Integration with Stochastic Point Processes Supplementary Material

A. CENGİZ ÖZTİRELİ
ETH Zurich

1. BIAS & VARIANCE OF SAMPLING PATTERNS

For reference, here we provide the essential formulae for bias and variance of sampling patterns, as derived in the paper.

1.1 The Integral Estimator

Given a function $f : \mathbb{R}^d \to \mathbb{R}^+$, 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:

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

for some positive weights w_i and sample points $\mathbf{x}_i$. We will assume that the support is a unit hypercube, $|V| = 1$ for simplicity.

1.2 General Point Processes

$$\mathbb{E}\hat{I} = \mathbb{E}\sum w(\mathbf{x}_i)f(\mathbf{x}_i) = \int_V w(\mathbf{x})f(\mathbf{x})\lambda(\mathbf{x})d\mathbf{x}. \tag{2}$$

$$\mathbb{E}\hat{I}^2 = \int_{V\times V} w(\mathbf{x})f(\mathbf{x})w(\mathbf{y})f(\mathbf{y})\varrho(\mathbf{x},\mathbf{y})d\mathbf{x}d\mathbf{y} + \int_V w^2(\mathbf{x})f^2(\mathbf{x})\lambda(\mathbf{x})d\mathbf{x}. \tag{3}$$

The bias and variance can then be computed using standard formulae: $bias(\hat{I}) = I - \mathbb{E}\hat{I}$, $var(\hat{I}) = \mathbb{E}\hat{I}^2 - (\mathbb{E}\hat{I})^2$.

1.3 Stationary Point Processes

The estimator with stationary point processes is unbiased. The variance is given by the following two alternative forms.

$$var(\hat{I}) = \frac{1}{\lambda}\int f^2(\mathbf{x})d\mathbf{x} + \int f(\mathbf{x})f(\mathbf{y})(g(\mathbf{x}-\mathbf{y})-1)d\mathbf{x}d\mathbf{y}. \tag{4}$$

$$var(\hat{I}) = \frac{1}{\lambda}\int f^2(\mathbf{x})d\mathbf{x} + \int a_f(\mathbf{h})g(\mathbf{h})d\mathbf{h} - \left(\int f(\mathbf{x})d\mathbf{x}\right)^2. \tag{5}$$

2. ADAPTIVE SAMPLING

Adaptive sampling refers to adjusting the density of points according to the integrand f such that the variance of the estimator $\hat{I}$ is decreased. It is obtained by assuming an interaction model between points, and setting a specialized spatially varying density for a given integrand.

Locally scaled processes One way to achieve this is via defining a distance measure adapted to the desired density such that distances become smaller in areas of high density. This is the idea behind adaptive and anisotropic blue noise sampling [Li et al. 2010; Wei and Wang 2011; Chen et al. 2013], and known as locally scaled point processes [Hahn et al. 2003] in statistics. Although this is useful for applications such as halftoning and stippling [Schmaltz et al. 2010; Fattal 2011], it involves a costly optimization with lots of evaluations of the density function, which translates into tracing many rays for rendering.

Intensity-reweighted stationary processes Instead, 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 $\varrho(\mathbf{x},\mathbf{y}) = \lambda(\mathbf{x})\lambda(\mathbf{y})g(\mathbf{x}-\mathbf{y})$. Defining $v(\mathbf{x}) = \lambda(\mathbf{x})w(\mathbf{x})f(\mathbf{x})$, and plugging the expression for ϱ into the equations for bias and variance of general point processes in Section 1.2, we get

$$bias(\hat{I}) = I - \int v(\mathbf{x})d\mathbf{x}, \tag{6}$$

$$var(\hat{I}) = \int \frac{v^2(\mathbf{x})}{\lambda(\mathbf{x})}d\mathbf{x} + \int v(\mathbf{x})v(\mathbf{y})(g(\mathbf{x}-\mathbf{y})-1)d\mathbf{x}d\mathbf{y}, \tag{7}$$

where we dropped the integration domains for brevity. A common variate of adaptive sampling is importance sampling where the weighting is chosen as $w(\mathbf{x}) = 1/\lambda(\mathbf{x})$. For this case, $v(\mathbf{x}) = f(\mathbf{x})$, bias vanishes, and Equation 7 shows how density and correlations can be individually adjusted according to the integrand f, by choosing the intensity in the first term and PCF in the second term, respectively. [1]

General unbiased importance sampling Although not common for rendering applications, some distributions may not be accurately described by second-order intensity-reweighted stationary point processes. For these processes, assuming an unbiased estimator (i.e. $w(\mathbf{x}) = 1/\lambda(\mathbf{x})$), the variance is given by the following expression:

$$var(\hat{I}) = \int \frac{f^2(\mathbf{x})}{\lambda(\mathbf{x})}d\mathbf{x} + \int f(\mathbf{x})f(\mathbf{y})\frac{\varrho(\mathbf{x},\mathbf{y})}{\lambda(\mathbf{x})\lambda(\mathbf{y})}d\mathbf{x}d\mathbf{y} - \left(\int f(\mathbf{x})d\mathbf{x}\right)^2. \tag{8}$$

[1] In a point process model, the number of points within a window can vary for different realizations. This is why this expression cannot be directly used to compute the variance of standard importance sampling, where the points are randomly distributed with $g(\mathbf{x}-\mathbf{y}) = 1$. In this case, although the second term vanishes, even if we set λ proportional to f, the variance does not vanish, since for a point process where the points are randomly distributed, the number of points in a window can wildly vary for different realizations especially when λ is low. For a fixed number of randomly distributed points within a window, we derive the variance in Section 3.1.

