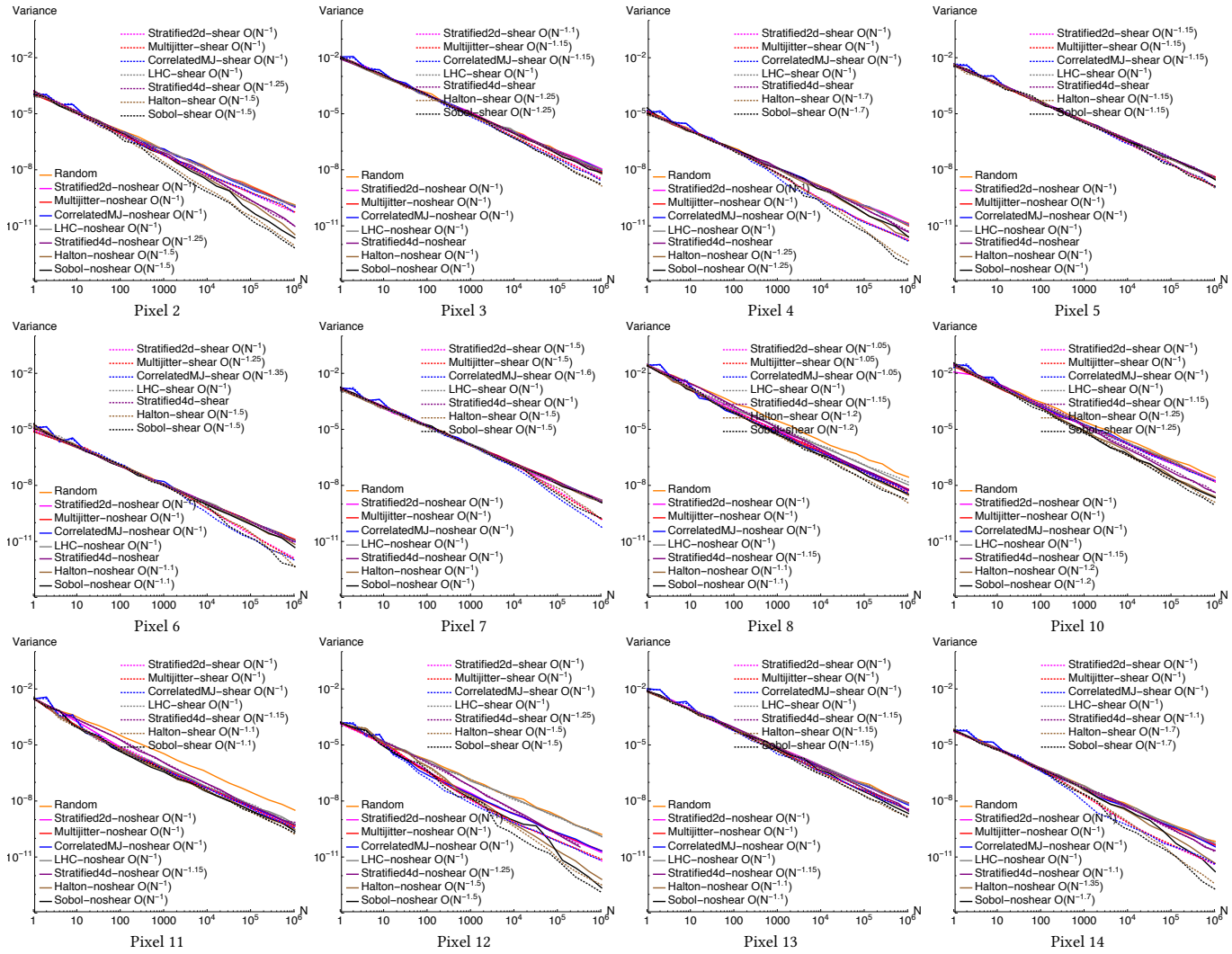


Fig. 12. Variance convergence plots are shown for all the pixels marked in the Cornell Box scene from Fig. 11. Pixels numbered from 2 to 7 are single depth pixels, whereas, Pixels ranging from 11 to 14 (in the bottom row) have continuously varying depth. Pixel 8 has one occluder (two depths) whereas Pixel 10 has two occluders (three depths). Shearing shows (asymptotic) improvement in almost all the scenarios with some pixels showing remarkable improvement in the convergence rates (Pixel 3, 4, 6, 7, 14). Note that, LHC correspond to Latin-hypercube sampler in the plots. Better illustration of these plots is shown in the accompanied supplemental.

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× 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

