

# Integration with Stochastic Point Processes

A. CENGİZ ÖZTİRELİ  
ETH Zurich

We present a novel comprehensive approach for studying error in integral estimation with point distributions based on point process statistics. We derive exact formulae for bias and variance of integral estimates in terms of the spatial or spectral characteristics of integrands, and first and second order product density measures of general point patterns. The formulae allow us to study and design sampling schemes adapted to different classes of integrands by analyzing the effect of sampling density, weighting, and correlations among point locations separately. We then focus on non-adaptive correlated stratified sampling patterns and specialize the formulae to derive closed-form and easy-to-analyze expressions of bias and variance for various stratified sampling strategies. Based on these expressions, we perform a theoretical error analysis for integrands involving the discontinuous visibility function. We show that significant reductions in error can be obtained by considering alternative sampling strategies instead of the commonly used random jittering or low discrepancy patterns. Our theoretical results agree with and extend various previous results, provide a unified analytic treatment of point patterns, and lead to novel insights. We validate the results with extensive experiments on benchmark integrands as well as real scenes with soft shadows.

Categories and Subject Descriptors: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—*Monte Carlo integration, ray tracing, shading*; G.1.4 [Numerical Analysis]: Quadrature and Numerical Differentiation—*Error analysis*

General Terms: Theory, Sampling, Quadrature, Rendering

Additional Key Words and Phrases: Monte Carlo sampling, integration, quadrature, stochastic sampling, point processes, stratified sampling, ray tracing, soft shadows

## 1. INTRODUCTION

Integrals of complex functions need to be estimated for rendering realistic scenes. The estimation is typically carried out by sampling the integrand at particular locations in space/time and taking a weighted average of the computed values. This method is general and can be made to converge to the true integral as more samples are taken. However, a prohibitively large number of samples might be needed to synthesize plausible images without noticeable noise. Hence, it is very important to optimally and adaptively choose the *weighting*, *density*, and arrangement of sampling points given by the interactions or *correlations* among sample locations.

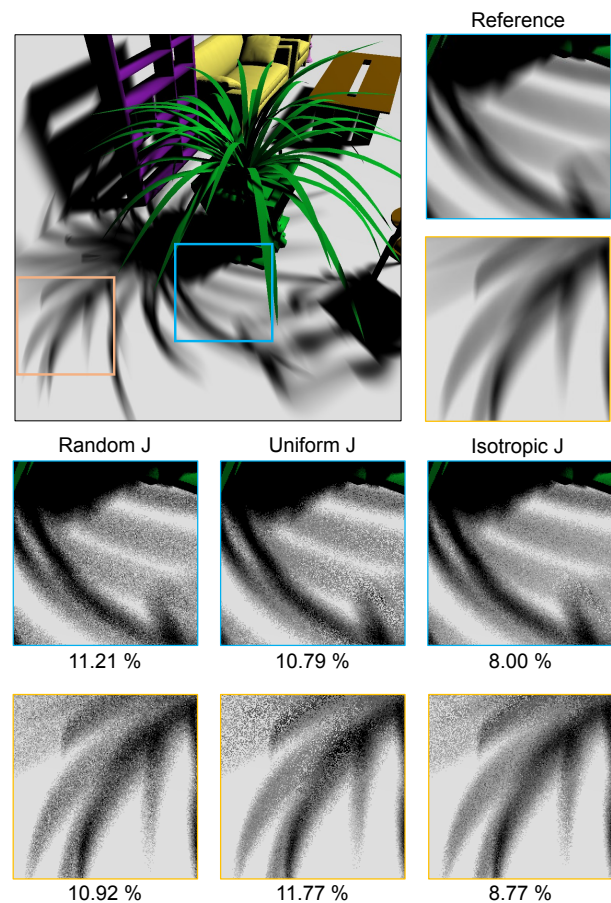


Fig. 1. Based on the proposed theoretical framework, we derive theoretical conditions under which certain sampling patterns will perform better. We show that alternative sampling patterns, such as *isotropic jittering* in this example, can result in errors lower than possible before under certain conditions on the scene geometry and light types. (The scene is rendered with a circle light source and an average of 7 visibility samples per pixel (spp)). The RMS errors are reported at the bottom of each inset.)

Functions with different characteristics need to be integrated for different applications. For example, rendering soft shadows requires an integration over the area of light sources involving the discontinuous occlusion function. Ideally, a sampling pattern should be studied and adapted according to the characteristics of the integrands for accurate estimations. However, such an analysis is challenging if correlations among point locations are taken into account, since deriving exact expressions of error tailored for general globally correlated sampling patterns has so far not been possible.

Indeed, most adaptive sampling techniques in rendering focus on adjusting the sampling density and weights according to the variance or the spectral bandwidths of the radiance to be integrated [Whitted 1980; Mitchell 1987; 1991; Hachisuka et al. 2008; Belcour et al. 2013; Ramamoorthi and Hanrahan 2004; Durand et al. 2005; Soler et al. 2009; Egan et al. 2009; Egan et al. 2011; Egan et al. 2011; Mehta et al. 2012]. However, as demonstrated in recent works [Durand 2011; Ramamoorthi et al. 2012; Subr and Kautz 2013; Subr et al. 2014], adapting weighting, density, and correlations to integrands simultaneously can significantly reduce the variance of integral estimates. Hence, it is very important to take correlations along with density and weighting into account when studying a point pattern for integration. Such a general theoretical analysis cannot be carried out with the existing approaches.

In this paper, we present a new theoretical framework that allows us to study and adapt general point distributions for reducing error in integration. Utilizing point process statistics, we derive exact bias and variance formulae for general distributions in terms of only the first and second order characteristics of the underlying point processes, and the spatial or spectral properties of integrands. We then specialize these formulae to stratified sampling patterns and derive closed-form expressions that are amenable to analysis and inference. We apply the theory to integrands with discontinuities and those involving the visibility function. In summary, the main contributions of this work are:

- Introduction of the most comprehensive theoretical framework for studying bias and variance of numerical integration with point distributions based on point processes. The framework (1) proves that only first and second order correlations contribute to error (2) allows us to analyze the effect of weighting, density, and correlations separately based on closed-form tailored expressions for bias and variance, (3) expresses bias and variance in terms of local or global, spatial or spectral properties of integrands, making analysis and inference easier, (4) applies to all point distributions where intensity and second order product density measures can be analytically computed or estimated.
- Closed-form and simple formulae for bias and variance of various globally *correlated* jittered sampling patterns. These include the more commonly used patterns as well as alternative patterns, which are chosen based on the insights gained from the variance formulae. Such exact formulae and analysis have so far only been possible for uncorrelated jittering patterns [Subr et al. 2014], or for 1D binary integrands under certain assumptions on the correlations [Ramamoorthi et al. 2012].
- An application of the theoretical framework to visibility sampling for computing soft shadows due to planar area lights. We analytically explain how light types and scene geometry affect the error, and when certain patterns perform better than others, by an analysis in the joint pixel-light space. We show that significant reductions in error can be obtained by considering alternative sampling patterns.

## 2. RELATED WORK

Our work relates to the fields of point processes and sampling patterns in computer graphics, in particular those used for integration in rendering applications.

**Point processes** The field of point processes provides powerful theoretical tools to study point distributions. A point process is defined as the generating process of a point pattern [Møller and Waagepetersen 2004; Illian et al. 2008]. Each realization of a point process is thus a point distribution. A point process can be characterized by considering correlations among point locations and associated marks. In particular, first and second order correlations, intuitively quantifying density and arrangement of points, have been proven to be sufficient to uniquely identify most processes [Illian et al. 2008; Öztireli and Gross 2012]. Recently, pair correlation function, measuring the probability of finding a pair of points at particular distances for isotropic point processes, has been introduced to analyze and synthesize point distributions with desired characteristics for graphics applications [Öztireli and Gross 2012; Heck et al. 2013]. Our theoretical framework is also based on point processes, with the goal of studying error in integration. We utilize Campbell's theorem (e.g. [Illian et al. 2008]) that gives expectations of sums of functions sampled at point locations in terms of integrals of the functions and statistical measures of the underlying point process. We show that bias and variance of a numerical integration scheme can be expressed in terms of first and second order measures and weighting. This offers a unified treatment of point distributions for integration, and closed-form and specific bias and variance expressions for different distributions.

**Analysis of point distributions** Several other analysis tools have been developed to understand the behavior of point distributions when used for reconstruction, object placement, stippling, simulation, or integration. A commonly used measure is the periodogram, computed by averaging the power spectra corresponding to different realizations of a point pattern. It is used extensively for determining antialiasing properties of distributions [Mitchell 1987; Lagae and Dutré 2008; Heck et al. 2013]. Since regularity and uniformity of distributions can be easily read from the shape of the spectrum, it has also been utilized to determine if point patterns are suitable for object placement [Wei 2010], stippling and halftoning [Schmaltz et al. 2010; Fattal 2011], among other applications. A similar analysis can also be performed on non-Euclidean domains and for non-constant densities [Bowers et al. 2010; Wei and Wang 2011]. Recently, Subr and Kautz [2013] has shown that bias and an upper bound on variance of numerical integration can be expressed in terms of the averaged Fourier spectrum of general point distributions. The bounds were further refined for the special case of homogenous distributions in a concurrent work [Pilleboue et al. 2015]. We also derive bias and variance of numerical integrators for point distributions. However, instead of spectral bounds, we strive to provide tailored closed-form expressions for general sampling patterns, which allows us to study how different patterns will perform for integrands with certain properties. Our framework also allows studying the effects of weighting, density, and correlations separately.

Apart from spectral diagrams, measures from point processes [Illian et al. 2008; Öztireli and Gross 2012; Heck et al. 2013], spatial characteristics such as the minimum distance between pairs of points [Lagae and Dutré 2008], or discrepancy [Shirley 1991] have been proposed for the analysis of point distributions in graphics. Among these spatial characteristics, point process statistics provide the most descriptive power. Building on the theory of point pro-

cesses, our approach allows us to assess error for integration, and to choose, adapt, and analyze distributions in relation to integrands.

**Sampling patterns for integration** The variance in Monte Carlo estimation with random point locations can be reduced by ensuring a more uniform distribution of points. Quasi-Monte Carlo methods achieve this by generating deterministic low discrepancy sampling patterns [Niederreiter 1992; Dick and Pillichshammer 2010; Keller et al. 2012]. Uniformity of distributions can be further improved while avoiding aliasing by using blue noise distributions where points are randomly distributed and each pair of points is a certain distance apart [Cook 1986; Lagae and Dutré 2008; Heck et al. 2013]. A large body of works exists to generate such distributions efficiently (e.g. [Dunbar and Humphreys 2006; Wei 2008; Ebeida et al. 2012]). However, generation of blue noise distributions is still more involved than the commonly used jittered distributions for integration. Jittered distributions follow the idea of stratified sampling, where the integration domain is divided into subdomains. They are obtained by dividing the domain into cells of equal size and placing a single point in each cell following a certain rule or probability distribution. Jittering provides high convergence rates and low computational complexity [Mitchell 1996; Owen 2013], can be combined with other variance reduction methods [Owen 2013; Subr et al. 2014], and hence is a popular choice in rendering. We derive exact expressions for bias and variance of various uncorrelated and globally correlated jittered sampling strategies to study their performance for different types of integrands. Following this analysis, we propose to use alternative correlated jittered distributions that can provide lower errors when computing integrals involving the visibility function. Our study extends the previous works [Ramamoorthi et al. 2012] that analyze how light types affect the performance of random and uniform jittering.

Correlations can be controlled further to reduce the error in integration. A prominent example is the method of antithetic sampling [Hammersley and Morton 1956], where the sampling points are generated in pairs to capture the negatively correlated parts of the integrand. However, this strategy can also increase the error if the correlated point pairs are not adapted to the integrand. Recent variants solve this problem with locally correlated pairs by combining antithetic sampling with low-discrepancy patterns [Owen 2008], or jittered sampling [Owen 2013; Subr et al. 2014]. We study a local variant of antithetic sampling and its performance for visibility sampling within our framework.

**Adapting density of sampling** Variance of an integral estimator can be reduced by adapting the density of points to integrands. In rendering, density adaptation has been extensively used to dramatically reduce sample counts [Whitted 1980; Mitchell 1987; 1991; Kirk and Arvo 1991; Hachisuka et al. 2008; Overbeck et al. 2009; Pharr and Humphreys 2010]. Following the frequency analysis of light transport [Ramamoorthi and Hanrahan 2004; Durand et al. 2005], several works adapt the sampling rate and density according to the derived bandwidths for efficient computation of soft shadows [Egan et al. 2011; Mehta et al. 2012], depth-of-field effects [Soler et al. 2009; Belcour et al. 2013], motion blur [Egan et al. 2009; Belcour et al. 2013], and directional occlusion [Egan et al. 2011]. In addition to density, our theoretical framework allows us to analyze and adapt correlations in sampling patterns to integrands.

### 3. PRELIMINARIES

We start with the fundamental definitions, theorems, and notation we use in the rest of the paper. We first define the numerical estimation scheme we are considering. Since we model the point distri-

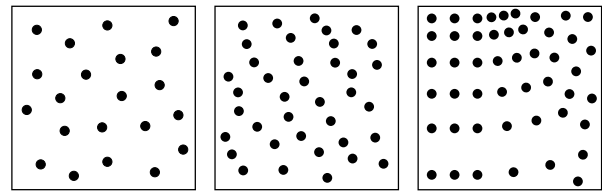


Fig. 2. An isotropic point process (left) generates point distributions with characteristics invariant to rotations. The more general stationary point process (middle) allows for directional dependency, while still producing translation invariant correlations and constant intensity. A general point process (right) can have different correlations and intensity in different parts of the space.

butions used for integration with point processes, we then provide the relevant definitions and Campbell’s theorem that allows us to write expected values over point distributions generated by a point process as integrals.

#### 3.1 Numerical Integration

Given a function  $f : \mathbb{R}^d \rightarrow \mathbb{R}^+$ , we would like to study the error in estimating the integral  $I := \frac{1}{|V|} \int_V f(\mathbf{x}) d\mathbf{x}$ , where  $V$  is the support of the function  $f$  such that  $\forall \mathbf{x} \notin V, f(\mathbf{x}) = 0$ , and  $|V|$  is its volume. We consider a general estimator of the form:

$$\hat{I} := \sum_{i=1}^n w_i f(\mathbf{x}_i), \quad (1)$$

for some positive weights  $w_i$  and sample points  $\mathbf{x}_i$ . We will assume that the support is a unit hypercube,  $|V| = 1$  in the rest of the paper for simplicity. The error in this estimator can be calculated by considering the bias  $bias(\hat{I}) = I - \mathbb{E}\hat{I}$  and variance  $var(\hat{I}) = \mathbb{E}\hat{I}^2 - (\mathbb{E}\hat{I})^2$  with  $\mathbb{E}$  denoting expectation. It is essential to choose the distribution of points and weights carefully so as to minimize the error. Such characteristics of point distributions can be conveniently studied within the framework of point processes, as we describe next.