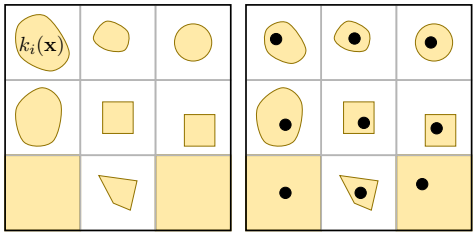

Fig. 1. A general uncorrelated random jittering where the location of the point in each stratum S_i follows a distribution $k_i(\mathbf{x})$ that is independent of those in the other strata.

The first term agrees with the well-known observation in importance sampling that the intensity should follow the integrand. To minimize the second term given a fixed intensity function, the second order product density ϱ should thus be as small as possible. One way to lower ϱ is by increasing the distances between the points, as is the case for blue noise distributions or low discrepancy patterns. However, note that what determines the variance is the integral of ϱ convolved and multiplied with f. Hence, it is possible to adapt ϱ further, going beyond the conventional methods, which we leave as future work.

3. BIAS & VARIANCE OF STRATIFIED SAMPLING

In this section, we provide further examples on how the analysis of stratified sampling patterns with point processes can be utilized to study important special cases.

3.1 Random Jitter Sampling

In random jitter sampling, the distribution of the point in each stratum in independent. More precisely, given a set of n strata S_i such that $S_i \cap S_j = \varnothing$ for $i \neq j$, the probability of finding a point in S_i at a location $\mathbf{x}$ is given by $k_i(\mathbf{x})$. We show this setting when the strata tile the unit square regularly in Figure 1. In Section 5.1 of the paper, the bias and variance of random jitter sampling is derived as follows:

$$bias(\hat{I}) = I - \sum_{i=1}^{n} \int w(\mathbf{x}) f(\mathbf{x}) k_i(\mathbf{x}) d\mathbf{x}, \tag{9}$$

$$var(\hat{I}) = \sum_{i=1}^{n} \int w^2(\mathbf{x}) f^2(\mathbf{x}) k_i(\mathbf{x}) d\mathbf{x} - \left(\int w(\mathbf{x}) f(\mathbf{x}) k_i(\mathbf{x}) d\mathbf{x} \right)^2, \tag{10}$$

where the integrations for the i^{th} term are over S_i, or equivalently V or $\mathbb{R}^d$.

Importance sampling We can easily derive the expressions for jittered importance sampling by setting $w(\mathbf{x}) = 1/\lambda(\mathbf{x}) = 1/\sum_{i=1}^{n} k_i(\mathbf{x})$, which is $w(\mathbf{x}) = 1/k_i(\mathbf{x})$ for $\mathbf{x} \in S_i$, since the supports of k_i are disjoint. Hence, we get an unbiased estimator with variance

$$var(\hat{I}) = \sum_{i=1}^{n} \int_{S_i} f^2(\mathbf{x}) / k_i(\mathbf{x}) d\mathbf{x} - \left(\int_{S_i} f(\mathbf{x}) d\mathbf{x} \right)^2. \tag{11}$$

This equation also gives the variance of standard importance sampling or random sampling by setting $n = 1$. [2]

[2] In Appendix C, we derive Equation 11 and consequently Equation 12 using standard statistical methods, and also explain why they differ slightly from the expressions in a recent work [Subr et al. 2014].

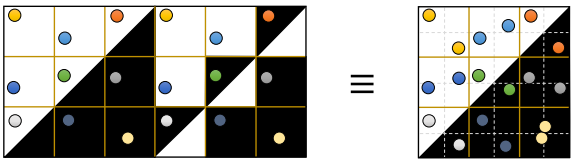

Fig. 2. Mirrored jitter sampling is based on the idea of locally mirroring parts of the integrand within each stratum and exploiting the negative correlation between the original and mirrored versions. Conceptually, the sampling pattern puts the original and locally mirrored versions side-by-side, and applies uniform jittering to each pair of original/mirrored stratum, independent of the other pairs of strata (left). In practice, this translates into taking a random point in a stratum, and mirroring it around the center of that stratum (right).

To simplify the analysis in the next sections, it is more convenient to define the functions $s_i(\mathbf{x}) = f(\mathbf{x} + \mathbf{o}_i)\Pi_S(\mathbf{x})$, where Π_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. We can then write the variance of standard random jittering, where the sampling point in each stratum is distributed uniformly in the stratum, as:

$$var(\hat{I}) = \frac{1}{n^2} \sum_{i=1}^{n} n \int s_i^2(\mathbf{x}) d\mathbf{x} - \left(n \int s_i(\mathbf{x}) d\mathbf{x} \right)^2, \tag{12}$$

which is also shown in Table I, Equation $T1$ for reference.

3.2 Mirrored Jitter Sampling

For some functions, each $s_i(\mathbf{x})$ is strongly negatively correlated with $s_i^-(\mathbf{x}) := s_i(-\mathbf{x})$, which is the local patch of f mirrored around the origin $\mathbf{o}_i$ of the stratum S_i. These correlations can be exploited to reduce the variance with a local variate of antithetic sampling [Owen 2013; Subr et al. 2014].