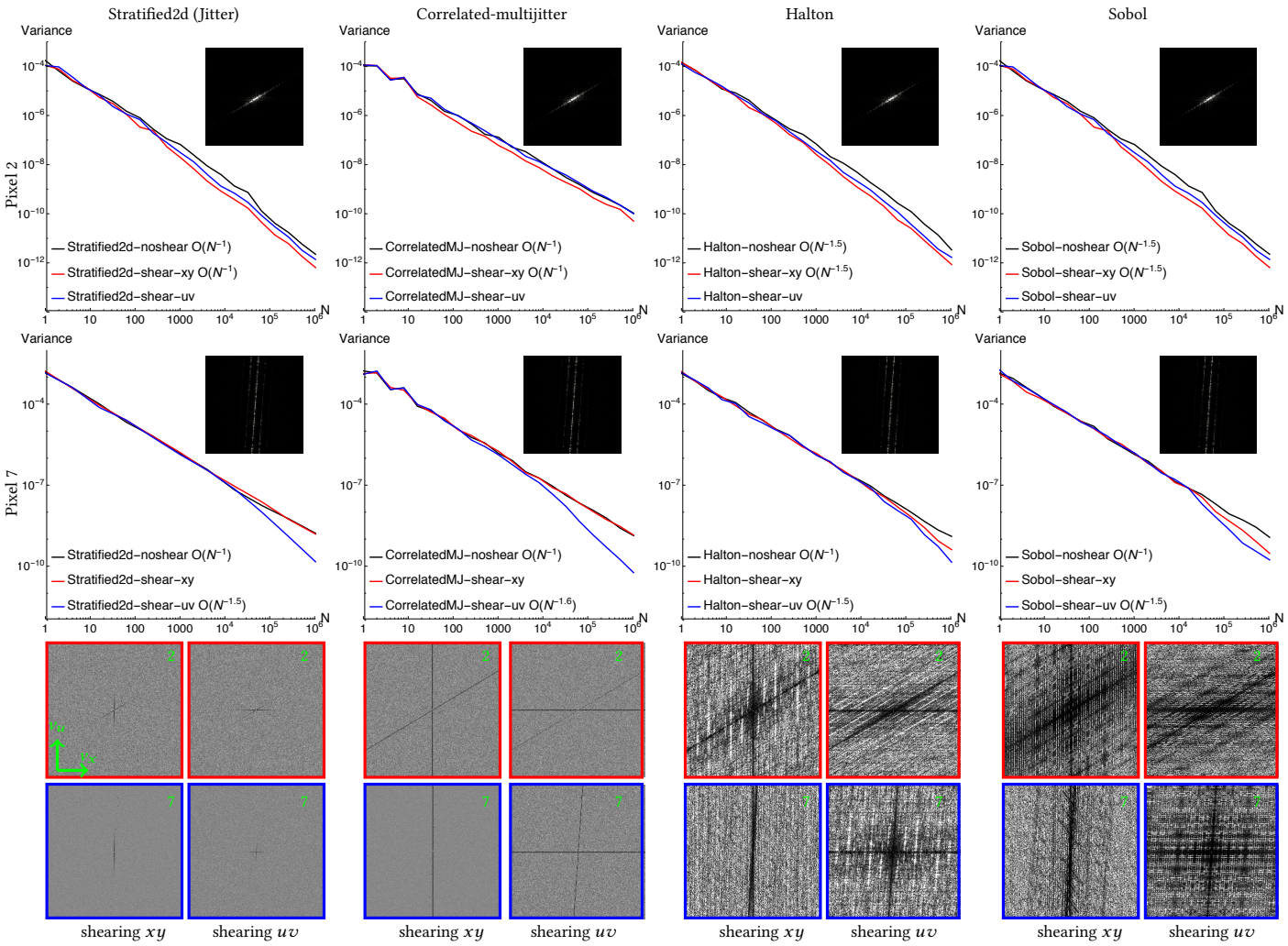


Fig. 13. **Pixels 2 & 7:** Variance convergence due to shearing the samples depends on the angle of the integrand power specturm from the horizontal v_x axis, which is directly correlated to the depth of the objects underneath each pixel. Pixel 2 from Fig. 11, which is very close to the focal plane directly benefits from shearing xy samples, whereas, Pixel 7 which is far from the focal plane shows improvement in convergence only for the uv samples.

2D convergence of $O\left(N^{-1.5}\right)$ whereas correlated-multijitter gives $O\left(N^{-1.6}\right)$ convergence for this 4D (x, y, u, v) integral. Halton and Sobol also show similar improvements in convergence for Pixel 7. As expected, shearing stratified4d does not show improvement in convergence and we obtain the 4D convergence rate of $O\left(N^{-1.25}\right)$. 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.

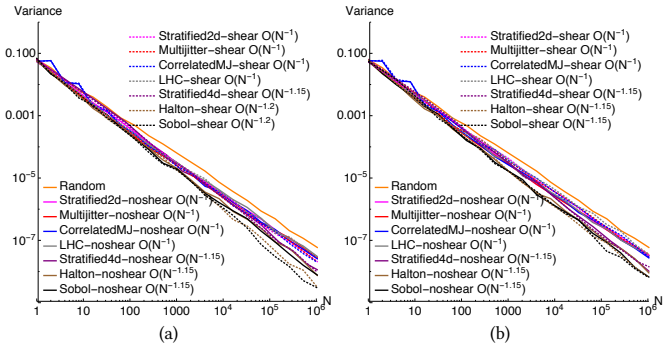


Fig. 14. **Pixel 9** has two depths which are visible as two high energy streaks forked from the DC (center of the image) in the spectrum (as shown in Fig. 11) with each line correspond to each depth. In (a) convergence is plotted by shearing the samples to align the sampling spectrum with the high energy orientation ($> 45°$), whereas in (b), sampling spectrum is aligned to the other orientation ($< 45°$).