#### 3.2 Point Processes

Point processes describe generating processes underlying observed point distributions [Møller and Waagepetersen 2004; Illian et al. 2008]. Each distribution can thus be considered as a realization of a point process. A point process can be a mathematical model for a complex process such as the environmental conditions that produce the distribution of trees in a forest, or defined directly in terms of an algorithm such as placing points randomly in space with a minimum distance between them. The distributions generated by a point process can be different but share common characteristics. The goal of point process statistics is to analyze these common characteristics to understand the underlying patterns. In the scope of this paper, we will consider point processes in  $\mathbb{R}^d$ .

A point process  $P$  can be defined by considering the joint probabilities of finding points generated by this process in infinitesimal volumes. One way to express these probabilities is via *product densities*. Let  $\mathbf{x}_i$  denote some arbitrary points in  $\mathbb{R}^d$ , and  $B_i$  infinitesimal spheres centered at these points with volumes  $dx_i = |B_i|$ . Then the  $n^{th}$  order product density  $\varrho^{(n)}$  is defined by  $p(\mathbf{x}_1, \dots, \mathbf{x}_n) = \varrho^{(n)}(\mathbf{x}_1, \dots, \mathbf{x}_n) dx_1 \dots dx_n$ , where  $p(\mathbf{x}_1, \dots, \mathbf{x}_n)$  is the joint probability of having a point of the point process  $P$  in each of the infinitesimal spheres  $B_i$ .

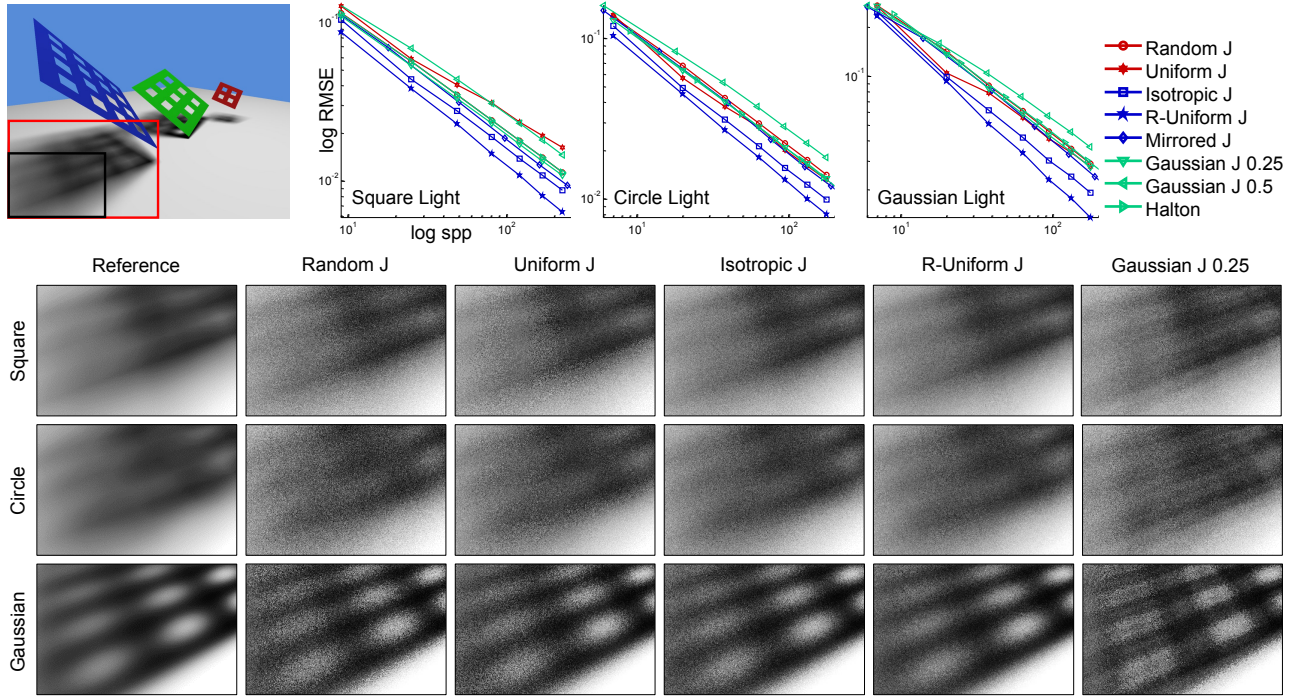


Fig. 3. RMSE vs. spp (average number of visibility samples per pixel) plots for the grids scene with different light sources. Based on our analysis, we propose to use randomly rotated and translated grids (*isotropic jitter sampling*), a variate of antithetic sampling (*mirrored jitter sampling*), and rotated and randomly translated grids (*R-Uniform J*) that can perform better than common patterns for integrals involving the visibility function, depending on the scene geometry and light types. All patterns, except r-uniform jittering, assume no knowledge on the actual integrand. For this scene, on average, isotropic jittering requires 37% less samples than random jittering for square lights, 37% and 44% less samples than uniform jittering for circle and Gaussian lights, respectively. (All error plots in the paper are computed by averaging over 50 instances of the scenes. *Gaussian J* refers to the Gaussian jittering pattern of Subr and Kautz [2013] with indicated parameters. The plots are computed for the area marked with the red rectangle, while the insets are from the area within the black rectangle.)

**Intensity and second order product density** For our derivations of bias and variance in this paper, we will just need the first  $\varrho^{(1)}$  and second  $\varrho^{(2)}$  order product densities. The first order product density is simply given by  $\varrho^{(1)}(\mathbf{x})d\mathbf{x} = p(\mathbf{x})$ . It can be shown that the expected number of points of  $P$  in a set  $B$  can be written as the integral of this expression:  $\mathbb{E}N(B) = \int_B \varrho^{(1)}(\mathbf{x})d\mathbf{x}$ , where  $N(B)$  is the (random) number of points of the process  $P$  in set  $B$ . Hence,  $\varrho^{(1)}(\mathbf{x})$  measures the local expected density of points of a point process. It is thus usually called the *intensity* of  $P$  and denoted by  $\lambda(\mathbf{x}) = \varrho^{(1)}(\mathbf{x})$ . Pairwise correlations are captured by the second order product density  $\varrho^{(2)}(\mathbf{x}, \mathbf{y})d\mathbf{x}d\mathbf{y} = p(\mathbf{x}, \mathbf{y})$ . In general, we need to consider higher order correlations as well to explain general point processes. However, for our case of integral estimation, we will show in Section 4 that the two statistics  $\lambda(\mathbf{x})$  and  $\varrho(\mathbf{x}, \mathbf{y})$  are sufficient to exactly express bias and variance of integration schemes for general point processes.

**Classification of point processes** A point process  $P$  can be defined to consist of finite or infinite number of points. Although observed point distributions are always in a finite window, they can be regarded as parts of an infinite point set that covers the whole space. Translation invariance (*stationary* point process) and rigid motion invariance (*isotropic* point process) of generated point patterns are other important considerations. A distribution generated by a stationary point process has the same characteristics regardless of the location of the observation window. Some examples are isotropic blue noise distributions, uniform jitter sampling, and random distribution. For a stationary point process, the intensity  $\lambda$  be-

comes a constant value independent of  $\mathbf{x}$ , and second order product density becomes a function of the difference between point locations  $\varrho(\mathbf{x} - \mathbf{y})$ , commonly written in terms of the intensity normalized *pair correlation function* (PCF)  $g$  [Öztireli and Gross 2012] as  $\varrho(\mathbf{x} - \mathbf{y}) = \lambda^2 g(\mathbf{x} - \mathbf{y})$ . Isotropy further simplifies it to  $\varrho(\|\mathbf{x} - \mathbf{y}\|)$ . We illustrate examples of distributions generated by different types of point processes in Figure 2. Finally, ergodicity will be an important concept for some of our definitions. Intuitively, for an *ergodic* point process  $P$ , averages in a finite window is equivalent to those defined by generating multiple instances of  $P$ . In other words, we can take a single distribution generated by the ergodic process  $P$ , and estimate the statistics of  $P$  by observing that distribution through a finite window. A familiar example of a non-ergodic point process can be generated by uniformly perturbing a regular grid (i.e. uniform jittering). This is a stationary process [Illian et al. 2008] with a constant  $\lambda$ . However, each realization of this point process is a translated regular grid. Hence, it is not possible to infer the translation invariance of this process by observing a single distribution.

### 3.3 Campbell's Theorem

A common operation needed to analyze point processes is computing expected values of sums of sampled values from functions  $\mathbb{E} \sum f(\mathbf{x}_i)$ , where the expectation is computed over different realizations of the point process  $P$ , and  $\mathbf{x}_1, \dots$  are the points in a given realization. Note that the number of points in a given realization is

also random, and hence is often omitted for the sums. This can also be generalized to functions of more than one variable. Campbell's theorem is a fundamental theorem that relates these sums to integrals of arbitrary positive functions  $f$  and the product densities of  $P$ . In particular, we will work with sums of the form  $\sum f(\mathbf{x}_i)$  and  $\sum_{i \neq j} f(\mathbf{x}_i, \mathbf{x}_j)$  to compute bias and variance of the integral estimate  $\hat{I}$ . These are given by the following expressions:

$$\mathbb{E} \sum f(\mathbf{x}_i) = \int_{\mathbb{R}^d} f(\mathbf{x}) \lambda(\mathbf{x}) d\mathbf{x}, \quad (2)$$

where  $f: \mathbb{R}^d \rightarrow \mathbb{R}^+$ , and

$$\mathbb{E} \sum_{i \neq j} f(\mathbf{x}_i, \mathbf{x}_j) = \int_{\mathbb{R}^d \times \mathbb{R}^d} f(\mathbf{x}, \mathbf{y}) \varrho(\mathbf{x}, \mathbf{y}) d\mathbf{x} d\mathbf{y}, \quad (3)$$

where  $f: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^+$  and under the common assumption [Illian et al. 2008] that no two points in a process can be at the same location almost surely. These equations make it clear that the variance and bias of  $\hat{I}$ , where the integrand is sampled at points in a realization of a point process, can be directly linked to the statistics of that point process. In the next section, we apply the theorem to derive formulae for different types of point processes.

#### 4. BIAS & VARIANCE OF SAMPLING PATTERNS

Campbell's theorem provides us the essential tool to express bias and variance of  $\hat{I}$  in terms of  $\lambda(\mathbf{x})$  and  $\varrho(\mathbf{x}, \mathbf{y})$  of the sampling pattern used to compute  $\hat{I}$ . Such expressions for other estimators in different fields have been derived starting from the Campbell's theorem under certain assumptions [Stoyan and Stoyan 1994; Illian et al. 2008; Guan 2008], which differ from our setting with the estimator in Equation 1. In this section, we will derive formulae for general and stationary processes, when they are used to estimate  $\hat{I}$ .

##### 4.1 General Point Processes

General point processes can be used to model all point distributions utilized for integration. In order to derive bias and variance, we need  $\mathbb{E}\hat{I}$  and  $\mathbb{E}\hat{I}^2$ . We treat weights as sampled from a continuous positive function such that  $w_i = w(\mathbf{x}_i)$ . For brevity, we sometimes write  $f_i = f(\mathbf{x}_i)$ . The expected value can then be easily derived as:

$$\mathbb{E}\hat{I} = \mathbb{E} \sum w(\mathbf{x}_i) f(\mathbf{x}_i) = \int_V w(\mathbf{x}) f(\mathbf{x}) \lambda(\mathbf{x}) d\mathbf{x}. \quad (4)$$

To derive  $\mathbb{E}\hat{I}^2$ , we can rewrite it as  $\mathbb{E} \sum_{i \neq j} w_i f_i w_j f_j + \mathbb{E} \sum (w_i f_i)^2$ . Using Equations 3 and 2 for the first and second terms, respectively:

$$\begin{aligned} \mathbb{E}\hat{I}^2 &= \int_{V \times V} w(\mathbf{x}) f(\mathbf{x}) w(\mathbf{y}) f(\mathbf{y}) \varrho(\mathbf{x}, \mathbf{y}) d\mathbf{x} d\mathbf{y} \\ &\quad + \int_V w^2(\mathbf{x}) f^2(\mathbf{x}) \lambda(\mathbf{x}) d\mathbf{x}. \end{aligned} \quad (5)$$

Based on these equations, we can easily write the expressions for bias  $bias(\hat{I}) = I - \mathbb{E}\hat{I}$  and variance  $var(\hat{I}) = \mathbb{E}\hat{I}^2 - (\mathbb{E}\hat{I})^2$ . Note that the integrals are over the support  $V$  of  $f$ , but we could have equivalently written them over  $\mathbb{R}^d$ . For the rest of the paper, we usually omit the integration domain when it is  $\mathbb{R}^d$ .

**Remarks** The following can be observed from the equations. (1) Bias purely depends on  $\lambda(\mathbf{x})$  and  $w(\mathbf{x})$ , and can always be eliminated by setting  $w(\mathbf{x}) = 1/\lambda(\mathbf{x})$  if  $\lambda(\mathbf{x}) \neq 0$  in  $V$ . (2) The effects of weights  $w$ , density given by  $\lambda$ , and correlations given by

$\varrho$  on variance can be analyzed separately. This is in contrast with spectral approaches [Durand 2011; Subr and Kautz 2013], where these terms are assimilated into a single measure. (3) As long as the statistics  $\lambda$  and  $\varrho$  can be analytically computed or estimated, we can get bias and variance for arbitrary point distributions for given integrands. Higher order correlations need not be considered. (4) For the sums, we did not explicitly write the number of points in  $V$ , since this is a random number in general that depends on  $\lambda$ . However, for most stratified distributions, the number of points in  $V$  stays constant. For blue noise, regular, or hard-core [Illian et al. 2008; Öztireli and Gross 2012] distributions, it is effectively constant when  $\lambda$  is high.