Denoting the mirrored strata with S_i^-, conceptually, mirrored jitter sampling puts the strata S_i and S_i^- side-by-side as in Figure 2, left. It then samples each pair S_i, S_i^- by uniform jittering, independent of the other pairs of strata. In practice, this can be realized by taking a random point $\mathbf{x}$ in each S_i, and placing 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 2, right). In Appendix A.1 of the paper, we derive (with $w = 1/n$) that the bias of the resulting estimator is zero, and the variance is:

$$\begin{aligned} var(\hat{I}) = \frac{1}{n^2} \sum_{i=1}^{m} n \int s_i(\mathbf{x}) s_i^-(\mathbf{x}) d\mathbf{x} \\ + n \int s_i^2(\mathbf{x}) d\mathbf{x} - \left(n \int s_i(\mathbf{x}) d\mathbf{x} \right)^2. \end{aligned} \tag{13}$$

Here, $m = n/2$, since there are actually only m strata and $n = 2m$ points. The equation confirms the intuition that negatively correlated s_i, s_i^- pairs provide low variance. In Section 4, we will utilize this expression and 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.

3.3 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

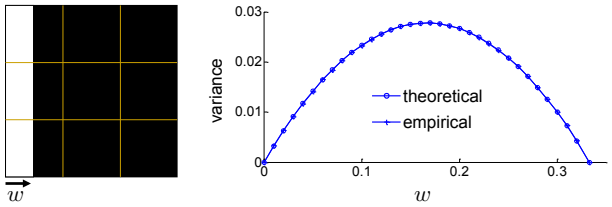

Fig. 3. The theoretical and empirically computed variance of the binary function on the left is plotted as a function of w, the width of the non-zero region. The empirical values match exactly with the theoretical ones.

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 $\lambda = n$. Hence, with $w = 1/n$, the resulting estimator is unbiased as is always the case for stationary processes. We derive the variance in Section 5.1 and Appendix A.2 of the paper as:

$$var(\hat{I}) = \frac{1}{n}\int f^2(\mathbf{x})d\mathbf{x} + \frac{1}{n^2}\sum_{i\neq j}\frac{a_f(\mathbf{h}_{ij})}{a_{\Pi_V}(\mathbf{h}_{ij})} - \left(\int f(\mathbf{x})d\mathbf{x}\right)^2, \tag{14}$$

where $\mathbf{h}_{ij} = \mathbf{x}_i - \mathbf{x}_j$, a_{Π_V} is the autocorrelation of the function Π_V, and Π_V is 1 inside V and zero otherwise.

Example We show a simple case where Equation 14 is utilized to derive an analytic formula for the variance of an integrand in Figure 3. The function f is 1 in a region of width w and height 1, and zero otherwise as shown in the figure. For $0 \leqslant w \leqslant 1/3$, we would like to compute how the variance changes. Assuming a uniformly jittered pattern with spacing of $1/3$, the length of each stratum is $1/3$. We can thus observe that the variance is zero for $w = 0$ and $w = 1/3$, since the function is aligned with the strata in those cases, and should reach its maximum when $w = 1/6$. To derive the exact variance formula, first note that the integral of f (and hence that of f^2, since $f = f^2$) is simply $I = w$, and $a_f(\mathbf{h})/a_{\Pi_V}(\mathbf{h}) = (w-|h_x|)(1-|h_y|)/(1-|h_x|)(1-|h_y|) = (w-|h_x|)/(1-|h_x|)$ for $|h_x| < w$ and zero otherwise. Hence, we get non-zero $a_f(\mathbf{h}_{ij})/a_{\Pi_V}(\mathbf{h}_{ij}) = w$ only for $\mathbf{h}_{ij} = \mathbf{x}_i - \mathbf{x}_j$ with zero x components, since for all other $\mathbf{h}_{ij}$, the absolute value of the x component is bigger than or equal to $1/3$, the spacing between points. As there are 18 $\mathbf{h}_{ij}$ with zero x components, variance is given by (Equation 14) $1/9w + 18/9^2w - w^2 = 1/3w - w^2$. This function is plotted along with the empirically computed variance for different w in Figure 3. They match perfectly as expected.

Strata-based variance formula The autocorrelation function a_f appearing in Equation 14 depends on the global properties of f, similar to spectral analysis. However, some cases of analysis involve assumptions on the local strata-based properties of f as in the previous example or in Section 4 for visibility sampling, or requires that variance of uniform jittering is compared with those of sampling patterns with strata-based variance expressions. For these cases, it is important to have a strata-based variance formula. We show in Appendix A.3 of the paper that such an expression can be derived starting from Equation 14:

$$var(\hat{I}) = \frac{1}{n^2}\sum_{ij} n\int s_i(\mathbf{x})s_j(\mathbf{x})d\mathbf{x} - n^2\int s_i(\mathbf{x})d\mathbf{x}\int s_j(\mathbf{x})d\mathbf{x}. \tag{15}$$

For brevity, we denote each term in this equation as c_{ij} such that $var(\hat{I}) = \frac{1}{n^2}\sum_{ij} c_{ij}$. This equation gives a practical way to analyze and compute variance of uniform jittering in terms of strata-based properties of the integrand as we will show below and in Section 4. So far, such an analysis has only been possible for 1D binary integrands under certain assumptions [Ramamoorthi et al. 2012] as we explain in the following example.

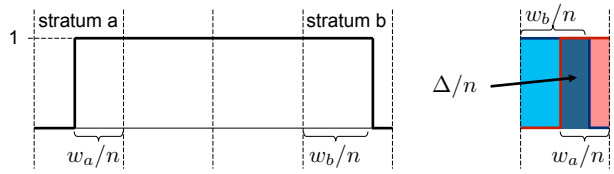

Fig. 4. The binary integrand considered by Ramamoorthi et al. [2012] (left) and the definitions to compute the variance with random and uniform jittering. The Δ denotes the intersection of stratum a and b.