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.

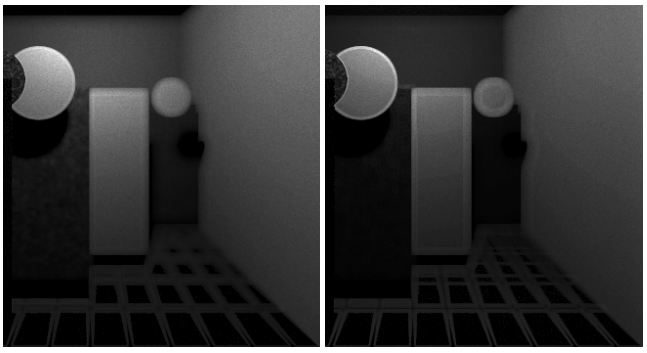
Pixel 3 & Pixel 5 show nominal improvements in the convergence. This is because the energy content of these relatively bright pixels is spread across higher frequencies compared to other pixels and therefore, the hairline anisotropic structures of the sampling spectra are not significantly improving 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 $O\left(N^{-1.7}\right)$ with Halton and Sobol after shearing, Pixel 6 shows convergence rates of $O\left(N^{-1.25}\right)$ with multijitter, $O\left(N^{-1.35}\right)$ with correlated-multijitter and $O\left(N^{-1.5}\right)$ 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 $< 45°$ from v_x whereas, the object behind orients at an angle $> 45°$. 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 Fig. 14 (a) and xy w.r.t. uv 14 (b)). Since Halton and Sobol have *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 LHC) show 10× improvement in variance for Pixel 12 at higher sample counts, whereas for Pixel 14 not only Sobol, multijitter, correlated-multijitter but also stratified2d show 10× improvement in variance (see supplemental material for better visualization of these variance plots). Halton's convergence improves from $O\left(N^{-1.35}\right)$ to $O\left(N^{-1.7}\right)$ 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