**Relation to spectral analysis** Bias and variance in the equations above depend on the spatial properties of the integrand  $f$ . Some of these derivations can also be performed in the Fourier domain. One way to derive frequency space formulae is to apply Campbell's theorem to the Fourier transform of the function  $s(\mathbf{x}) = \sum w_i \delta(\mathbf{x} - \mathbf{x}_i)$  given by  $\mathcal{F}(\sum w_i \delta(\mathbf{x} - \mathbf{x}_i)) = \sum w_i e^{-i2\pi\omega^T \mathbf{x}_i}$  as follows:  $\mathbb{E} \sum w_i e^{-i2\pi\omega^T \mathbf{x}_i} = \int w(\mathbf{x}) \lambda(\mathbf{x}) e^{-i2\pi\omega^T \mathbf{x}} d\mathbf{x} = (W * \Lambda)(\omega)$ , where  $W$  and  $\Lambda$  are the Fourier transforms of  $w$  and  $\lambda$  respectively, and  $*$  denotes convolution. Then, by using the frequency space expression for bias [Subr and Kautz 2013], we can write  $bias(\hat{I}) = I - \int (W * \Lambda)(\omega) F(-\omega) d\omega$  with  $F = \mathcal{F}(f)$ . A similar derivation can be used to compute  $\mathbb{E}|\mathcal{F}(s)|^2$ , which leads to upper bounds for variance [Subr and Kautz 2013]. Another possibility is to transform the integrals in Equations 4 and 5 into their Fourier domain counterparts [Durand 2011]. However, it is challenging to analyze the error and adapt the sampling patterns to integrands in the frequency domain, especially for correlated patterns and when integrands contain discontinuities, as is the case for the functions we consider in Sections 5.2 and 6. In contrast, working in the spatial domain generates more intuitive, simple, and exact formulae that are easy to analyze, as we will demonstrate for stratified sampling patterns in Section 5. In particular, we can study and adapt local or global properties of sampling patterns in combination with integrands with intuitive and transparent expressions, leading to novel insights.

##### 4.2 Stationary Point Processes

An important special case of point distributions is when the pattern is translation invariant (i.e. stationary, Section 3.2). For this case, the expressions for bias and variance simplify significantly, since we can now write  $\lambda(\mathbf{x}) = \lambda$ , and  $\varrho(\mathbf{x}, \mathbf{y}) = \lambda^2 g(\mathbf{x} - \mathbf{y})$ , where  $g$  is the pair correlation function (Section 3.2). To get an unbiased estimator, we need to set  $w = 1/\lambda$ , as can be observed in Equation 4. Plugging these into the expression for variance, we get

$$\begin{aligned} var(\hat{I}) &= \mathbb{E}\hat{I}^2 - (\mathbb{E}\hat{I})^2 \\ &= \int \int \frac{f(\mathbf{x})}{\lambda} \frac{f(\mathbf{y})}{\lambda} \lambda^2 g(\mathbf{x} - \mathbf{y}) d\mathbf{x} d\mathbf{y} \\ &\quad + \int \frac{1}{\lambda^2} f^2(\mathbf{x}) \lambda d\mathbf{x} - \left( \int \frac{1}{\lambda} f(\mathbf{x}) \lambda d\mathbf{x} \right)^2 \\ &= \frac{1}{\lambda} \int f^2(\mathbf{x}) d\mathbf{x} + \int \int f(\mathbf{x}) f(\mathbf{y}) (g(\mathbf{x} - \mathbf{y}) - 1) d\mathbf{x} d\mathbf{y}. \end{aligned} \quad (6)$$

The second term in Equation 6 can also be expressed in terms of the autocorrelation  $a_f(\mathbf{h}) = [f(\mathbf{x}) * f(-\mathbf{x})](\mathbf{h})$  of  $f$ . By making the change of variables  $\mathbf{h} = \mathbf{x} - \mathbf{y}$ , we can rewrite the second term in Equation 6 as  $\int \int f(\mathbf{x}) f(\mathbf{x} - \mathbf{h}) g(\mathbf{h}) d\mathbf{x} d\mathbf{h} - \int \int f(\mathbf{x}) f(\mathbf{y}) d\mathbf{x} d\mathbf{y} = \int g(\mathbf{h}) \int f(\mathbf{x}) f(\mathbf{x} - \mathbf{h}) d\mathbf{x} d\mathbf{h} - (\int f(\mathbf{x}) d\mathbf{x})^2 = \int a_f(\mathbf{h}) g(\mathbf{h}) d\mathbf{h} -$

$(\int f(\mathbf{x})d\mathbf{x})^2$ , resulting in the expression

$$\text{var}(\hat{I}) = \frac{1}{\lambda} \int f^2(\mathbf{x})d\mathbf{x} + \int a_f(\mathbf{h})g(\mathbf{h})d\mathbf{h} - \left( \int f(\mathbf{x})d\mathbf{x} \right)^2. \quad (7)$$

The first term in Equation 7 gives the  $1/\lambda$  dependency of variance, which cannot be improved by changing the arrangement of points. The second term illustrates that, if we have minimal information about the integrand  $f$ , we need to make the pair correlation function  $g$  as low as possible. This is especially important for  $\mathbf{h}$  with a small length since the autocorrelation  $a_f$  will necessarily vanish to zero as  $\|\mathbf{h}\|$  increases due to the finite support  $V$  of  $f$ . For a random distribution,  $g = 1$  and we get no reduction due to the second term. For distributions with clustered points,  $g$  is high for lower  $\|\mathbf{h}\|$  and  $g > 1$  for most  $\mathbf{h}$  [Öztireli and Gross 2012] and we get extra variance. To make  $g < 1$ , we thus need to increase distances between points and generate hard-core distributions (such as blue noise) or low discrepancy patterns. Note that for such distributions, one cannot arbitrarily increase  $\lambda$  without changing  $g$  since the distances between points will necessarily decrease for increasing number of points. The space of possible  $g$ 's is also rather small [Öztireli and Gross 2012; Heck et al. 2013] and hence it is not trivial to adapt  $g$  to a given  $f$  or class of  $f$ 's.

The form of variance in Equation 7 also makes the connection to spectral analysis clear. Since the Fourier transform of  $g(\mathbf{h})$  is directly related to the average power spectrum of  $\sum \delta(\mathbf{x} - \mathbf{x}_i)$  [Heck et al. 2013], and  $\mathcal{F}(a_f) = |\mathcal{F}(f)|^2$ , we can also express Equation 7 in the frequency domain, leading to the spectral formula in a concurrent work [Pilleboue et al. 2015] for homogenous ( $\equiv$  stationary) patterns (see Section 4.1 for a discussion of spectral and spatial formulae for general distributions, and Appendix D of the supplementary material for a derivation of the formula in Pilleboue et al. [2015]). However, as we will show in the next sections, working with spatial point process statistics allows us to derive specific and intuitive formulae, in contrast to the spectral bounds derived in other works [Subr and Kautz 2013; Pilleboue et al. 2015].

## 5. BIAS & VARIANCE OF STRATIFIED SAMPLING

Jittered stratified sampling, where the integration domain is divided into disjoint strata and a point is picked in each of them, is one of the most commonly used sampling strategies in rendering especially for real-time applications, due to its effectiveness in improving error and the simple, efficient, and parallelizable nature of the generating algorithms [Mehta et al. 2012; Mehta et al. 2013; Mehta et al. 2014; Subr et al. 2014]. In this section, we study bias and variance formulae for various jittered distributions.

The new framework allows us to derive formulae for the full spectrum of jittered sampling patterns. We particularly focus on *correlated jittering*, and for the first time derive variance formulae for globally correlated patterns with inter-strata correlations. This allows us to study patterns in relation to the integrands, and to propose to use alternative patterns that can significantly improve the error for certain classes of integrands.

### 5.1 Variance Formulae

**Random Jitter Sampling** In random jitter sampling, the distribution of the point in each stratum is independent. More precisely, given a set of  $n$  strata  $S_i$  such that  $S_i \cap S_j = \emptyset$  for  $i \neq j$ , the probability density of finding a point in  $S_i$  at a location  $\mathbf{x}$  is given by  $k_i(\mathbf{x})$  (Figure 4, left). This covers standard random jittering, where  $k_i(\mathbf{x}) = n$  for  $\mathbf{x} \in S_i$  (of volume  $1/n$ ) is a uniform dis-

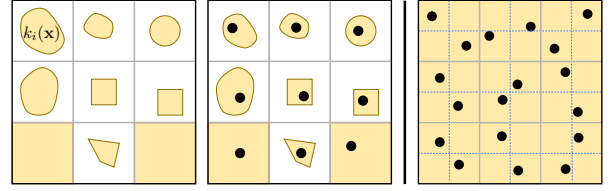


Fig. 4. (Left) A general uncorrelated random jittering where the location of the point in each stratum  $S_i$  follows a distribution  $k_i(\mathbf{x})$  that is independent of those in the other strata. (Right) In mirrored jitter sampling, there are two points in each stratum, one is the mirrored version of the other around the center of the stratum.

tribution within  $S_i$ , and also the recently proposed variants [Subr and Kautz 2013; Subr et al. 2014] as long as the supports of  $k_i$  stay disjoint. Note that due to the independence of the point locations in different strata, the bias and variance formulae for this case could also be derived using standard statistical tools. We include this case as a reference for comparisons and an example for the derivation and utilization of point process statistics.

In order to compute bias and variance, we first need to derive  $\lambda$  and  $\varrho$  for this sampling pattern. The probability of having a point in an infinitesimal  $d\mathbf{x}$  around  $\mathbf{x}$  in  $S_i$  is  $k_i(\mathbf{x})d\mathbf{x}$ . Since the supports of  $k_i$  are disjoint, for an arbitrary  $\mathbf{x}$ , this probability can be written as  $\sum_{i=1}^n k_i(\mathbf{x})d\mathbf{x}$ . Hence, the intensity is given by  $\lambda(\mathbf{x}) = \sum_{i=1}^n k_i(\mathbf{x})$ . Note that the number of points of the underlying process in domain  $V$  (unit square) is always equal to  $n$  in this case. The expected number of points given by the integral of  $\lambda(\mathbf{x})$  over the whole domain  $V$  is again equal to  $n$ , as expected. The joint probability of having a point in  $d\mathbf{x}$  around  $\mathbf{x} \in S_i$  and in  $d\mathbf{y}$  around  $\mathbf{y} \in S_j$  is zero if  $i = j$  (since a stratum can only have a single point), and  $k_i(\mathbf{x})k_j(\mathbf{y})d\mathbf{x}d\mathbf{y}$  if  $i \neq j$ . This probability can be compactly written as:  $[\lambda(\mathbf{x})\lambda(\mathbf{y}) - \sum k_i(\mathbf{x})k_i(\mathbf{y})]d\mathbf{x}d\mathbf{y}$ . Hence,  $\varrho(\mathbf{x}, \mathbf{y}) = \lambda(\mathbf{x})\lambda(\mathbf{y}) - \sum k_i(\mathbf{x})k_i(\mathbf{y})$ .

Plugging these expressions into Equations 4 and 5, the following formulae for bias and variance can be derived (note that this is not a stationary process and hence we cannot use Equation 6):

$$\begin{aligned} \text{bias}(\hat{I}) &= I - \mathbb{E}\hat{I} \\ &= I - \int w(\mathbf{x})f(\mathbf{x}) \sum_{i=1}^n k_i(\mathbf{x})d\mathbf{x} \\ &= I - \sum_{i=1}^n \int w(\mathbf{x})f(\mathbf{x})k_i(\mathbf{x})d\mathbf{x}, \end{aligned} \quad (8)$$

$$\begin{aligned} \text{var}(\hat{I}) &= \mathbb{E}\hat{I}^2 - (\mathbb{E}\hat{I})^2 \\ &= \int w(\mathbf{x})f(\mathbf{x})w(\mathbf{y})f(\mathbf{y}) \left( \lambda(\mathbf{x})\lambda(\mathbf{y}) - \sum_{i=1}^n k_i(\mathbf{x})k_i(\mathbf{y}) \right) d\mathbf{x}d\mathbf{y} \\ &\quad + \int w^2(\mathbf{x})f^2(\mathbf{x})\lambda(\mathbf{x})d\mathbf{x} - \left( \int w(\mathbf{x})f(\mathbf{x})\lambda(\mathbf{x})d\mathbf{x} \right)^2 \\ &= \left( \int w(\mathbf{x})f(\mathbf{x})\lambda(\mathbf{x})d\mathbf{x} \right)^2 - \sum_{i=1}^n \left( \int w(\mathbf{x})f(\mathbf{x})k_i(\mathbf{x})d\mathbf{x} \right)^2 \\ &\quad + \sum_{i=1}^n \int w^2(\mathbf{x})f^2(\mathbf{x})k_i(\mathbf{x})d\mathbf{x} - \left( \int w(\mathbf{x})f(\mathbf{x})\lambda(\mathbf{x})d\mathbf{x} \right)^2 \\ &= \sum_{i=1}^n \int w^2(\mathbf{x})f^2(\mathbf{x})k_i(\mathbf{x})d\mathbf{x} - \left( \int w(\mathbf{x})f(\mathbf{x})k_i(\mathbf{x})d\mathbf{x} \right)^2, \end{aligned} \quad (9)$$

where the integrations for the  $i^{\text{th}}$  term are over  $S_i$ , or equivalently  $V$  or  $\mathbb{R}^d$ . Since the locations of points in different strata are independent, each stratum contributes an independent amount to the total variance. Exactly the same formulae can also be derived using standard definitions of variance and bias due to this independence.

For unbiased estimators,  $w(\mathbf{x}) = 1/\lambda(\mathbf{x}) = 1/\sum_{i=1}^n k_i(\mathbf{x})$ , which is  $w(\mathbf{x}) = 1/k_i(\mathbf{x})$  for  $\mathbf{x} \in S_i$ , since the supports of  $k_i$  are disjoint. By setting  $k_i(\mathbf{x}) = n$  and thus  $w(\mathbf{x}) = 1/n$  in Equation 9, we can write the variance of standard unbiased random jittering, where the sampling point in each stratum is distributed uniformly in the stratum as in Equation T1. In this equation,  $s_i(\mathbf{x}) = f(\mathbf{x} + \mathbf{o}_i)\Pi_S(\mathbf{x})$  is the part of the integrand in  $S_i$ , where  $\Pi_S$  is 1 inside  $S$  and 0 otherwise,  $S$  is the hypercube of length  $1/n^{1/d}$  centered around the origin, and  $\mathbf{o}_i$  is the center of the stratum  $S_i$ .

**Mirrored Jitter Sampling** Some integrands can exhibit negatively correlated parts, which can be exploited to reduce the error in integration. Antithetic sampling [Hammersley and Morton 1956] is a classical method to take advantage of such correlations. The idea is generating a pair of samples, one of which has the reflected position of the other, e.g.  $\mathbf{x}$  and  $1 - \mathbf{x}$ , to capture the negatively correlated parts of the integrand. However, for many integrands, this strategy can also increase the variance. As a remedy, recent works consider combining jittering with antithetic sampling to capture local correlations [Owen 2013; Subr et al. 2014]. We consider a local variate of antithetic sampling that we call mirrored jitter sampling, due to our interpretation of it with mirrored strata. In this pattern, each stratum  $S_i$  contains a random point  $\mathbf{x}$ , and another point at  $-\mathbf{x} + 2\mathbf{o}_i$  that is the mirrored version of the initial point around the stratum center  $\mathbf{o}_i$  (Figure 4, right). The resulting variance formula is derived in Appendix A.1 and shown in Table I, Equation T2. Note that in the formula,  $m = n/2$ , since there are actually only  $m$  strata and  $n = 2m$  points. Due to the first term in this expression, variance will be reduced when  $s_i(\mathbf{x})$ , the local patch of  $f$  in the stratum  $S_i$ , and its mirrored version  $s_i^-(\mathbf{x}) = s_i(-\mathbf{x})$  around the origin of  $S_i$ , are negatively correlated. In Section 6, we will compare the variance of mirrored jitter sampling to that of random jittering, and derive theoretical conditions on when one will perform better for binary integrands.