Examples In their careful analysis of random and uniform jitter sampling, Ramamoorthi et al. [2012] derive the expected variance for the case of 1D binary integrands with two discontinuities in different strata (as summarized in Figure 6 of the mentioned paper [Ramamoorthi et al. 2012]). The type of the integrand considered is shown in Figure 4. Equation 15 allows us to easily compute the analytical variance in this case. Note that $c_{ij} = 0$ whenever the integrand is constant (0 or 1) in stratum S_i or S_j. Hence, the only terms that contribute to the variance are $c_{aa} = w_a - w_a^2$, $c_{ab} = c_{ba} = \Delta - w_a w_b$, and $c_{bb} = w_b - w_b^2$ (see Figure 4 for the notation). The intersection area is given by $\Delta = w_a + w_b - 1$ for $w_a + w_b > 1$, and 0 otherwise. Note that $c_{aa} + c_{bb}$ gives the variance of random jittering (Equation 12), and adding c_{ab} and c_{ba} to that sum results in the variance of uniform jittering (both up to the factor $1/n^2$). To find the average variance, Ramamoorthi et al. [2012] assume a uniform distribution for the locations of the discontinuities. Hence, the average variance is given by

$$\int_0^1\int_0^1 \frac{1}{n^2}\left[w_a - w_a^2 + w_b - w_b^2 + 2(\Delta - w_a w_b)\right] dw_a dw_b$$
$$= \frac{2}{n^2}\left(\frac{1}{2} - \frac{1}{3}\right) + \frac{2}{n^2}\left(\frac{1}{6} - \frac{1}{4}\right) = \frac{1}{3n^2} - \frac{1}{6n^2},$$

where the first term $1/3n^2$ is the variance of random jittering, and the result $1/6n^2$ is that of uniform jittering, agreeing with the previous results [Ramamoorthi et al. 2012].

In practice, it is possible to have more than one discontinuity within the strata. We present an example to illustrate good and bad cases for uniform jittering when there are two discontinuities within strata in Figure 5. In the example, two samples are used for integrating each of the two integrands integrand 1 and integrand 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 vs. random jittering As illustrated by the examples above, Equations 12 and 15 give us a direct way to analyze and compare uniform and random jittering. Let us define a matrix $\mathbf{C}_{ij} = c_{ij}$. By subtracting variance expressions for uniform and random jittering, it is easy to see that uniform sampling produces a lower variance if and only if $\varepsilon_{\mathbf{C}} = \mathbf{1}^T\mathbf{C}\mathbf{1} - \mathrm{Tr}(\mathbf{C}) < 0$, where $\mathrm{Tr}(\mathbf{C})$ denotes the trace of $\mathbf{C}$ and $\mathbf{1} = [1, \cdots]^T$, i.e. the sum of the

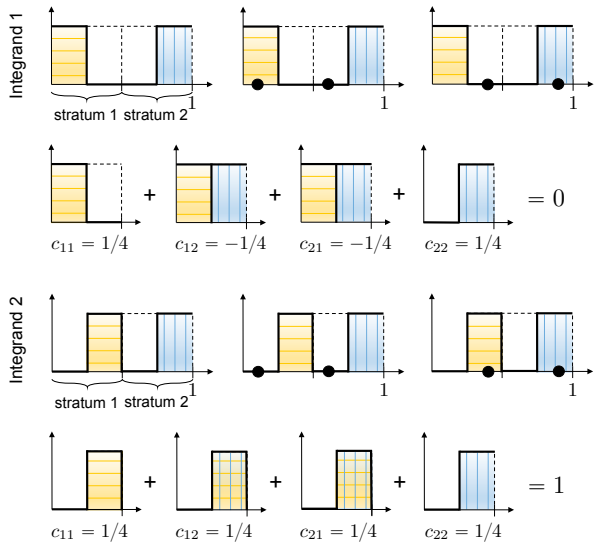

Fig. 5. We can compute exact variances for the integrands 1 and 2. For each case, the integrand itself is in the top-left figures. Top-middle and top-right figures illustrate two realizations of uniform jittering. Exact variance computations (up to a constant factor of $1/2^2$) are shown on the bottom rows. Due to the negatively correlated strata for integrand 1, we get zero variance. This can also be observed by considering different realizations of uniform jittering: the estimator always gives the correct result of $1/2$. This is in contrast with the estimator for integrand 2, where it can give 0 or 1.

non-diagonal elements is smaller than zero. We show four simple binary integrands along with their $\varepsilon_{\mathbf{C}}$ in Figure 6. For the top-left integrand, the strata containing the discontinuity are heavily correlated to each other, giving $\varepsilon_{\mathbf{C}} > 0$. The two edges of the top-middle integrand introduce negative correlations, leading to $\varepsilon_{\mathbf{C}} < 0$ and hence an advantage for uniform jittering. It has a similar advantage for the disk on the top-right, due to the negative correlations among the cells containing the boundaries of the disk. Finally, we show a square along with its $\mathbf{C}$ matrix on the bottom. We will present a similar theoretical analysis of visibility sampling in Section 4.