(a) Correlated-multjitter with no shear (b) Correlated-multijitter with shear

Fig. 15. Following our algorithm of shearing samples prior to MC integration, we observe modest improvement in variance (b) at depth discontinuities, and dramatic improvement in variance at other regions when rendered using $N = 100489$ sheared correlated-multijittered samples at each pixel. Brighter pixels correspond to relatively high variance value.

& UV subspaces which are then randomly permuted to solve the 4D (x, y, u, v) defocus integral. Our shearing algorithm reliably decreases variance across the whole image. Constant depth regions (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 improvements 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 predicting 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 reconstruction approaches and the recent work by Georgiev and Fajardo [2016]—which exploits inter-pixel correlation to improve the spectral 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 shuffling 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:

$$\mathcal{R}_a^b(\mathbf{n}_k) = \int_a^b \rho^{d-1} \langle \mathcal{P}_S(\rho \mathbf{n}_k) \rangle \mathcal{P}_f(\rho \mathbf{n}_k) \mathrm{d}\rho \,, \tag{19}$$

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

$$Var(\boldsymbol{I}_N) < \mathcal{R}_0^\infty(\mathbf{n}_1) + \sum_{k=0}^{m} \mathcal{R}_0^\infty(\mathbf{n}_k) \,. \tag{20}$$

Lets assume that we have a *constant* radial profile along direction $\mathbf{n}_1$ and a monomial behaviour of degree $b_k > 0$ along all other directions. This would allow us to approximate $\mathcal{R}_0^\infty(\mathbf{n}_1)$ as $O\left(N^{-1}\right)$ irrespective of the integrand. The second part $\mathcal{R}_0^\infty(\mathbf{n}_k)$ can be expanded similar to Eq. (11) as follows:

$$Var(\boldsymbol{I}_N) < O\left(N^{-1}\right) + \lim_{m\to\infty} \sum_{k=2}^{m} (\mathcal{R}_0^{\rho_k}(\mathbf{n}_k) + \mathcal{R}_{\rho_k}^\infty(\mathbf{n}_k)) \tag{21}$$

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

$$Var(I_N) < \begin{cases} \sum_{k=2}^{m} O\left(\frac{1}{N N_k{}^{b_k}}\right) + O\left(\frac{1}{N}\right) & 0 < b_k < 1 \\ \sum_{k=2}^{m} O\left(\frac{1}{N N_k}\right) + O\left(\frac{1}{N}\right) & b_k \geq 1 \end{cases} . \tag{22}$$

In a similar fashion, we can obtain the variance convergence rates for the best case, in all $\mathbf{n}_k / \{k = 1\}$ directions:

$$Var(I_N) < \sum_{k=2}^{m} O\left(\frac{1}{N N_k{}^{b_k}}\right) + O\left(\frac{1}{N}\right). \tag{23}$$

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: $Var(I_N) < O\left(\frac{1}{N}\right)$.

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:

$$\boxed{Var(I_N) < O\left(\frac{1}{N N_k^{b_k}}\right) \ \forall \ \underset{\text{infimum}}{b_k > 0}} \tag{24}$$

□

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.