**Uniform Jitter Sampling** In contrast to random jitter sampling, uniform jitter sampling uses the same random perturbation for all points in different strata. This can be thought of as randomly perturbing a regular grid. Since we consider the unit hypercube as the integration domain, this process is equivalent to assuming an arbitrary perturbation of an infinite grid. Hence, uniform jittering is generated by a stationary process [Illian et al. 2008]. The intensity is simply given by  $\lambda = n$ , and we derive the pair correlation function in Appendix A.2 as

$$g(\mathbf{h}) = \frac{1}{n^2} \sum_{i \neq j} \frac{\delta(\mathbf{h} - \mathbf{h}_{ij})}{a_{\Pi_V}(\mathbf{h}_{ij})} \quad a_{\Pi_V}(\mathbf{h}) \neq 0, \quad (10)$$

where  $\mathbf{h}_{ij} = \mathbf{x}_i - \mathbf{x}_j$ ,  $a_{\Pi_V}$  is the autocorrelation of the function  $\Pi_V$ , and  $\Pi_V$  is 1 inside  $V$  and zero otherwise. With  $w = 1/n$ , the resulting estimator is unbiased as is always the case for stationary processes (Section 4.2). Substituting the expression for the pair correlation function in Equation 10 into Equation 7, the variance can be derived as in Table I, Equation T3. It is also possible to express the resulting formula in terms of  $s_i$ , as shown in Table I, Equation T4 (Appendix A.3). This form is particularly important to analyze and compute variance of uniform jittering in terms of strata-based local properties of the integrand as we will investigate in Section 6. So far, such an analysis has only been possible for 1D

Table I. Variance formulae for the stratified patterns considered. Here,  $s_i(\mathbf{x}) = f(\mathbf{x} + \mathbf{o}_i)\Pi_S(\mathbf{x})$ , where  $\Pi_S$  ( $\Pi_V$ ) is 1 inside  $S$  ( $V$ ) and 0 otherwise,  $S$  is the hypercube of length  $1/n^{1/d}$  centered around the origin,  $\mathbf{o}_i$  is the center of the stratum  $S_i$ ,  $s_i^-(\mathbf{x}) := s_i(-\mathbf{x})$  is the local patch of  $f$  mirrored around  $\mathbf{o}_i$ ,  $\mathbf{h}_{ij} = \mathbf{x}_i - \mathbf{x}_j$ ,  $a_f$  is the autocorrelation function of  $f$ ,  $F = \mathcal{F}(f)$  is the Fourier transform of  $f$ ,  $\omega_k$  is the set of points of the infinite regular grid of spacing  $n^{1/d}$  centered at the origin in the frequency domain,  $|\partial V_d|$  is the area of the unit hypersphere in  $d$  dimensions. For mirrored jitter sampling, there are  $m$  strata and  $n = 2m$  points.

Random jitter sampling

$$\frac{1}{n^2} \sum_{i=1}^n n \int s_i^2(\mathbf{x}) d\mathbf{x} - \left( n \int s_i(\mathbf{x}) d\mathbf{x} \right)^2 \quad (T1)$$

Mirrored jitter sampling

$$\frac{1}{n^2} \sum_{i=1}^m n \int s_i(\mathbf{x}) s_i^-(\mathbf{x}) d\mathbf{x} + n \int s_i^2(\mathbf{x}) d\mathbf{x} - \left( n \int s_i(\mathbf{x}) d\mathbf{x} \right)^2 \quad (T2)$$

Uniform jitter sampling

$$\frac{1}{n^2} \sum_{i \neq j} \frac{a_f(\mathbf{h}_{ij})}{a_{\Pi_V}(\mathbf{h}_{ij})} + J(f, n) \quad (T3)$$

$$\frac{1}{n^2} \sum_{ij} n \int s_i(\mathbf{x}) s_j(\mathbf{x}) d\mathbf{x} - n^2 \int s_i(\mathbf{x}) d\mathbf{x} \int s_j(\mathbf{x}) d\mathbf{x} \quad (T4)$$

$$\sum_{\omega_k \neq 0} |F(\omega_k)|^2 \quad (T5)$$

Uniform + random jitter sampling

$$\frac{1}{n^2} \sum_{i \neq j} \frac{[f(\mathbf{x}) * (a_k * f)(-\mathbf{x})](\mathbf{h}_{ij})}{a_{\Pi_V}(\mathbf{h}_{ij})} + J(f, n) \quad (T6)$$

Isotropic jitter sampling

$$\frac{1}{\lambda^2 |\partial V_d|} \sum_{i \neq j} \frac{\int a_f(\mathbf{h}) \delta(\|\mathbf{h}\| - \|\mathbf{h}_{ij}\|) d\mathbf{h}}{a_{\Pi_V}(\mathbf{h}_{ij}) \|\mathbf{h}_{ij}\|^{d-1}} + J(f, n) \quad (T7)$$

$$J(f, n) = \frac{1}{n} \int f^2(\mathbf{x}) d\mathbf{x} - \left( \int f(\mathbf{x}) d\mathbf{x} \right)^2 \quad (T8)$$

binary integrands under certain assumptions [Ramamoorthi et al. 2012]. For brevity, we denote each term in this equation as  $c_{ij}$  such that  $\text{var}(\hat{I}) = \frac{1}{n^2} \sum_{i,j} c_{ij}$ . Note that  $c_{ii}$  are the terms appearing in the variance formula for random jitter sampling (Equation T1). Finally, utilizing the relation  $\mathcal{F}(a_f) = |\mathcal{F}(f)|^2$ , we can also derive (Appendix A.3) a spectral formula for variance (Table I, Equation T5). Due to the translation invariance of the process, the variance does not depend on the phase of the Fourier transform of the integrand (please see the supplementary material for an analysis of variance in the spectral domain).

We present an example to illustrate good and bad cases for uniform jittering in Figure 5. In the example, two samples are used for integrating each of the two integrands, integrand 1 and integrand

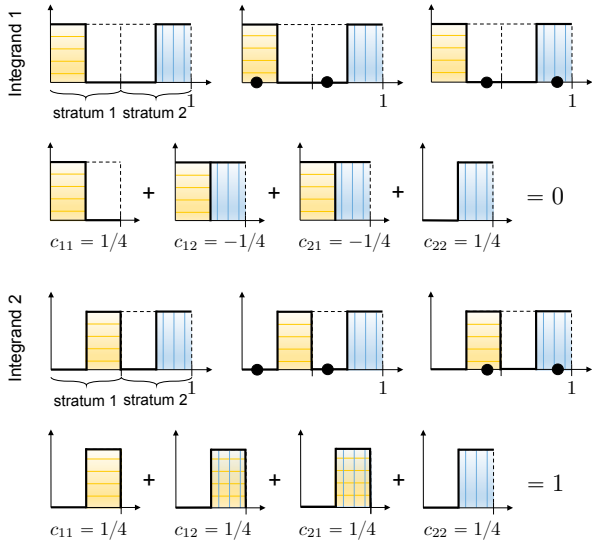


Fig. 5. We can compute exact variances for the integrands 1 and 2. For each case, the integrand itself is in the top-left figures. Top-middle and top-right figures illustrate two realizations of uniform jittering. Exact variance computations (up to a constant factor of  $1/2^2$ ) are shown on the bottom rows.

2. For each integrand, we can directly compute variance by calculating the terms  $c_{ij}$  and summing them, giving a zero variance for the first integrand, and  $1/2^2$  for the second. This result can be easily confirmed by considering some realizations of uniform jittering, as is done in the figure. For integrand 1, it is always true that  $\hat{I} = 0.5 = I$ , while for integrand 2, it can sometimes be 0 and sometimes 1. In summary, the negatively correlated  $s_1$  and  $s_2$  for integrand 1 makes the variance vanish, while the opposite is true for integrand 2.

**Uniform followed by Random Jittering** We can think of combining properties of uniform and random jittering by first applying a uniform jittering to generate a jittered grid, and then random jittering to each point. When  $k_i$ 's for random jittering are shifted versions of a function  $k$  (as in standard random or Gaussian jittering [Subr and Kautz 2013]), we can derive (Appendix A.4) that this only modifies the first term in Equation T3 such that the autocorrelation  $a_f(\mathbf{h}) = [f(\mathbf{x}) * f(-\mathbf{x})](\mathbf{h})$  turns into  $[f(\mathbf{x}) * (a_k * f)(-\mathbf{x})](\mathbf{h})$ , where  $a_k$  is the autocorrelation function of  $k$ . Hence, the correlations will be between  $f$  and a smoothed version of it (Table I, Equation T6). Although this reduces correlations to some degree, we did not find this reduction effective in practice.

**Isotropic Jitter Sampling** The main disadvantage of uniform jitter sampling results from positively correlated strata, leading to high  $c_{ij}$ . Even when it performs better than random jitter sampling, these correlations can increase the variance. A case where uniform jittering performs better than random jittering in spite of the correlations present among strata is shown in Figure 6, top-left. A look at the autocorrelation function (Figure 6, top-right) reveals that large terms along the vertical axis will contribute to variance in Equation T3. Pictorially, variance for uniform jittering consists of sums of  $a_f$  at given locations (Figure 6, bottom-left), some of which take the large values along the vertical axis. Instead, the idea of isotropic jitter sampling is having a variance equation that involves circular averages of the autocorrelation function (Figure 6, bottom-middle).

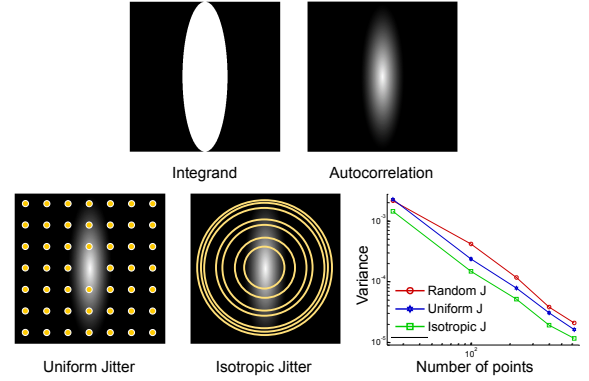


Fig. 6. Although uniform jittering performs better than random jittering in this example, the anisotropy of the autocorrelation function limits the reduction in variance due to the sum of the strong terms along the vertical axis. Instead, the variance due to isotropic jitter sampling consists of terms that are radial averages of the autocorrelation function, which are lower than the values on the vertical axis.

Such a sampling pattern can be obtained by introducing a global random translation (as in uniform jitter sampling), *and* independent rotation. Each instance of this process is thus a randomly translated and rotated regular grid. Rotated grids have so far only been considered to supersample pixels for antialiasing [Beets and Barron 2000; Hasselgren et al. 2005]. The generating point process for a randomly rotated and translated regular grid is isotropic, i.e. translation and rotation invariant, and hence we get an unbiased estimator for constant weights  $w = 1/\lambda$ . The variance can be derived (Appendix A.5) as shown in Table I, Equation T7. This expression shows that the only change to the variance of uniform jittering is that we now take radial averages of  $a_f$  at radii  $\|\mathbf{h}_{ij}\|$ , instead of values of  $a_f$  at  $\mathbf{h}_{ij} = \mathbf{x}_i - \mathbf{x}_j$ .

By a random rotation, we are avoiding problematic cases for uniform jittering, i.e. when there are strong positive correlations among the strata, while still having the maximum possible distance between the sampling points, in contrast with random jitter sampling, which has been considered as the alternative for those cases [Ramamoorthi et al. 2012]. Note, however, that random rotations can also increase the variance of uniform jittering. As the pattern of uniform jittering is rotated, we get a minimum and a maximum variance depending on the rotation. Since the term in Equation T7 is an average of variances of uniform jittering for all rotations, it lies between this minimum and maximum.

**R-Uniform Jitter Sampling** We call the rotated uniform jittering pattern with the minimum variance as *rotated-uniform* jittering. Of course, knowing this exact rotation requires information about the integrand. Hence, it is not an unadaptive method anymore. Note that here not the density, but the *correlations* are adapted to the integrand. In the scope of this paper, we estimate this rotation by trying different rotations. Hence, the errors for r-uniform jittering are reported to inform the reader on the lowest possible error that can be obtained by rotating the uniform jittering pattern. As an example, for the scene in Figure 3, for the same error level, on average 40%, 75%, and 101% less samples than uniform jittering are required for the square, circle, and Gaussian lights, respectively. If less information is available, it is also possible to limit the rotations to a certain range. Since the resulting process is still stationary, we get an unbiased estimator with less variance. We will present an analysis of how rotation can affect the variance for visibility sampling in Section 6. Please see Appendix A.6 for an al-



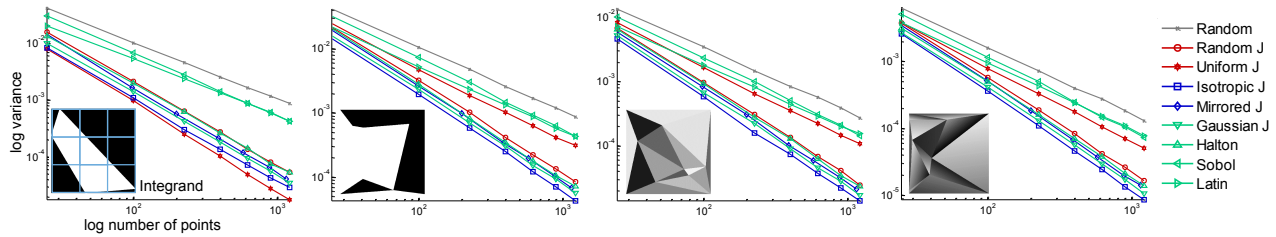


Fig. 7. An instance of the 50 random integrands used for each case is shown in the insets, with nine strata laid over the leftmost one for reference. Except the first integrand with strong negative correlations among the edges, isotropic jitter sampling performs the best, followed by Gaussian jittering (with  $\sigma = 0.25/\sqrt{n}$ ) and mirrored jittering sampling.

gorithm to generate rotated and translated grids with stratum-wise operations.

**Summary and Discussion** Our framework allows us to derive, for the first time, variance formulae for jittered sampling with global inter-strata correlations (Equations T3-T7). Thanks to operating in the spatial domain, the formulae are simple and amenable to analysis, inference, and comparisons, leading to novel insights and design choices, as we will illustrate in the next sections.