Spectral variance formula The relation $\mathscr{F}(a_f) = |\mathscr{F}(f)|^2$ with the Fourier transform $\mathscr{F}$ allows us to express the variance in Equation 14 in the spectral domain, as we derive in Appendix A.3 of the paper. Setting $F = \mathscr{F}(f)$, and denoting the set of points of the infinite regular grid of spacing $n^{1/d}$ centered at the origin in the frequency domain with ω_k, the resulting variance expression is

$$var(\hat{I}) = \sum_{\omega_k \neq 0} |F(\omega_k)|^2. \tag{16}$$

Due to the translation invariance of the process, the variance does not depend on the phase of F. This is in contrast with the mean squared error (MSE) of regular sampling, where the sampling points are placed on a regular grid. For a regular grid with a translation $\mathbf{t}$ with respect to the origin, the MSE of regular sampling is given by $mse(\hat{I}) = [\sum_{\omega_k \neq 0} |F(\omega_k)| \cos(\phi(\omega_k) + 2\pi\omega_k^T \mathbf{t})]^2$ (Appendix A.1), where ϕ denotes the phase of F. Hence, it is possible to get much lower MSE with uniform jitter sampling, for instance, for functions symmetric around one of the sampling points of the regular grid. To see this, note that for this case, the integrand is given by $f(\mathbf{x}) = f_s(\mathbf{x} - \mathbf{t})$ for an even function $f_s(\mathbf{x})$. Thus, $F(\omega) = F_s(\omega) e^{-i2\pi\omega^T \mathbf{t}}$ for the real Fourier transform F_s of f_s. The phase of $F(\omega)$, $\phi(\omega) = -2\pi\omega^T \mathbf{t}$ cancels out in the formula

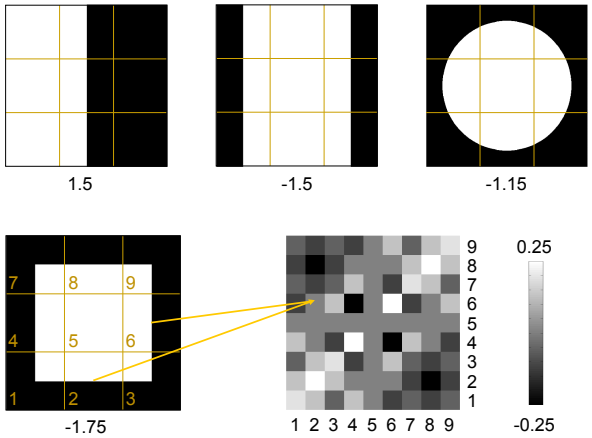

Fig. 6. Basic binary integrands with the strata overlaid and their $\varepsilon_{\mathbf{C}}$ below each integrand (also the matrix $\mathbf{C}$ for the bottom one). Negatively correlated strata give a clear advantage for uniform jittering (negative $\varepsilon_{\mathbf{C}}$). Each entry in the matrix $\mathbf{C}_{ij}$ quantifies how much the pair S_i and S_j contributes to the variance.

of $mse(\hat{I})$ for regular sampling, making all cosine terms equal to 1, and thus resulting in higher error than uniform jittering (please see Appendix A.1 for a derivation of the MSE of regular sampling and a comparison of the MSEs for Gaussian integrands).

3.4 Isotropic and R-Uniform Jitter Sampling

Isotropic jitter sampling is 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. 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 with $w = 1/\lambda$. In Appendix A.5 of the paper, we prove that the variance can be estimated by modifying the second term of variance of uniform jitter sampling (Equation 14) as:

$$\frac{1}{\lambda^2 |\partial V_d|} \sum_{i \neq j} \frac{\int a_f(\mathbf{h}) \delta(||\mathbf{h}|| - ||\mathbf{h}_{ij}||) d\mathbf{h}}{a_{\Pi_V}(\mathbf{h}_{ij}) ||\mathbf{h}_{ij}||^{d-1}}, \tag{17}$$

where $|\partial V_d|$ is the area of the unit hypersphere in d dimensions, and ϱ is the Dirac delta. 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 $\mathbf{h}_{ij}$.

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 17 is an average of variances of uniform jittering for all rotations, it lies between this minimum and maximum.

R-uniform jittering We call the rotated uniform jittering pattern with the minimum variance as *r(otated)-uniform* jittering. 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 anal-

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 Π_S (Π_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, $F = \mathscr{F}(f)$ is the Fourier transform of f, ω_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

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

Mirrored jitter sampling

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

Uniform jitter sampling

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

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

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

Uniform + random jitter sampling

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

Isotropic jitter sampling

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

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

ysis of how rotation can affect the variance for visibility sampling in Section 4.

4. APPLICATION TO VISIBILITY SAMPLING

We detail the analytic analysis of stratified sampling patterns when applied to visibility sampling for computing soft shadows in this section. We start with an introduction to the setting we are considering.

4.1 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. The image intensity at a point $\mathbf{x}$ on the image plane is computed by the following integral:

$$I(\mathbf{x}) = \int_V v(\mathbf{x},\mathbf{y})l(\mathbf{y})t(\mathbf{x},\mathbf{y})d\mathbf{y}, \quad (18)$$

where I is the intensity (radiance), V is the unit square as before, v is the binary visibility function, l is the lighting from the area light source, and t is a function that models light transport including BRDFs. We make the common assumption [Ramamoorthi et al. 2012; Mehta et al. 2012] that t can be taken out of the integral. Hence, for a given point $\mathbf{x}$ on the image plane, we are interested in the following integral:

$$I(\mathbf{x}) = \int_V v(\mathbf{x},\mathbf{y})l(\mathbf{y})d\mathbf{y}. \quad (19)$$

Once this is computed, it can be multiplied with $t(\mathbf{x})$ for all $\mathbf{x}$ to get the final rendering. 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.

4.2 Error Analysis and Comparisons

We need to consider the lighting l along with the visibility function v to analyze the error in Equation 19 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). 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:

$$e = \int_R var(\hat{I}(\mathbf{x}))d\mathbf{x}, \quad (20)$$

(we consider unbiased jittering patterns and hence mean squared error is equal to variance) over some area R of the pixel space. In the next sections, we compare different sampling strategies analytically and empirically to understand under what conditions on the light types and scene geometries they will perform well.