Abdalla G. M. Ahmed, Hélène Perrier, David Coeurjolly, Victor Ostromoukhov, Jianwei Guo, Dong-Ming Yan, Hui Huang, and Oliver Deussen. 2016. Low-discrepancy Blue Noise Sampling. *ACM Trans. Graph. (Proc. SIGGRAPH)* 35, 6, Article 247 (Nov. 2016), 13 pages.

Pierre Alliez, David Cohen-Steiner, Olivier Devillers, Bruno Lévy, and Mathieu Desbrun. 2003. Anisotropic Polygonal Remeshing. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)* 22, 3 (July 2003), 9.

Tom M. Apostol. 1974. *Mathematical Analysis.* Addison-Wesley Publishing Company.

Michael Balzer, Thomas Schlömer, and Oliver Deussen. 2009. Capacity-Constrained Point Distributions: A Variant of Lloyd's Method. *ACM Trans. Graph. (Proc. SIGGRAPH)* 28, 3 (2009).

Laurent Belcour, Cyril Soler, Kartic Subr, Nicolas Holzschuch, 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.

Luca Brandolini, Leonardo Colzani, and Andrea Torlaschi. 2001. Mean square decay of Fourier transforms in Euclidean and non Euclidean spaces. *Tohoku Math. J. (2)* 53, 3 (2001).

Luca Brandolini, Steve Hofmann, and Alex Iosevich. 2003. Sharp rate of average decay of the Fourier transform of a bounded set. *Geometric & Functional Analysis GAFA* 13, 4 (2003).

Robert L. Cook. 1986. Stochastic Sampling in Computer Graphics. *ACM Transactions on Graphics* 5, 1 (Jan. 1986).

Robert L. Cook, Thomas Porter, and Loren Carpenter. 1984. Distributed Ray Tracing. *Computer Graphics (Proc. SIGGRAPH)* 18, 3 (Jan. 1984).

Fernando de Goes, Katherine Breeden, Victor Ostromoukhov, and Mathieu Desbrun. 2012. Blue Noise through Optimal Transport. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)* 31 (2012). Issue 6.

Mark A. Z. Dippé and Erling Henry Wold. 1985. Antialiasing Through Stochastic Sampling. *Computer Graphics (Proc. SIGGRAPH)* (1985).

Mark A. Z. Dippé and Erling Henry Wold. 1992. Progress in Computer Graphics (Vol. 1). Ablex Publishing Corp., Norwood, NJ, USA, Chapter Stochastic Sampling: Theory and Application.

Frédo Durand. 2011. *A frequency analysis of Monte-Carlo and other numerical integration schemes.* Technical Report TR-2011-052. MIT CSAIL.

Frédo Durand, Nicolas Holzschuch, Cyril Soler, Eric Chan, and François X. Sillion. 2005. A Frequency Analysis of Light Transport. *ACM Trans. Graph. (Proc. SIGGRAPH)* 24, 3 (July 2005), 1115–1126.

Kevin Egan, Frédo Durand, and Ravi Ramamoorthi. 2011. Practical Filtering for Efficient Ray-traced Directional Occlusion. *ACM Trans. Graph.* 30, 6, Article 180 (Dec. 2011), 10 pages.

Kevin Egan, Yu-Ting Tseng, Nicolas Holzschuch, Frédo Durand, and Ravi Ramamoorthi. 2009. Frequency Analysis and Sheared Reconstruction for Rendering Motion Blur. *ACM Trans. Graph. (Proc. SIGGRAPH)* 28, 3, Article 93 (July 2009), 13 pages.

Louis Feng, Ingrid Hotz, Bernd Hamann, and Kenneth I Joy. 2008. Anisotropic noise samples. *Visualization and Computer Graphics, IEEE Transactions on* 14, 2 (2008).

Andrea Gabrielli and Salvatore Torquato. 2004. Voronoi and void statistics for superhomogeneous point processes. *Physical Review E* 70, 4 (2004).