The variance formulae shows that the choice of the sampling pattern depends on the correlations present in the integrand. Local intra-strata correlations can be exploited by utilizing mirrored jitter sampling. It can be observed in Equation T2 that if a local patch in a stratum and its mirrored version are negatively correlated, variance will be reduced. Greater reductions can be obtained when there are global correlations. As revealed by Equation T4, negatively correlated strata leads to considerable reduction in variance when using uniform jitter sampling. For functions with positively correlated strata (leading to high  $a_f(\mathbf{h}_{ij})$  terms in Equation T3 as in Figure 6), it is beneficial to use isotropic jitter sampling instead. The random rotations in this sampling pattern can be constrained to further reduce the variance (r-uniform jittering), when the anisotropy in the integrand can be estimated. Hence, we can choose the pattern to use based on our prior knowledge of the possible correlations present in the integrands considered.

## 5.2 Empirical Analysis

The analysis so far provides the theoretical conditions for correlated jittering patterns to produce low-variance estimates for given classes of integrands. In this section, we investigate these conditions in practice via computational experiments.

As a benchmark, we chose to use the setting and integrands used in the previous work by Subr and Kautz [2013] since they present challenging cases with discontinuities for integral estimation, and resemble the integrands encountered in practice. We have four different types of integrands with random points determining the regions [Subr and Kautz 2013]. One instance of each is shown in Figure 7 insets. We use 50 random instances of each integrand type and 100 different point distributions for each of the tested sampling algorithms to compute the variance of the estimator. To allow averaging the variances for different instances of an integrand type, we normalized each estimated variance by the squared mean of the integral estimates [Subr and Kautz 2013].

As shown in Figure 7, left, the first integrand contains strongly negatively correlated strata pairs, such as the pair of the middle stratum and bottom-left stratum in the figure. Hence, uniform jitter sampling works best for this case due to the negative  $c_{ij}$ , followed by isotropic jittering. However, for all other integrands, uniform

jittering performs quite poorly, due to the linear discontinuities between the regions, while isotropic jittering always gives the lowest variance. Note that this variance is even lower than those of the biased methods tested. For all four integrand types, the variance of isotropic jitter sampling is followed by those of Gaussian jittering (with  $\sigma = 0.25/\sqrt{n}$  for  $n$  points) [Subr and Kautz 2013] and mirrored jitter sampling. Note that for this  $\sigma$ , Gaussian jittering has apparent bias (see Figure 3), while mirrored jittering is unbiased. Mirrored jittering performs well due to the linear boundaries of the regions, which result in negative correlation between a stratum and its mirrored version. This can be observed, for example, for the bottom-left stratum shown for the first integrand.

## 6. APPLICATION TO VISIBILITY SAMPLING

The analysis in the previous sections shows how the error in integration depends on the sampling patterns as well as the integrands. We particularly focused on the difficult case of integrands with discontinuities as they arise in practice for rendering due to visibility, and jittering patterns as they are commonly used for distributed ray tracing. Indeed, the main inspiration of this work is previous results on visibility sampling for soft shadows [Ramamoorthi et al. 2012] and integration of functions with discontinuities with jittered patterns in general [Subr and Kautz 2013]. In this section, we present an application of our analysis to visibility sampling for computing soft shadows with ray tracing, by considering the joint pixel-light space. We show that careful selection of sampling strategies can result in soft shadows with significantly lower variances than previously possible [Ramamoorthi et al. 2012]. We study sampling patterns in relation with light types and scene geometry, to understand when, why, and how much certain patterns can perform better. Here we focus on some of the important conclusions. For theoretical results with detailed comparisons, and derivation details please refer to the supplementary material.

### 6.1 Soft Shadows from Planar Area Lights

We assume that there are planar area light sources in the scene to be rendered. These emit light from each point  $\mathbf{y}$  on the light plane. Under common assumptions [Ramamoorthi et al. 2012; Mehta et al. 2012], the image intensity at a point  $\mathbf{x}$  on the image plane is computed by the multiplication of the function  $t(\mathbf{x})$  that models light transport including BRDFs, and the following integral:

$$I(\mathbf{x}) = \int_V v(\mathbf{x}, \mathbf{y})l(\mathbf{y})d\mathbf{y}, \quad (11)$$

where  $I$  is the intensity (radiance),  $V$  is the unit square as before,  $v$  is the binary visibility function, and  $l$  is the lighting from the area

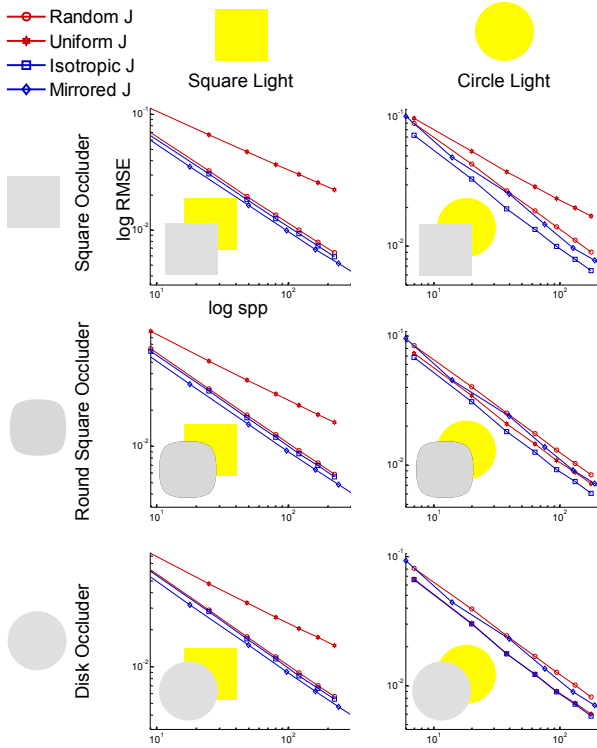


Fig. 8. RMSE plots for soft shadows with different combinations of basic occluders and light types.

light source. The binary function  $v(\mathbf{x}, \mathbf{y})$  gives which points on the light plane are occluded and hence is determined by the scene geometry. Common light types considered for rendering are square lights:  $l(\mathbf{y}) = 1$  in  $V$ , circle lights:  $l(\mathbf{y}) = 1$  for  $\mathbf{y}$  in a disk of a certain radius, and Gaussian lights:  $l$  is a Gaussian function.

## 6.2 Error Analysis and Comparisons

We need to consider the lighting  $l$  along with the visibility function  $v$  to analyze the error in Equation 11 for different sampling strategies. For each pixel location  $\mathbf{x}$ ,  $v(\mathbf{x}, \mathbf{y})$  changes, and hence there is a different integrand on the light plane. Thus, we need to take into account that the overall error in soft shadows in a scene is the sum of errors at all pixels. We assume square or circular light sources such that the integrand  $v(\mathbf{x}, \mathbf{y})l(\mathbf{y})$  becomes a binary function (Gaussian lights result in error behaving very similarly to that of circular lights, as we illustrate in the supplementary material). By defining the estimator at pixel  $\mathbf{x}$  as  $\hat{I}(\mathbf{x})$ , we thus want to compare sampling patterns in terms of the following total error:  $e = \int_R \text{var}(\hat{I}(\mathbf{x}))d\mathbf{x}$  (we consider unbiased jittering patterns and hence mean squared error is equal to variance) over some area  $R$  of the pixel space. We compare different sampling strategies analytically and empirically to understand under what conditions on the light types and scene geometries they will perform well.

**Mirrored vs. Random Jitter Sampling** Under common [Egan et al. 2011; Ramamoorthi et al. 2012; Mehta et al. 2012] and further mild assumptions as described in the supplementary material, and a square light source ( $l(\mathbf{y}) = 1 \forall \mathbf{y} \in V$ ), the ratio of total errors of random jitter  $e^{RND}$  and mirrored jitter  $e^{MIR}$  sampling can be derived as  $e^{RND}/e^{MIR} = 2^{2-p}$ , where  $p$  is the convergence rate of random jittering. It has been observed that  $p \in [1, 2]$  [Mitchell

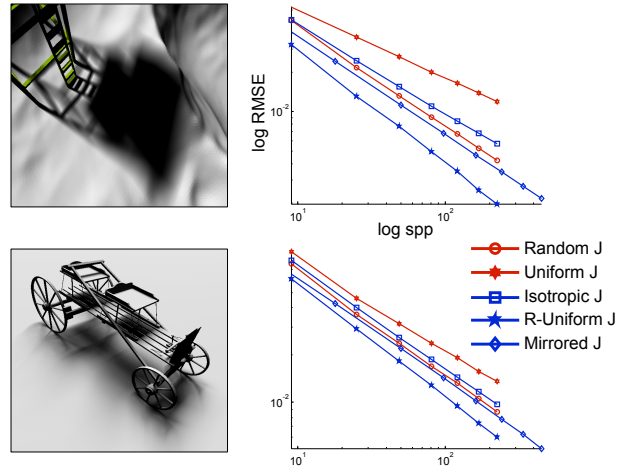


Fig. 9. RMS errors are plotted for two different scenes with a square light source. The correlations among the strata make uniform and isotropic sampling disadvantaged. Among the non-adaptive sampling methods (i.e. excluding R-Uniform J), mirrored jitter sampling results in the least error. Its advantage diminishes as the shadows get more complex (bottom).

1996; Ramamoorthi et al. 2012], and stay around 1.5 for integrands with discontinuities. Hence, mirrored jittering will always perform better than random jittering under the mentioned assumptions. For the typical convergence rate of  $p = 1.5$ , random jitter sampling will have  $\sqrt{2}$  times as much variance as mirrored jitter sampling. For the same error level, random jittering needs to use  $2^{2/p-1}$  times as much samples as mirrored jittering. Random jittering thus requires approximately 25% more samples than mirrored jittering for  $p = 1.5$ , with the maximum 100% for  $p = 1$ . We illustrate several cases where these assumptions hold in Figure 8, square light. As expected, the ratio of the variances approaches the estimated  $\sqrt{2}$  for these cases. In practice, even if the assumptions are not exactly satisfied, the ratio approximately holds. We illustrate such cases in Figures 3 and 9. For the simpler occluding geometry in Figure 9, top, random jittering requires 22% more samples than mirrored jittering for the same error level, while this drops to 11% for the complex shadow in Figure 9, bottom. In summary, for square lights and simpler scene geometries, mirrored jitter sampling performs better than random jitter sampling, requiring around 25% less samples. However, for other light types, or more complex occluding geometries, its advantage disappears.

**Uniform Jitter Sampling** As elaborated in Section 5.1, and can be observed in Equation T4, uniform jitter sampling exploits negatively correlated strata to reduce variance. This is captured by the terms  $c_{ij}$ . These terms will thus favor uncorrelated random jittering for positively correlated strata pairs. In particular, if we assume a square light source, and that for a given pixel location  $\mathbf{x}$  the visibility function has a linear boundary in the whole domain  $V$  of the light plane, we can analytically prove that random jittering will perform better (supplementary material). In general, the slopes of the boundary curves in different strata can be different. Local boundary lines with similar slopes increase  $c_{ij}$  and hence the error of uniform jittering, while the opposite is true for disparate slopes. This is illustrated in Figure 8 for simple flat occluder shapes. For both light types, uniform jittering performs worst for the square, and best for the disk occluder. Since the circle light breaks long boundary lines, uniform jittering performs better for the circle light, as previously observed [Ramamoorthi et al. 2012].

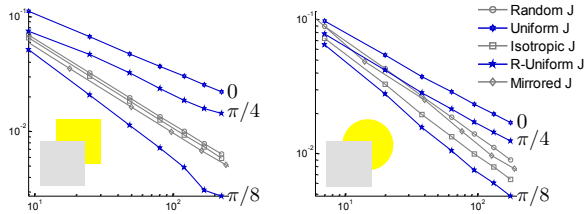


Fig. 10. R-uniform jittering is obtained by rotating the grid pattern of uniform jittering by different angles. R-uniform jittering can give lower errors than possible with other non-adaptive methods, but requires estimating the rotation angle.

Note, however, that the error of uniform jittering is still significantly higher than those of other sampling patterns for the circle light - square occluder pair, deviating from previous observations [Ramamoorthi et al. 2012]. More interestingly, by simply rotating the regular grid pattern that is uniformly jittered by a constant angle, uniform jittering can perform significantly better than the other non-adaptive sampling patterns. This is due to the strong correlations introduced by the changing slopes in the strata. We illustrate this in Figure 10, where different rotation angles are applied, and also in Figures 3 and 9 with the optimal angle under the name r-uniform jittering. In summary, even for square light sources, uniform jittering can be by far the best alternative if a global angle of rotation can be estimated. (Please see the supplementary images for more examples with uniform and r-uniform jitter sampling, in Figures 8 and 10 uniform jittering refers to the one with rotation for maximum variance, in other figures the rotation is random.)

**Isotropic Jitter Sampling** Isotropic jitter sampling tries to reduce the variance of uniform jitter sampling by introducing a random rotation. The error difference will be largest when the radial average of  $a_f(\mathbf{h})$  at  $\mathbf{h}_{ij}$  deviates from  $a_f(\mathbf{h}_{ij})$  most, which happens when there are directional structures. These are especially prominent for square lights and simple geometries (Figure 8, left and Figure 9, top) but also for circle lights and long lines of low curvature (Figure 8, circle light - square occluder). For these cases, uniform jitter sampling has higher error, while isotropic sampling performs significantly better. For more symmetric occluders and symmetric lights, the difference in error diminishes. In the special case of a flat receiver parallel to the light plane, we can analytically compute the total error (supplementary material). In particular, for a disk occluder and circle light, we get the same error for both uniform and isotropic jittering. Such a case is illustrated in Figure 8, right, third row (circle light - disk occluder pair).

By constraining the random rotation of the uniformly jittered pattern (r-uniform jittering), errors lower than isotropic jittering can be obtained (Figure 10). Unlike all other jittering methods considered, estimating ranges of right angles, however, requires knowledge on the light types and scene geometry. Isotropic jittering is thus an unadaptive alternative to capture such angles by random sampling. In contrast to random jitter sampling, distributions generated by isotropic jittering have well distributed points, which reduces variance. Hence, it is a better unadaptive alternative.