4.2.1 *Random Jitter Sampling.* The error of random jitter sampling (Equation 12) depends on stratum-wise characteristics without inter-strata correlation terms, and hence is of the form $e^{RND}(\mathbf{x}) = 1/n^2\sum_{i=1}^{n} e_i^{RND}(\mathbf{x})$. We thus first derive an expression for the contribution of the i^{th} stratum for pixel $\mathbf{x}$, i.e. $e_i^{RND}(\mathbf{x})$, to the variance. For further analysis and comparisons, we would like to write this term as a function of the normalized integral of the integrand within this stratum: $A_i(\mathbf{x}) := n\int s_i(\mathbf{x},\mathbf{y})d\mathbf{y}$, where $s_i(\mathbf{x},\mathbf{y})$ is the part of the integrand $v(\mathbf{x},\mathbf{y})l(\mathbf{y})$ in the stratum $S_i \in V$ on the light plane. Since the integrand is binary, the $A_i(\mathbf{x}) \in [0,1]$ is thus the normalized area of the stratum S_i covered by the non-zero part of the integrand.

Utilizing the fact that $s_i = s_i^2$ for binary integrands, we can write the variance (and thus MSE) terms for random jittering as

$$\begin{aligned} e_i^{RND}(\mathbf{x}) &= n\int s_i^2(\mathbf{x},\mathbf{y})d\mathbf{y} - \left(n\int s_i(\mathbf{x},\mathbf{y})d\mathbf{y}\right)^2 \\ &= A_i(\mathbf{x})(1 - A_i(\mathbf{x})). \end{aligned} \quad (21)$$

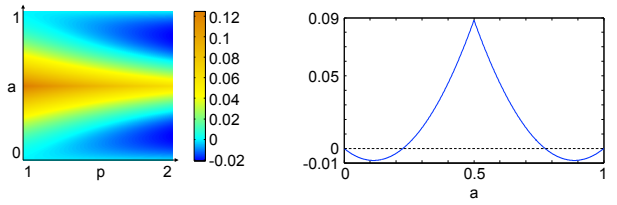

Fig. 7. MSE difference $2^{-p}e^{RND}(a) - e^{MIR}(a)$ is plotted as a function of the normalized area (a) covered by the support of the integrand in a given stratum, and the convergence rate p of random jitter sampling. The right graph shows the cross section for $p = 1.5$, the typical convergence rate of random jittering for discontinuous integrands.

This term will be largest when $A_i(\mathbf{x}) = 0.5$, and will get smaller as $A_i(\mathbf{x})$ becomes smaller or larger. Since only a single sample is taken in this stratum, it is better for random jitter sampling to have an integrand that mostly covers or is mostly nonexistent in the stratum.

4.2.2 *Mirrored Jitter Sampling.* Similar to random jitter sampling, the error for mirrored jitter sampling (Equation 13) is of the form $e^{MIR}(\mathbf{x}) = 1/n^2 \sum_{i=1}^{n} e_i^{MIR}(\mathbf{x})$ (note that in contrast to Equation 13, here we have n strata, and absorb the resulting constants into $e_i^{MIR}(\mathbf{x})$). We show in Appendix A.2 that under the assumption of a locally linear boundary of the support region of the integrand in a given stratum, the error of mirrored jitter sampling can be derived in terms of $A_i(\mathbf{x})$ as:

$$e_i^{MIR}(\mathbf{x}) = \begin{cases} \frac{1}{2}A_i(\mathbf{x}) - A_i^2(\mathbf{x}) & : A_i(\mathbf{x}) \in [0, \frac{1}{2}] \\ \frac{3}{2}A_i(\mathbf{x}) - \frac{1}{2} - A_i^2(\mathbf{x}) & : A_i(\mathbf{x}) \in [\frac{1}{2}, 1] \end{cases} \quad (22)$$

In contrast to random jitter sampling, mirrored jitter sampling has the least error $e_i^{MIR}(\mathbf{x}) = 0$ when $A_i(\mathbf{x}) = 0.5$, due to using two mirrored samples in the stratum.

Random jitter vs. mirrored jitter sampling There are n strata for both sampling patterns and hence n points for random jittering, and $2n$ points for mirrored jittering. This allows us to consider consistent strata. To account for the difference in the error due to the different number of points, we scale the resulting variance for random jittering by 2^{-p}, where p is the convergence rate for random jittering.

For a single stratum, we plot the resulting error difference $2^{-p}e_i^{RND}(A_i(\mathbf{x})) - e_i^{MIR}(A_i(\mathbf{x}))$ as a function of the area $A_i(\mathbf{x})$ and p in Figure 7, left. It has been observed that $p \in [1, 2]$ [Mitchell 1996; Ramamoorthi et al. 2012], and stay around 1.5 for integrands with discontinuities. We also plot the error difference for $p = 1.5$ in Figure 7, right, as a function of the area. Although random jitter sampling can perform slightly better for some of the a, p combinations, especially when the convergence rate p is high, mirrored jitter sampling has lower error for a much larger region in the a, p space, and performs considerably well for the common convergence rate $p = 1.5$.

Total error The analysis so far provides us with the exact error terms for given areas $A_i(\mathbf{x})$, but it does not give the total error computed over all pixels, and requires the knowledge of the area of the support of the integrand within each stratum. The total error for both random jitter and mirrored jitter sampling is given by $e = \int_R var(\hat{I}(\mathbf{x}))d\mathbf{x} = 1/n^2 \sum_{i=1}^{n} \int_R e_i(\mathbf{x})d\mathbf{x}$. Hence, we need to integrate the error over all pixels by computing $\int_R e_i(\mathbf{x})d\mathbf{x}$ for both random jitter and mirrored jitter sampling. We can compute this integral by considering how the area $A_i(\mathbf{x})$ changes for a given stratum S_i, as $\mathbf{x}$ moves on the image plane.