Iliyan Georgiev and Marcos Fajardo. 2016. Blue-noise Dithered Sampling. In *ACM SIGGRAPH 2016 Talks.* ACM, New York, NY, USA, Article 35, 1 pages.

Toshiya Hachisuka, Wojciech Jarosz, Richard Peter Weistroffer, Kevin Dale, Greg Humphreys, Matthias Zwicker, and Henrik Wann Jensen. 2008. Multidimensional Adaptive Sampling and Reconstruction for Ray Tracing. *ACM Trans. Graph. (Proc. SIGGRAPH)* 27, 3, Article 33 (Aug. 2008), 10 pages.

John H. Halton. 1960. On the Efficiency of Certain Quasi-random Sequences of Points in Evaluating Multi-dimensional Integrals. *Numer. Math.* 2, 1 (Dec. 1960), 7.

Daniel Heck, Thomas Schlömer, and Oliver Deussen. 2013. Blue Noise Sampling with Controlled Aliasing. *ACM Transactions on Graphics* 32, 3 (2013).

Janine Illian, P. Antti Penttinen, Helga Stoyan, and Dietrich Stoyan. 2008. *Statistical Analysis and Modelling of Spatial Point Patterns.* Wiley.

Howard J. Keisler. 2012. *Elementary Calculus: An Infinitesimal Approach.* Dover Publications.

Andrew Kensler. 2013. Correlated multi-jittered sampling. (2013).

Ares Lagae and Philip Dutré. 2008. A Comparison of Methods for Generating Poisson Disk Distributions. *Comp. Graph. Forum* 27, 1 (2008).

Jaakko Lehtinen, Timo Aila, Jiawen Chen, Samuli Laine, and Frédo Durand. 2011. Temporal Light Field Reconstruction for Rendering Distribution Effects. *ACM Trans. Graph.* 30, 4, Article 55 (July 2011), 12 pages.

Oscar AZ Leneman. 1966. Random sampling of random processes: Impulse processes. *Information and Control* 9, 4 (1966).

Bruno Lévy and Yang Liu. 2010. Lp Centroidal Voronoi Tessellation and Its Applications. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)* 29, 4, Article 119 (July 2010), 11 pages.

Hongwei Li, Li-Yi Wei, Pedro V. Sander, and Chi-Wing Fu. 2010. Anisotropic Blue Noise Sampling. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)* 29, 6 (Dec. 2010).

Soham Uday Mehta, Brandon Wang, and Ravi Ramamoorthi. 2012. Axis-aligned Filtering for Interactive Sampled Soft Shadows. *ACM Trans. Graph. (Proc. SIGGRAPH)* 31, 6, Article 163 (Nov. 2012), 10 pages.

Soham Uday Mehta, JiaXian Yao, Ravi Ramamoorthi, and Fredo Durand. 2014. Factored Axis-aligned Filtering for Rendering Multiple Distribution Effects. *ACM Trans. Graph. (Proc. SIGGRAPH)* 33, 4, Article 57 (July 2014), 12 pages.

Don P. Mitchell. 1991. Spectrally Optimal Sampling for Distributed Ray Tracing. *Computer Graphics (Proc. SIGGRAPH)* 25, 4 (1991).

Don P. Mitchell. 1996. Consequences of Stratified Sampling in Graphics. In *Annual Conference Series (Proc. SIGGRAPH).* ACM.

A. Cengiz Öztireli. 2016. Integration with Stochastic Point Processes. *ACM Transactions on Graphics* 35, 5, Article 160 (Aug. 2016), 16 pages.

Matt Pharr, Wenzel Jakob, and Greg Humphreys. 2016. *Physically Based Rendering: From Theory To Implementation* (3rd ed.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Adrien Pilleboue, Gurprit Singh, David Coeurjolly, Michael Kazhdan, and Victor Ostromoukhov. 2015. Variance Analysis for Monte Carlo Integration. *ACM Trans. Graph. (Proc. SIGGRAPH)* 34, 4 (July 2015).