## 7. CONCLUSIONS AND FUTURE DIRECTIONS

We presented a new approach to study error in integration with stochastic point patterns. The theoretical framework gives exact bias and variance expressions for general sampling patterns, as long as they can be interpreted as stochastic. The expressions depend on

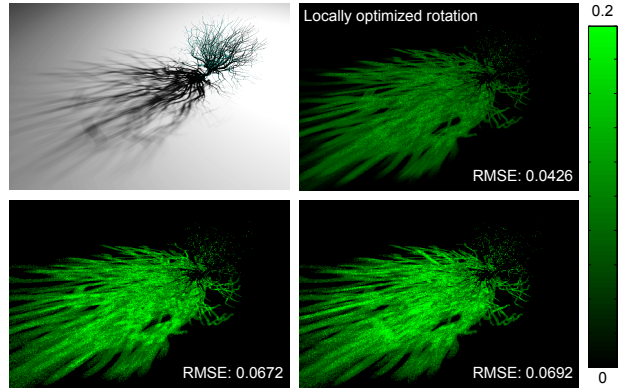


Fig. 11. For a scene with thin structures and complex shadows (top-left), the minimum and maximum RMS errors (bottom-left and -right) obtained by rotating the uniform jittering pattern do not differ much. In contrast, adapting the rotation to each pixel significantly reduces the error (top-right).

spatial or spectral properties of the integrands, and first and second order correlations of the sampling patterns. We illustrated the utility of the theory by applying it to the analysis of existing sampling patterns, and in particular to the commonly used jittered patterns. The analysis allowed us to propose to use and study alternative correlated patterns optimized for visibility sampling. We believe similar analyses based on the proposed framework will be useful in reducing error and improving properties of sampling algorithms for rendering and further applications. We mention a few of these applications and future directions below.

**Adaptive density** For visibility sampling, the analyzed jittered patterns can be immediately used with adaptive methods that determine number of visibility samples per pixel and sample the light plane with stratified distributions [Mehta et al. 2012]. In general, our focus on optimizing correlations among points is orthogonal to the adaptive sampling methods which optimize density. Correlations and density are captured by different terms in our bias and variance expressions, which makes performing similar analyses for adaptive sampling patterns possible. Adaptive sampling algorithms used for integration in graphics estimate measures such as the variation of the integrand, and distribute more samples in those regions, often iteratively [Whitted 1980; Hachisuka et al. 2008; Overbeck et al. 2009; Belcour et al. 2013]. If the underlying distribution can be assumed as stationary, the variance of the resulting point patterns can be accurately described by second-order intensity-reweighted stationary point processes [Illian et al. 2008]. For these processes, the second order product density is given by  $\varrho(\mathbf{x}, \mathbf{y}) = \lambda(\mathbf{x})\lambda(\mathbf{y})g(\mathbf{x} - \mathbf{y})$ , clearly separating the intensity and PCF terms. We elaborate on how bias and variance expressions for adaptive distributions with given statistics can be derived and analyzed in the supplementary material.

**Adaptive correlations** Although density adaptation is extensively studied, adjusting correlations has been limited to distributing sampling points as uniformly as possible while avoiding aliasing artifacts, with the exception of antithetic sampling and a few recent works investigating adapting correlations to integrands [Subr and Kautz 2013; Subr et al. 2014]. Adapting correlations has been hindered by the difficulty of computing error via classical methods when the sample points are not independent. The proposed theoretical framework will be instrumental in overcoming this difficulty and developing patterns with correlations adapted to inte-

grands with certain properties. As an example, we illustrated how simply rotating a uniformly jittered grid according to an integrand can result in significant reductions in variance. This rotation can be estimated by shooting initial rays, similar to adaptive density estimation (e.g. [Mehta et al. 2012]).

**Blue noise and general sampling** Our bias and variance equations for general point processes apply to blue noise distributions, as well as to more general distributions such as combinations of deterministic patterns with randomization [Chiu et al. 1994; Kollig and Keller 2002; Kensler 2013]. Various methods can be used to estimate the point process statistics [Illian et al. 2008; Öztireli and Gross 2012; Heck et al. 2013], which can then be used to derive the expressions for bias and variance. Similar to our analysis for visibility sampling and stratified patterns, point patterns tailored to other applications can be developed.

**Local scene analysis** In this paper, we consider a globally optimal sampling pattern for the whole scene. However, local or even pixel-wise statistics can be used to determine different sampling patterns for different parts of the scene, as is currently done for importance sampling in rendering. An example of such an adaptation for soft shadows is shown in Figure 11. For the complex thin shadows in this figure, a global rotation of uniform jittering does not affect the error much, while using a pixel-wise optimum rotation on the light plane significantly reduces it. Our stratum-wise formulas allow such a local analysis in the joint light-pixel space.

**Antialiasing and reconstruction of images** In addition to integration, the proposed framework can be utilized to study antialiasing properties of sampling patterns, and the error introduced in reconstructions. Although the general desired characteristics of sampling patterns for antialiasing are known, it is an open problem to develop patterns with better correlations, and to combine the patterns with adaptive densities [Heck et al. 2013]. Traditionally, aliasing is studied in the spectral domain, but Heck et al. [2013] recently showed the utility of point process statistics for this problem. We can further understand and develop new point patterns with the proposed framework.

**Analysis in general domains** Although we presented the theory for Euclidean domains, it extends to non-Euclidean domains and general measure spaces as well [Illian et al. 2008]. Analysis of point patterns in general spaces, especially spherical, hemispherical, or warped domains is important for rendering applications. Such an analysis is challenging when starting from the spectral domain [Pilleboue et al. 2015]. By utilizing analytic or estimated measures or distances, the proposed theory can be lifted to such cases.

## APPENDIX

### A. PCF, BIAS, VARIANCE, AND GENERATION OF JITTERED POINT PATTERNS

#### A.1 Bias and Variance of Mirrored Jitter Sampling

To derive bias and variance of mirrored jitter sampling, we consider the conceptual setting of putting the original and mirrored versions of the integrand  $f$  side-by-side, as illustrated in Figure 12. We assume there are  $n$  points in total, which implies that there are  $m := n/2$  points for each of the original and the mirrored version of  $f$ . We define the functions  $f_i(\mathbf{x}) := s_i(\mathbf{x} - \mathbf{o}_i) + s_i^-(\mathbf{x} - \mathbf{o}_i^-)$ , where  $\mathbf{o}_i^-$  is the center of the stratum  $S_i^-$ ,  $s_i(\mathbf{x}) = f(\mathbf{x} + \mathbf{o}_i)\mathbb{I}_S(\mathbf{x})$ , and  $s_i^-(\mathbf{x}) = s_i(-\mathbf{x})$ . Each  $f_i$  is thus the sum of the part of the function  $f$  in  $S_i$  and its mirrored ver-

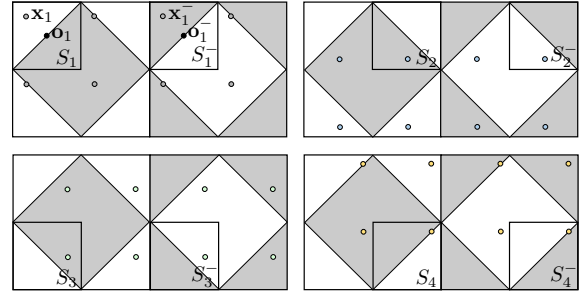


Fig. 12. In mirrored jitter sampling, conceptually, the integral  $I_i$  of the sum of the integrand in stratum  $S_i$  and its mirrored version in  $S_i^-$  is estimated by uniform jitter sampling. For different  $i$ , these uniformly jittered patterns are independent.

sion in  $S_i^-$ , as depicted in Figure 12. The integral of this function is given by  $I_i = \int f_i(\mathbf{x})d\mathbf{x} = \int [s_i(\mathbf{x} - \mathbf{o}_i) + s_i^-(\mathbf{x} - \mathbf{o}_i^-)]d\mathbf{x} = 2 \int_{S_i} s_i(\mathbf{x} - \mathbf{o}_i)d\mathbf{x} = 2 \int_{S_i} f(\mathbf{x})d\mathbf{x}$ .

Hence, the integral we seek is

$$I = \int f(\mathbf{x})d\mathbf{x} = \sum_{i=1}^m \int_{S_i} f(\mathbf{x})d\mathbf{x} = \frac{1}{2} \sum_{i=1}^m I_i, \quad (12)$$

and an unbiased estimator of this integral is:  $\hat{I} = \sum_{i=1}^m \hat{I}_i/2$  for unbiased  $\hat{I}_i$ . We will see that this estimator is indeed the estimator used in mirrored jitter sampling. Below, we will first derive the expected value and variance of  $\hat{I}_i$ , and then use them to compute the bias and variance of the estimator  $\hat{I}$ .

In mirrored jitter sampling, we assume that the estimator  $\hat{I}_i$  for each  $I_i$  is computed by uniform jitter sampling (we will see why this assumption leads to mirrored jitter sampling below). Note that the intensity of this pattern is  $\lambda = m = n/2$ . The expected value of  $\hat{I}_i$ , as given by Equation 4, is thus  $\mathbb{E}\hat{I}_i = \lambda \int w(\mathbf{x})f_i(\mathbf{x})d\mathbf{x}$ , and an unbiased estimator is obtained by setting  $w = 1/\lambda = 2/n$ :

$$\mathbb{E}\hat{I}_i = \int f_i(\mathbf{x})d\mathbf{x} = 2 \int_{S_i} f(\mathbf{x})d\mathbf{x} = I_i. \quad (13)$$

Hence, we get an unbiased estimator for  $\hat{I}$ :

$$\text{bias}(\hat{I}) = I - \mathbb{E} \frac{1}{2} \sum_{i=1}^m \hat{I}_i = I - \frac{1}{2} \sum_{i=1}^m 2 \int_{S_i} f(\mathbf{x})d\mathbf{x} = 0. \quad (14)$$

To compute the variance of  $\hat{I}_i$ , we can use Equation T4 that gives the variance of uniform jittering. Since there are only two strata,  $S_i$  and  $S_i^-$  where the function  $f_i(\mathbf{x})$  is non-zero, the variance can simply be written as:

$$\begin{aligned} \text{var}(\hat{I}_i) &= 2 \left( \frac{1}{\lambda} \int s_i(\mathbf{x})s_i^-(\mathbf{x})d\mathbf{x} - \int s_i(\mathbf{x})d\mathbf{x} \int s_i^-(\mathbf{x})d\mathbf{x} \right) \\ &\quad + 2 \left( \frac{1}{\lambda} \int s_i^2(\mathbf{x})d\mathbf{x} - \left( \int s_i(\mathbf{x})d\mathbf{x} \right)^2 \right) \\ &= 4 \left( \frac{1}{n} \int s_i(\mathbf{x})s_i^-(\mathbf{x})d\mathbf{x} + \frac{1}{n} \int s_i^2(\mathbf{x})d\mathbf{x} - \left( \int s_i(\mathbf{x})d\mathbf{x} \right)^2 \right), \end{aligned} \quad (15)$$

where we used the fact that  $\int s_i(\mathbf{x})d\mathbf{x} = \int s_i^-(\mathbf{x})d\mathbf{x}$ . Since the point processes to compute the estimators  $\hat{I}_i$  are independent, we

can write the variance of  $\hat{I}$  as  $var(\hat{I}) = var(\frac{1}{2} \sum_{i=1}^m \hat{I}_i) = \frac{1}{4} \sum_{i=1}^m var(\hat{I}_i)$ . Plugging in the formula for  $var(\hat{I}_i)$ , we obtain:

$$var(\hat{I}) = \frac{1}{n^2} \sum_{i=1}^m n \int s_i(\mathbf{x}) s_i^-(\mathbf{x}) d\mathbf{x} + n \int s_i^2(\mathbf{x}) d\mathbf{x} - \left( n \int s_i(\mathbf{x}) d\mathbf{x} \right)^2. \quad (16)$$

Finally, we show that the estimator  $\hat{I}$  corresponds to the estimator we propose in Section 5.1 (depicted in Figure 4, right), which is obtained by selecting a random point and its mirrored version around the origin of the stratum in each stratum. Recall that the estimator we use so far in this section has the form  $\hat{I} = \sum_{i=1}^m \hat{I}_i/2$ . Each estimator  $\hat{I}_i$  is obtained by uniform jitter sampling for  $f_i$  and hence an instance of this estimator is computed by  $\hat{I}_i = w(\mathbf{x}_i) f_i(\mathbf{x}_i) + w(\mathbf{x}_i^-) f_i(\mathbf{x}_i^-)$  for a sample point  $\mathbf{x}_i$  in  $S_i$  and another one  $\mathbf{x}_i^-$  in  $S_i^-$ , as illustrated in Figure 12. Substituting the weights  $w = 1/\lambda = 2/n$ , and noting that  $\mathbf{x}_i^- = \mathbf{x}_i + \mathbf{o}_i^- - \mathbf{o}_i$  and  $\Pi_S$  is an even function, we get the following expression for  $\hat{I}$ :

$$\begin{aligned} \hat{I} &= \frac{1}{n} \sum_{i=1}^m f_i(\mathbf{x}_i) + f_i(\mathbf{x}_i^-) \\ &= \frac{1}{n} \sum_{i=1}^m s_i(\mathbf{x}_i - \mathbf{o}_i) + s_i^-(\mathbf{x}_i^- - \mathbf{o}_i^-) \\ &= \frac{1}{n} \sum_{i=1}^m f(\mathbf{x}_i) \Pi_S(\mathbf{x}_i - \mathbf{o}_i) + f(\mathbf{o}_i + \mathbf{o}_i^- - \mathbf{x}_i^-) \Pi_S(\mathbf{o}_i^- - \mathbf{x}_i^-) \\ &= \frac{1}{n} \sum_{i=1}^m [f(\mathbf{x}_i) + f(2\mathbf{o}_i - \mathbf{x}_i)] \Pi_S(\mathbf{x}_i - \mathbf{o}_i) \\ &= \frac{1}{n} \sum_{i=1}^m f(\mathbf{x}_i) + f(2\mathbf{o}_i - \mathbf{x}_i), \end{aligned} \quad (17)$$

which is equivalent to taking a point  $\mathbf{x}_i$  and its mirrored point  $-\mathbf{x}_i + 2\mathbf{o}_i$  (around  $\mathbf{o}_i$ ), with weights  $1/n$  in each stratum  $S_i$ .

## A.2 PCF of Uniform Jitter Sampling Process

Since we assume a hypercube as the domain  $V$ , there will always be the same number of points in this cube for different realizations of uniform jitter sampling. This implies that  $\lambda = n$ , and we will always have the same difference vectors  $\mathbf{h}_{ij} := \mathbf{x}_i - \mathbf{x}_j = \mathbf{o}_i - \mathbf{o}_j$ , where  $\mathbf{x}_i$  is a point in a realization of the point process and  $\mathbf{o}_i$  is the origin of the stratum  $S_i$ . Hence the expected value of the following sum will be simply given by its own value:

$$\mathbb{E} \sum_{i=1}^n \sum_{j=1, j \neq i}^n \delta((\mathbf{x} - \mathbf{y}) - \mathbf{h}_{ij}) = \sum_{i=1}^n \sum_{j=1, j \neq i}^n \delta((\mathbf{x} - \mathbf{y}) - \mathbf{h}_{ij}), \quad (18)$$

where  $\delta$  is the Dirac delta.<sup>1</sup> On the other hand, this expected value can also be computed by using Campbell's theorem as follows:

$$\begin{aligned} &\mathbb{E} \sum_{i=1}^n \sum_{j=1, j \neq i}^n \delta((\mathbf{x} - \mathbf{y}) - \mathbf{h}_{ij}) \\ &= n^2 \int_V \delta((\mathbf{x} - \mathbf{y}) - (\mathbf{z} - \mathbf{t})) g(\mathbf{z} - \mathbf{t}) d\mathbf{z} d\mathbf{t} \end{aligned}$$

<sup>1</sup>Technically, we need a series of functions that converge to Dirac delta here.