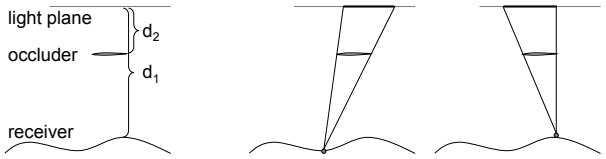

Fig. 8. (Left) The scene geometry assumed for some parts of the error analysis for soft shadows. (Right) For this geometry, the occluded area on the light plane translates and scales as the pixel location $\mathbf{x}$ moves in the image plane.

Under common assumptions on the occluders [Egan et al. 2011; Ramamoorthi et al. 2012; Mehta et al. 2012], the visibility function takes the form $v(\mathbf{x}, \mathbf{y}) = v(\tau(\mathbf{x})\mathbf{x} + (1 - \tau(\mathbf{x}))\mathbf{y})$, where $\tau(\mathbf{x}) = d_2/d_1(\mathbf{x})$, $d_1(\mathbf{x})$ is the distance from the receiver to the light plane, and d_2 is the (assumed to be constant) distance from the occluder to the light plane, as illustrated in Figure 8, left. Hence, at each pixel location $\mathbf{x}$, the visibility function just scales and translates on the light plane (Figure 8, right).

Assuming a square light source with $l(\mathbf{y}) = 1\ \forall \mathbf{y} \in V$, a smooth τ that does not scale the visibility function such that its support is smaller than a stratum's size, a sufficiently smooth boundary for the support of v, and that the whole soft shadow due to the occluder is rendered (i.e. the area R on the image plane is sufficiently large), each stratum is gradually completely occluded and unoccluded by $v(\mathbf{x}, \mathbf{y})$ several times as it slides over the light plane for different pixel locations $\mathbf{x}$. Hence, the area $A_i(\mathbf{x})$ of the unoccluded region in stratum S_i gradually changes from 0 to 1, or 1 to 0 several times, as $\mathbf{x}$ moves. The total error can thus be estimated by taking the integral $\int_0^1 e_i(\mathbf{x})dA_i(\mathbf{x})$. Denoting $a = A_i(\mathbf{x})$, for random jitter sampling, this gives $e_i^{RND} = \int_0^1 a(1-a)da = 1/6$, and similarly for mirrored jitter sampling $e_i^{MIR} = 1/24$. Hence, the ratio of the MSE's is $e^{RND}/e^{MIR} = 2^{2-p}$. For the typical convergence rate of $p = 1.5$, random jitter sampling will thus have $\sqrt{2}$ times as much variance as mirrored jitter sampling. Since $p \in [1, 2]$, mirrored jittering will always perform better than random jittering under the mentioned assumptions. For the same error level, random jittering needs to use $2^{2/p-1}$ times as much samples as mirrored jittering. Hence, for $p = 1.5$, random jittering requires approximately 25% more samples than mirrored jittering.

We illustrate several cases where these assumptions hold in Figure 9, 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, most strata go under almost complete occlusion and unocclusion, and the ratio approximately holds. We illustrate such cases in Figures 10, and 13. For the simpler occluding geometry in Figure 10, 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 10, bottom. For square lights, we thus observed an advantage of using mirrored jitter sampling in all the examples. However, for other light types, or more complex occluding geometries, this advantage disappears. (Please see Appendix B for an analytical derivation of the errors for mirrored jitter sampling and random jitter sampling for flat occluders and receivers, i.e. constant τ.)

4.2.3 *Uniform Jitter Sampling.* As elaborated in Section 3.3, uniform jitter sampling exploits negatively correlated strata to reduce variance. For the binary integrand we consider, this is more

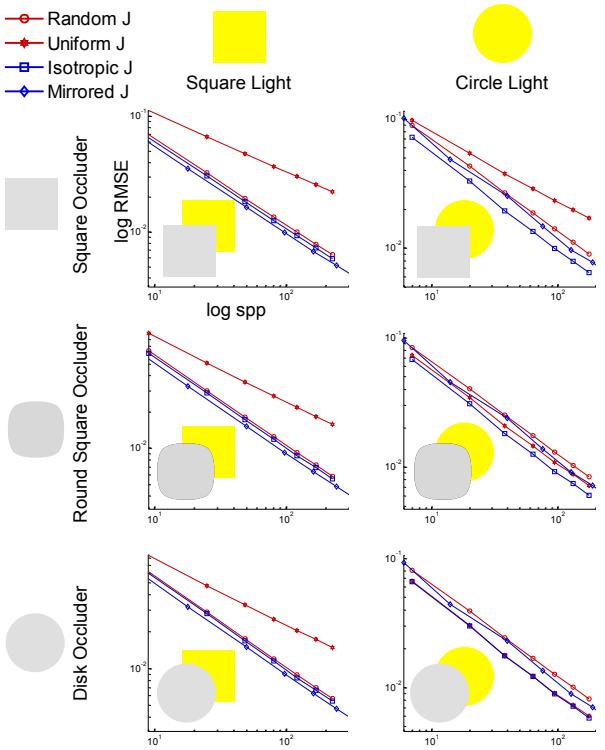

Fig. 9. RMSE plots for soft shadows with different combinations of basic occluders and light types (for all cases, $d_1 = 2d_2$).

apparent by writing the error as:

$$e^{UNI}(\mathbf{x}) = \frac{1}{n^2}\sum_{ij} A_{ij}(\mathbf{x}) - A_i(\mathbf{x})A_j(\mathbf{x}), \qquad (23)$$

where $A_{ij}(\mathbf{x}) = n\int s_i(\mathbf{x},\mathbf{y})s_j(\mathbf{x},\mathbf{y})d\mathbf{y}$ is the normalized area of the intersection of the supports of s_i and s_j. For negatively correlated strata, for example when s_i and s_j have disjoint supports, $A_{ij}(\mathbf{x}) = 0$, and the term $c_{ij}(\mathbf{x}) = A_{ij}(\mathbf{x}) - A_i(\mathbf{x})A_j(\mathbf{x})$ will decrease the error, conversely if the supports completely match, the error will increase by $1 - A_i(\mathbf{x})A_j(\mathbf{x})$.