Ravi Ramamoorthi, John Anderson, Mark Meyer, and Derek Nowrouzezahrai. 2012. A Theory of Monte Carlo Visibility Sampling. *ACM Trans. Graph. (Proc. SIGGRAPH)* 31, 5 (2012).

Bernhard Reinert, Tobias Ritschel, Hans-Peter Seidel, and Iliyan Georgiev. 2016. Projective Blue Noise Sampling. *Comp. Graph. Forum* (2016).

Peter .S. Shirley. 1991. Discrepancy as a Quality Measure for Sample Distributions. In *Proc. Eurographics.*

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, 4 (1967), 86 – 112.

Cyril Soler, Kartic Subr, Frédo Durand, Nicolas Holzschuch, and François Sillion. 2009. Fourier Depth of Field. *ACM Transactions on Graphics* 28, 2, Article 18 (May 2009), 12 pages.

Kartic Subr and Jan Kautz. 2013. Fourier Analysis of Stochastic Sampling Strategies for Assessing Bias and Variance in Integration. *ACM Trans. Graph. (Proc. SIGGRAPH)* 32, 4 (July 2013).

Kartic Subr, Derek Nowrouzezahrai, Wojciech Jarosz, Jan Kautz, and Kenny Mitchell. 2014. Error analysis of estimators that use combinations of stochastic sampling strategies for direct illumination. *Comp. Graph. Forum (Proc. EGSR)* 33, 4 (June 2014).

Kartic Subr, Gurprit Singh, and Wojciech Jarosz. 2016. Fourier Analysis of Numerical Integration in Monte Carlo Rendering: Theory and Practice. In *ACM SIGGRAPH 2016 Courses.* ACM, New York, NY, USA, Article 10, 40 pages.

Robert Ulichney. 1987. *Digital Halftoning.* MIT Press, Cambridge, MA, USA.

Karthik Vaidyanathan, Jacob Munkberg, Petrik Clarberg, and Marco Salvi. 2015. Layered Light Field Reconstruction for Defocus Blur. *ACM Transactions on Graphics* 34, 2, Article 23 (March 2015), 12 pages.

Florent Wachtel, Adrien Pilleboue, David Coeurjolly, Katherine Breeden, Gurprit Singh, Gaël Cathelin, Fernando de Goes, Mathieu Desbrun, and Victor Ostromoukhov. 2014. Fast Tile-Based Adaptive Sampling with User-Specified Fourier Spectra. *ACM Trans. Graph. (Proc. SIGGRAPH)* 33, 4 (2014).

Li-Yi Wei, Chia-Kai Liang, Graham Myhre, Colvin Pitts, and Kurt Akeley. 2015. Improving Light Field Camera Sample Design with Irregularity and Aberration. *ACM Trans. Graph. (Proc. SIGGRAPH)* 34, 4, Article 152 (July 2015), 11 pages.

Li-Yi Wei and Rui Wang. 2011. Differential Domain Analysis for Non-uniform Sampling. *ACM Trans. Graph.* 30, 4, Article 50 (July 2011), 10 pages.

Turner Whitted. 1980. An Improved Illumination Model for Shaded Display. *Commun. ACM* 23, 6 (June 1980), 343–349.

Yahan Zhou, Haibin Huang, Li-Yi Wei, and Rui Wang. 2012. Point sampling with general noise spectrum. *ACM Trans. Graph. (Proc. SIGGRAPH)* 31, 4 (2012).

Matthias Zwicker, Wojciech Jarosz, Jaakko Lehtinen, Bochang Moon, Ravi Ramamoorthi, Fabrice Rousselle, Pradeep Sen, Cyril Soler, and Sung-Eui Yoon. 2015. Recent Advances in Adaptive Sampling and Reconstruction for Monte Carlo Rendering. *Comp. Graph. Forum (Proc. Eurographics)* 34, 2 (May 2015), 667–681.