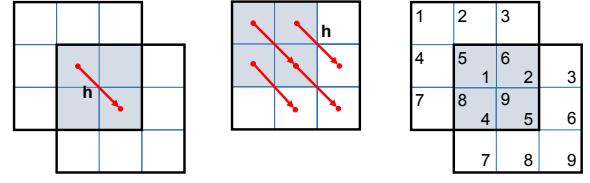


Fig. 13. (Left) The shaded area shows the autocorrelation  $a_{\Pi_V}(\mathbf{h})$  of  $\Pi_V$  (unit square) at  $\mathbf{h}$ . In this 2 dimensional case, it is equal to  $(1 - |h_x|)(1 - |h_y|) = 4/9$ . (Middle) There are  $(1 - |h_x|)(1 - |h_y|)n = 4$   $\mathbf{h}$  vectors between origins of different strata in Equation 20. (Right) Each different  $\mathbf{h}$  results in correlations between different  $s_i$  and  $s_j$ .

$$= n^2 g(\mathbf{x} - \mathbf{y}) a_{\Pi_V}(\mathbf{x} - \mathbf{y}), \quad (19)$$

where  $a_{\Pi_V}(\mathbf{x} - \mathbf{y}) = \prod_{i=1}^d (1 - |(\mathbf{x} - \mathbf{y})_i|)$  is the autocorrelation function of  $\Pi_V$  at  $\mathbf{x} - \mathbf{y}$  (Figure 13, left) with  $(\mathbf{x} - \mathbf{y})_i$  denoting the  $i^{\text{th}}$  component of the  $d$  dimensional vector  $\mathbf{x} - \mathbf{y}$ . Combining Equations 18 and 19 and substituting  $\mathbf{h} = \mathbf{x} - \mathbf{y}$ , the pair correlation function  $g$  of uniform jitter sampling process can be written as:

$$g(\mathbf{h}) = \frac{1}{n^2} \sum_{i \neq j} \frac{\delta(\mathbf{h} - \mathbf{h}_{ij})}{a_{\Pi_V}(\mathbf{h}_{ij})} \quad a_{\Pi_V}(\mathbf{h}) \neq 0. \quad (20)$$

We can also derive a more refined version of this equation by cancelling the effect of autocorrelation. Consider a unique  $\mathbf{h}$  among the  $\mathbf{h}_{ij}$ 's. The number of times that particular  $\mathbf{h}$  appears in the sum in Equation 20 is given by  $n \prod_{i=1}^d (1 - |(\mathbf{h})_i|) = n a_{\Pi_V}(\mathbf{h})$ , where  $(\mathbf{h})_i$  denotes the  $i^{\text{th}}$  component of  $\mathbf{h}$ , as illustrated in Figure 13, left and middle. Hence, we can write Equation 20 as

$$\begin{aligned} g(\mathbf{h}) &= \frac{1}{n^2} \sum_{\text{unique } \mathbf{h}_k} n a_{\Pi_V}(\mathbf{h}_k) \frac{\delta(\mathbf{h} - \mathbf{h}_k)}{a_{\Pi_V}(\mathbf{h}_k)} \\ &= \frac{1}{n} \sum_{\text{unique } \mathbf{h}_k} \delta(\mathbf{h} - \mathbf{h}_k) \quad a_{\Pi_V}(\mathbf{h}) \neq 0. \end{aligned} \quad (21)$$

## A.3 Variance of Uniform Jitter Sampling

**Strata-based variance formula** In order to derive Equation T4, we need to express the integrals of  $f$  and the autocorrelation function  $a_f$  in terms of the integrals of  $s_i$ . The former is trivial, while the latter can be realized by utilizing the expression for the pair correlation function of uniform jittering process in Equation 21. Recall that the variance of a stationary point process is given by (Equation 7):

$$var(\hat{I}) = \frac{1}{\lambda} \int f^2(\mathbf{x}) d\mathbf{x} + \int a_f(\mathbf{h}) g(\mathbf{h}) d\mathbf{h} - \left( \int f(\mathbf{x}) d\mathbf{x} \right)^2. \quad (22)$$

Substituting the PCF (Equation 21) into the second term in this expression, we obtain  $\frac{1}{n} \sum_{\text{unique } \mathbf{h}_k} a_f(\mathbf{h}_k)$ . Hence, it is sufficient to express  $a_f(\mathbf{h}_k)$  in terms of the integrals of  $s_i(\mathbf{x})$ .

The function  $f$  can be written as:  $f(\mathbf{x}) = \sum_i s_i(\mathbf{x} - \mathbf{o}_i)$ . The autocorrelation  $a_f$  can then be expanded as:

$$\begin{aligned} a_f(\mathbf{h}) &= \int f(\mathbf{x}) \sum_i s_i(\mathbf{x} - \mathbf{h} - \mathbf{o}_i) d\mathbf{x} \\ &= \sum_i \int f(\mathbf{x} + \mathbf{h} + \mathbf{o}_i) s_i(\mathbf{x}) d\mathbf{x} \end{aligned}$$

$$\begin{aligned}
&= \sum_i \int \sum_j s_j(\mathbf{x} + \mathbf{h} + \mathbf{o}_i - \mathbf{o}_j) s_i(\mathbf{x}) d\mathbf{x} \\
&= \sum_{i,j} \int s_i(\mathbf{x}) s_j(\mathbf{x} + \mathbf{h} - \mathbf{h}_{ji}) d\mathbf{x}. \quad (23)
\end{aligned}$$

In this sum, for a given  $\mathbf{h} \in \{\text{unique } \mathbf{h}_k\}$ , only the terms with  $\mathbf{h}_{ji} = \mathbf{h}$  will be non-zero due to the limited supports of  $s_i$ . Hence, each  $\mathbf{h} \in \{\text{unique } \mathbf{h}_k\}$  results in several unique terms  $\int s_i(\mathbf{x}) s_j(\mathbf{x}) d\mathbf{x}$ , as also illustrated in Figure 13, right. Summing over all  $\mathbf{h}_k$  gives:

$$\begin{aligned}
&\frac{1}{n} \sum_{\text{unique } \mathbf{h}_k} a_f(\mathbf{h}_k) \\
&= \frac{1}{n} \sum_{i,j} \sum_{\text{unique } \mathbf{h}_k} \int s_i(\mathbf{x}) s_j(\mathbf{x} + \mathbf{h}_k - \mathbf{h}_{ji}) d\mathbf{x} \quad (24) \\
&= \frac{1}{n} \sum_{i \neq j} \int s_i(\mathbf{x}) s_j(\mathbf{x}) d\mathbf{x}.
\end{aligned}$$

Finally, adding this term to the first and last terms in Equation 22 gives us Equation T4. Note that the first term can be written as  $\frac{1}{n} \sum_i \int s_i^2(\mathbf{x}) d\mathbf{x}$  due to the disjoint supports of  $s_i$ .

**Spectral variance formula** To express the variance in the spectral domain, we need to write the second term in Equation 22 involving the autocorrelation function  $a_f$  in terms of the Fourier transform  $F$  of the integrand  $f$ . Starting from the form of this term in Equation 24, and noting that the set  $\{\text{unique } \mathbf{h}_k\}$  in Equation 24 can be extended to the set of points in an infinite regular grid centered at the origin  $\mathbf{0}$  except the origin itself (since  $a_f$  is zero for all  $\mathbf{h}_k \neq \mathbf{0}$  not in the set  $\{\text{unique } \mathbf{h}_k\}$  due to the finite support  $V$  of  $f$ ), we can write

$$\begin{aligned}
&\frac{1}{n} \sum_{\mathbf{h}_k \neq \mathbf{0}} a_f(\mathbf{h}_k) \\
&= \frac{1}{n} \sum_{\mathbf{h}_k} a_f(\mathbf{h}_k) - \frac{1}{n} a_f(\mathbf{0}) \\
&= \int a_f(\mathbf{h}) \frac{1}{n} \sum_{\mathbf{h}_k} \delta(\mathbf{h} - \mathbf{h}_k) d\mathbf{h} - \frac{1}{n} a_f(\mathbf{0}) \\
&= \mathcal{F} \left[ a_f(\mathbf{h}) \frac{1}{n} \sum_{\mathbf{h}_k} \delta(\mathbf{h} - \mathbf{h}_k) \right] (\mathbf{0}) - \frac{1}{n} a_f(\mathbf{0}) \\
&= \int |F(\omega)|^2 \sum_{\omega_k} \delta(\omega - \omega_k) d\omega - \frac{1}{n} a_f(\mathbf{0}) \\
&= \sum_{\omega_k} |F(\omega_k)|^2 - \frac{1}{n} \int f^2(\mathbf{x}) d\mathbf{x}. \quad (25)
\end{aligned}$$

Adding this term to the first and last terms in Equation 22 gives

$$\begin{aligned}
\text{var}(\hat{I}) &= \sum_{\omega_k} |F(\omega_k)|^2 - \left( \int f(\mathbf{x}) d\mathbf{x} \right)^2 \\
&= \sum_{\omega_k} |F(\omega_k)|^2 - |F(\mathbf{0})|^2. \quad (26)
\end{aligned}$$

#### A.4 Variance of Uniform + Random Jitter Sampling

Following uniform jittering by random jittering can be considered as performing a clustering process on the initial stationary point

process underlying uniform jittering. For each realization of this process, we first generate a jittered grid, and then add a further independent jittering to each point  $\mathbf{x}_i$  of this jittered grid following a distribution  $k_i = k(\mathbf{x} - \mathbf{x}_i)$ , where the supports of  $k_i$  are assumed to be disjoint (as in Section 5.1, random jitter sampling).

For such a point process, we can utilize a result that gives the PCF of a clustering process where the cluster centers are chosen at points from an underlying stationary process [Felsenstein 1975], [Illian et al. 2008] (p. 370). This result assumes that around each point of the underlying stationary process, called the parent process, with  $\lambda_p$ ,  $\varrho_p$ , and  $g_p$ , a cluster is formed without including the parent point. Defining  $\lambda_c$ ,  $\varrho_c$ , and  $g_c$  for the cluster process, and denoting the mean number of points in a cluster by  $\bar{c}$ , the intensity and second order product density of the resulting process can be given by  $\lambda = \lambda_p \bar{c}$ , and:

$$\begin{aligned}
\varrho(\mathbf{h}) &= \lambda_p \int \varrho_c(\mathbf{h} + \mathbf{x}, \mathbf{x}) d\mathbf{x} + \\
&\int \int \varrho_p(\mathbf{h} - \mathbf{x} + \mathbf{y}) \lambda_c(\mathbf{x}) \lambda_c(\mathbf{y}) d\mathbf{x} d\mathbf{y}. \quad (27)
\end{aligned}$$

The second term in this expression can also be written as:

$$\begin{aligned}
&\int \int \varrho_p(\mathbf{h} - \mathbf{x} + \mathbf{y}) \lambda_c(\mathbf{x}) \lambda_c(\mathbf{y}) d\mathbf{x} d\mathbf{y} \\
&= \int \varrho_p(\mathbf{z}) \int \lambda_c(\mathbf{x}) \lambda_c(\mathbf{z} - \mathbf{h} + \mathbf{x}) d\mathbf{x} d\mathbf{z} \\
&= \int \varrho_p(\mathbf{z}) a_{\lambda_c}(\mathbf{h} - \mathbf{z}) d\mathbf{z} \\
&= (\varrho_p * a_{\lambda_c})(\mathbf{h}), \quad (28)
\end{aligned}$$

where we made the substitution  $\mathbf{z} = \mathbf{h} - \mathbf{x} + \mathbf{y}$ . For our case, each cluster has a single point distributed according to  $k$ , hence  $\varrho_c = 0$ ,  $\lambda_c(\mathbf{x}) = k(\mathbf{x})$ , and  $\lambda = \lambda_p$ . Thus,  $g(\mathbf{h})$  can be written as:

$$g(\mathbf{h}) = \varrho(\mathbf{h}) / \lambda^2 = (g_p * a_k)(\mathbf{h}). \quad (29)$$

Substituting the PCF of uniform jittering process in Equation 21 into this expression, we can get the following equation for the PCF of the new process:

$$g(\mathbf{h}) = \frac{1}{n} \sum_{\text{unique } \mathbf{h}_i} a_k(\mathbf{h} - \mathbf{h}_i) \quad a_{\Pi_V}(\mathbf{h}) \neq 0. \quad (30)$$

Finally, we substitute this expression into the second term in Equation 7:

$$\begin{aligned}
&\int a_f(\mathbf{h}) g(\mathbf{h}) d\mathbf{h} \\
&= \frac{1}{n} \sum_{\text{unique } \mathbf{h}_i} \int a_f(\mathbf{h}) a_k(\mathbf{h} - \mathbf{h}_i) d\mathbf{h} \\
&= \frac{1}{n} \sum_{\text{unique } \mathbf{h}_i} (a_f(\mathbf{h}') * a_k(-\mathbf{h}'))(\mathbf{h}_i) \quad (31)
\end{aligned}$$

Comparing this expression with Equation 24, first row, we can see that  $a_f(\mathbf{h})$  in the variance expressions for uniform jitter sampling is replaced by  $(a_f(\mathbf{h}') * a_k(-\mathbf{h}'))(\mathbf{h}) = [f(\mathbf{h}') * (a_k * f)(-\mathbf{h}')](\mathbf{h})$ .

#### A.5 Variance of Isotropic Jitter Sampling

For the derivation of the pair correlation function (PCF) of this point process, we will utilize an unbiased estimator for PCFs of

isotropic point processes [Illian et al. 2008; Öztireli and Gross 2012]. This estimator can be written as:

$$g(\|\mathbf{h}\|) = \frac{1}{\lambda^2} \frac{1}{|\partial V_d| \|\mathbf{h}\|^{d-1}} \sum_{i \neq j} \frac{k(\|\mathbf{h}\| - \|\mathbf{h}_{ij}\|)}{a_{\Pi_V}(\mathbf{h}_{ij})}, \quad (32)$$

where  $|\partial V_d|$  is the area of the unit hypersphere in  $d$  dimensions, and  $k$  is a kernel function such as a Gaussian for smoothing the estimate. Assuming that we are using this estimator with many points in the point sets generated by isotropic jitter sampling, we can approximate that we always get the same number of points and thus the same set of distances. Under this assumption, and since the estimator is unbiased, we can get a zero variance and hence zero error estimator by setting  $k$  to be the Dirac delta. Inserting Equation 32 with  $k = \delta$  into the second term in Equation 7, we obtain the following:

$$\begin{aligned} \int a_f(\mathbf{h})g(\mathbf{h})d\mathbf{h} &= \frac{1}{\lambda^2 |\partial V_d|} \sum_{i \neq j} \int \frac{a_f(\mathbf{h})\delta(\|\mathbf{h}\| - \|\mathbf{h}_{ij}\|)}{a_{\Pi_V}(\mathbf{h}_{ij})\|\mathbf{h}\|^{d-1}} d\mathbf{h} \\ &= \frac{1}{\lambda^2} \sum_{i \neq j} \frac{\int a_f(\mathbf{h})\delta(\|\mathbf{h}\| - \|\mathbf{h}_{ij}\|)d\mathbf{h}}{a_{\Pi_V}(\mathbf{h}_{ij})|\partial V_d| \|\mathbf{h}_{ij}\|^{d-1}}. \end{aligned} \quad (33)$$

Using the same equations for  $a_{\Pi_V}(\mathbf{h}_{ij})$  and number of unique  $\mathbf{h}$  vectors as in Section A.3, we can simplify this into:

$$\frac{1}{\lambda} \sum_{\text{unique } \mathbf{h}_k} \frac{1}{|\partial V_d| \|\mathbf{h}_k\|^{d-1}} \int a_f(\mathbf{h})\delta(\|\mathbf{h}\| - \|\mathbf{h}_k\|)d\mathbf{h}. \quad (34)$$

This is the sum of radial averages of  $a_f$  at each unique  $\mathbf{h}_k$ . In practice, the number of points and hence distances within the support of  $V$  can change for different rotations, introducing a slight deviation from this theoretical value.

## A.6 Point in a Stratum for Translated & Rotated Grids

Each instance of isotropic or r-uniform jitter sampling is a translated and rotated grid. One way to generate such a distribution is taking a grid larger than the unit hypercube we are considering, and applying a random global translation and rotation to all points. Although this process is simple and generalizable to high dimensions, it is more convenient and for some cases more efficient to generate the pattern with stratum-wise operations, i.e. generating the point(s) in a stratum given the coordinates of that stratum. In this section, we derive the necessary algorithm for the two dimensional case. Although the algorithm introduces extra steps after generating the random numbers for jittering, this additional cost becomes negligible even for moderately complex scenes due to the cost of tracing rays.

We would like to find the point(s) of a rotated and translated grid that fall into a given stratum. Let us denote the length of an edge of the stratum by  $\Delta$  (we assume square strata), the coordinates of the lower-end corner of the stratum by  $\Delta \mathbf{m} = \Delta[m_x \ m_y]^T$  (so that  $m_x$  and  $m_y$  are integers), and those of the point(s) that fall into this stratum by  $\mathbf{x} = [x \ y]^T$ . Then, for a random translation  $\Delta \mathbf{t}$  and rotation angle  $0 \leq \theta \leq \pi/2$ , and a vector (with integer  $i$  and  $j$ )  $\mathbf{c} = [i + 0.5 \ j + 0.5]^T$ :

$$\mathbf{x} = \mathbf{R}_\theta \Delta \mathbf{c} + \Delta \mathbf{t}, \quad (35)$$

with  $\mathbf{R}_\theta$  representing rotation of angle  $\theta$ . Hence, we assume that the point that was originally in the center of the stratum with the lower-left corner  $\Delta [i \ j]^T$  falls into the stratum we are considering with the lower-left corner  $\Delta \mathbf{m}$  after translation and rotation.

For  $\mathbf{x}$  to be in the stratum,  $\Delta m_x \leq x \leq \Delta(m_x + 1)$  and  $\Delta m_y \leq y \leq \Delta(m_y + 1)$ . Substituting the expressions for  $x$  and  $y$  into these inequalities, we can get the following conditions for  $i$  and  $j$ :

$$\begin{aligned} \mathbf{r}_1^T(\mathbf{m} - \mathbf{t}) - \frac{1}{2} \leq i \leq \mathbf{r}_1^T(\mathbf{m} - \mathbf{t}) - \frac{1}{2} + \sin(\theta) + \cos(\theta) \\ \mathbf{r}_2^T(\mathbf{m} - \mathbf{t}) - \frac{1}{2} - \sin(\theta) \leq j \leq \mathbf{r}_2^T(\mathbf{m} - \mathbf{t}) - \frac{1}{2} + \cos(\theta), \end{aligned} \quad (36)$$

where  $\mathbf{r}_1$  and  $\mathbf{r}_2$  denote the first and second columns of the rotation matrix  $\mathbf{R}_\theta$ . We can then plug the integers  $i, j$  within these limits into Equation 35 to compute the  $\mathbf{x}$ '(s) that are in the current stratum.

## REFERENCES

- BEETS, K. AND BARRON, D. 2000. Super-sampling anti-aliasing analyzed. *Beyond3D*, 1–22.
- BELCOUR, L., SOLER, C., SUBR, K., HOLZSCHUCH, N., AND DURAND, F. 2013. 5d covariance tracing for efficient defocus and motion blur. *ACM Trans. Graph.* 32, 3 (July), 31:1–31:18.
- BOWERS, J., WANG, R., WEI, L.-Y., AND MALETZ, D. 2010. Parallel poisson disk sampling with spectrum analysis on surfaces. *ACM Trans. Graph.* 29, 166:1–166:10.
- CHIU, K., SHIRLEY, P., AND WANG, C. 1994. Graphics gems iv. Academic Press Professional, Inc., San Diego, CA, USA, Chapter Multi-jittered Sampling, 370–374.
- COOK, R. L. 1986. Stochastic sampling in computer graphics. *ACM Trans. Graph.* 5, 1, 51–72.
- DICK, J. AND PILLICHSHAMMER, F. 2010. *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*. Cambridge University Press, New York, NY, USA.
- DUNBAR, D. AND HUMPHREYS, G. 2006. A spatial data structure for fast poisson-disk sample generation. *ACM Trans. Graph.* 25, 503–508.
- DURAND, F. 2011. A frequency analysis of monte-carlo and other numerical integration schemes. Tech. Rep. MIT-CSAILTR-2011-052, CSAIL, MIT, MA, February.
- DURAND, F., HOLZSCHUCH, N., SOLER, C., CHAN, E., AND SILLION, F. X. 2005. A frequency analysis of light transport. *ACM Trans. Graph.* 24, 3 (July), 1115–1126.
- EBEIDA, M. S., MITCHELL, S. A., PATNEY, A., DAVIDSON, A. A., AND OWENS, J. D. 2012. A simple algorithm for maximal poisson-disk sampling in high dimensions. *Comput. Graph. Forum* 31, 2pt4, 785–794.
- EGAN, K., DURAND, F., AND RAMAMOORTHY, R. 2011. Practical filtering for efficient ray-traced directional occlusion. *ACM Trans. Graph.* 30, 6 (Dec.), 180:1–180:10.
- EGAN, K., HECHT, F., DURAND, F., AND RAMAMOORTHY, R. 2011. Frequency analysis and sheared filtering for shadow light fields of complex occluders. *ACM Trans. Graph.* 30, 2 (Apr.), 9:1–9:13.
- EGAN, K., TSENG, Y.-T., HOLZSCHUCH, N., DURAND, F., AND RAMAMOORTHY, R. 2009. Frequency analysis and sheared reconstruction for rendering motion blur. *ACM Trans. Graph.* 28, 3 (July), 93:1–93:13.
- FATTAL, R. 2011. Blue-noise point sampling using kernel density model. *ACM Trans. Graph.* 30, 4 (July), 48:1–48:12.
- FELSENSTEIN, J. 1975. A pain in the torus: Some difficulties with models of isolation by distance. *The American Naturalist* 109, 967, 359–368.
- GUAN, Y. 2008. Variance estimation for statistics computed from inhomogeneous spatial point processes. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 70, 1, pp. 175–190.
- HACHISUKA, T., JAROSZ, W., WEISTROFFER, R. P., DALE, K., HUMPHREYS, G., ZWICKER, M., AND JENSEN, H. W. 2008. Multi-dimensional adaptive sampling and reconstruction for ray tracing. *ACM Trans. Graph.* 27, 3 (Aug.), 33:1–33:10.

- HAMMERSLEY, J. M. AND MORTON, K. W. 1956. A new monte carlo technique: antithetic variates. *Mathematical Proceedings of the Cambridge Philosophical Society* 52, 449–475.
- HASSELGREN, J., AKENINE-MILLER, T., AND LAINE, S. 2005. A family of inexpensive sampling schemes. *Comput. Graph. Forum* 24, 4, 843–848.
- HECK, D., SCHLÖMER, T., AND DEUSSEN, O. 2013. Blue noise sampling with controlled aliasing. *ACM Trans. Graph.* 32, 3 (July), 25:1–25:12.
- ILLIAN, J., PENTTINEN, A., STOYAN, H., AND STOYAN, D., Eds. 2008. *Statistical Analysis and Modelling of Spatial Point Patterns*. John Wiley and Sons, Ltd.
- KELLER, A., PREMOZE, S., AND RAAB, M. 2012. Advanced (quasi) monte carlo methods for image synthesis. In *ACM SIGGRAPH 2012 Courses*. SIGGRAPH '12. ACM, New York, NY, USA, 21:1–21:46.
- KENSLER, A. 2013. Correlated multi-jittered sampling. Tech. Rep. Pixar Technical Memo 13-01, Pixar. March.
- KIRK, D. AND ARVO, J. 1991. Unbiased sampling techniques for image synthesis. *SIGGRAPH Comput. Graph.* 25, 4 (July), 153–156.
- KOLLIG, T. AND KELLER, A. 2002. Efficient multidimensional sampling. *Comput. Graph. Forum* 21, 3, 557–563.
- LAGAE, A. AND DUTRÉ, P. 2008. A comparison of methods for generating Poisson disk distributions. *Comput. Graph. Forum* 27, 1 (March), 114–129.
- MEHTA, S. U., WANG, B., AND RAMAMOORTHY, R. 2012. Axis-aligned filtering for interactive sampled soft shadows. *ACM Trans. Graph.* 31, 6 (Nov.), 163:1–163:10.
- MEHTA, S. U., WANG, B., RAMAMOORTHY, R., AND DURAND, F. 2013. Axis-aligned filtering for interactive physically-based diffuse indirect lighting. *ACM Trans. Graph.* 32, 4 (July), 96:1–96:12.
- MEHTA, S. U., YAO, J., RAMAMOORTHY, R., AND DURAND, F. 2014. Factored axis-aligned filtering for rendering multiple distribution effects. *ACM Trans. Graph.* 33, 4 (July), 57:1–57:12.
- MITCHELL, D. P. 1987. Generating antialiased images at low sampling densities. *SIGGRAPH Comput. Graph.* 21, 4 (Aug.), 65–72.
- MITCHELL, D. P. 1991. Spectrally optimal sampling for distribution ray tracing. *SIGGRAPH Comput. Graph.* 25, 4 (July), 157–164.
- MITCHELL, D. P. 1996. Consequences of stratified sampling in graphics. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '96. ACM, New York, NY, USA, 277–280.
- MØLLER, J. AND WAAGEPETERSEN, R. P. 2004. *Statistical inference and simulation for spatial point processes*. Chapman & Hall/CRC, 2003, Boca Raton (Fl.), London, New York.
- NIEDERREITER, H. 1992. *Quasi-Monte Carlo Methods*. John Wiley & Sons, Ltd.
- OVERBECK, R. S., DONNER, C., AND RAMAMOORTHY, R. 2009. Adaptive wavelet rendering. *ACM Trans. Graph.* 28, 5 (Dec.), 140:1–140:12.
- OWEN, A. B. 2008. Local antithetic sampling with scrambled nets. *Annals of Statistics* 36, 2319–2343.
- OWEN, A. B. 2013. *Monte Carlo Theory, Methods and Examples*. To be published.
- ÖZTIRELI, A. C. AND GROSS, M. 2012. Analysis and synthesis of point distributions based on pair correlation. *ACM Trans. Graph.* 31, 6 (Nov.), 170:1–170:10.
- PHARR, M. AND HUMPHREYS, G. 2010. *Physically Based Rendering, Second Edition: From Theory To Implementation*, 2nd ed. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
- PILLEBOUE, A., SINGH, G., COEURJOLLY, D., KAZHDAN, M., AND OSTROMOUKHOV, V. 2015. Variance analysis for monte carlo integration. *ACM Trans. Graph.* 34, 4 (July), 124:1–124:14.
- RAMAMOORTHY, R., ANDERSON, J., MEYER, M., AND NOWROUZEZAHRAI, D. 2012. A theory of monte carlo visibility sampling. *ACM Trans. Graph.* 31, 5 (Sept.), 121:1–121:16.
- RAMAMOORTHY, R. AND HANRAHAN, P. 2004. A signal-processing framework for reflection. *ACM Trans. Graph.* 23, 4 (Oct.), 1004–1042.
- SCHMALTZ, C., GWOSDEK, P., BRUHN, A., AND WEICKERT, J. 2010. Electrostatic halftoning. *Comput. Graph. Forum* 29, 8, 2313–2327.
- SHIRLEY, P. 1991. Discrepancy as a quality measure for sample distributions. In *In Eurographics '91*. Elsevier Science Publishers, 183–194.
- SOLER, C., SUBR, K., DURAND, F., HOLZSCHUCH, N., AND SILLION, F. 2009. Fourier depth of field. *ACM Trans. Graph.* 28, 2 (May), 18:1–18:12.
- STOYAN, D. AND STOYAN, H. 1994. *Fractals, random shapes, and point fields : methods of geometrical statistics*. Wiley series in probability and mathematical statistics. Wiley, Chichester, New York.
- SUBR, K. AND KAUTZ, J. 2013. Fourier analysis of stochastic sampling strategies for assessing bias and variance in integration. *ACM Trans. Graph.* 32, 4 (July), 128:1–128:12.
- SUBR, K., NOWROUZEZAHRAI, D., JAROSZ, W., KAUTZ, J., AND MITCHELL, K. 2014. Error analysis of estimators that use combinations of stochastic sampling strategies for direct illumination. *Comput. Graph. Forum* 33, 4, 93–102.
- WEI, L.-Y. 2008. Parallel poisson disk sampling. *ACM Trans. Graph.* 27, 20:1–20:9.
- WEI, L.-Y. 2010. Multi-class blue noise sampling. *ACM Trans. Graph.* 29, 4 (July), 79:1–79:8.
- WEI, L.-Y. AND WANG, R. 2011. Differential domain analysis for non-uniform sampling. *ACM Trans. Graph.* 30, 4 (July), 50:1–50:10.
- WHITTED, T. 1980. An improved illumination model for shaded display. *Commun. ACM* 23, 6 (June), 343–349.

Received Month YYYY; accepted Month YYYY