Uniform vs. random jittering The error difference between uniform and random jittering is given by $e^{UNI}(\mathbf{x}) - e^{RND}(\mathbf{x}) = \frac{1}{n^2}\sum_{i\neq j} c_{ij}(\mathbf{x})$. These terms will favor random jittering for positively correlated strata. 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 (Figure 11, left), we can analytically prove that random jittering will perform better. As illustrated in Figure 11, left, for this case, the intersection area $A_{ij}(\mathbf{x})$ will be equal to the smaller of $A_i(\mathbf{x})$ and $A_j(\mathbf{x})$. Denoting the smaller one with the index j, we can thus set $A_{ij}(\mathbf{x}) = A_j(\mathbf{x})$, giving $c_{ij}(\mathbf{x}) = A_j(\mathbf{x})(1 - A_i(\mathbf{x}))$. Since $A_i(\mathbf{x}) \in [0,1]$ and $A_j(\mathbf{x}) \in [0,1]$, it is also true that $c_{ij} \in [0,1]$. Thus, each term will increase the error difference, resulting in $e^{UNI}(\mathbf{x}) - e^{RND}(\mathbf{x}) > 0$.

In general, the slopes of the boundary lines in different strata can be different (we assume a sufficiently smooth boundary that can be well represented by local lines in each stratum). Under the assumption [Egan et al. 2011; Ramamoorthi et al. 2012; Mehta et al. 2012] that the visibility function takes the form $v(\mathbf{x},\mathbf{y}) = v(\tau(\mathbf{x})\mathbf{x} + (1-\tau(\mathbf{x}))\mathbf{y})$ as above and in Figure 8, as $\mathbf{x}$ moves on the image plane, the slopes of the boundary lines in the strata will not change, since the visibility function can only scale or translate. Only the intersection point of the boundary lines in different strata, when any two strata are aligned, is a function of the pixel location $\mathbf{x}$. Hence, to estimate the total error, we can average c_{ij} over all intersection points, for each given pair of slopes (in Appendix B, we also present an analysis for constant $\tau(\mathbf{x})$).

In Figure 11, right, we plot the averaged c_{ij}, as a function of the angles the boundary lines make with one of the axes of the light plane. Local boundary lines with similar slopes increase c_{ij} and

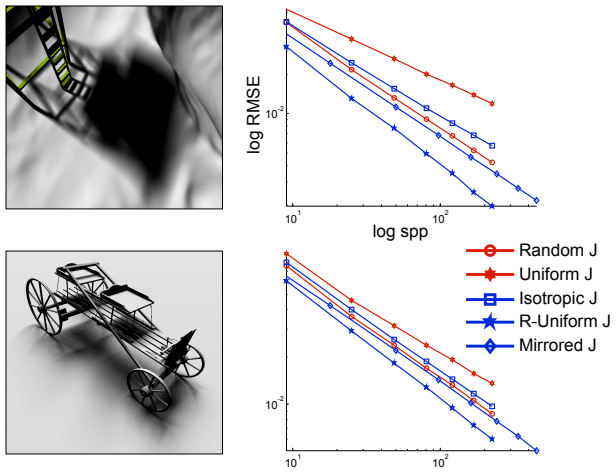

Fig. 10. RMS errors are plotted for two different scenes with a square light source. The correlations among the strata make uniform and isotropic sampling disadvantaged. Among the non-adaptive sampling methods (i.e. excluding R-Uniform J), mirrored jitter sampling results in the least error, although its advantage diminishes as the shadows get more complex.

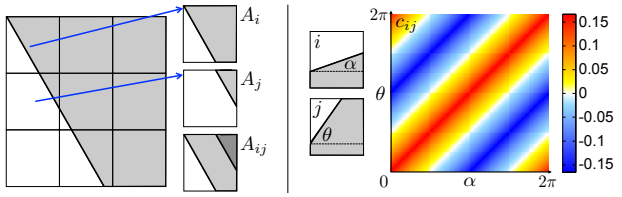

Fig. 11. (Left) When the boundary consists of a single line on the light plane, the intersection area A_{ij} is equal to the smaller of A_i and A_j. (Right) The averaged c_{ij} is plotted as a function of the slopes of the boundary lines in the strata.

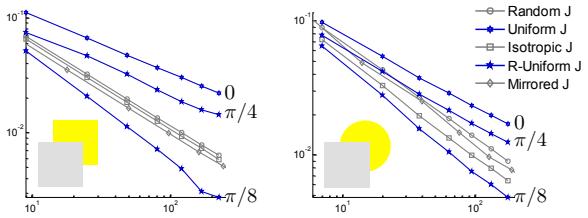

Fig. 12. R-uniform jittering is obtained by rotating the grid pattern of uniform jittering by different angles. R-uniform jittering can give lower errors than possible with other non-adaptive methods, but requires estimating a rotation angle or a range of angles.

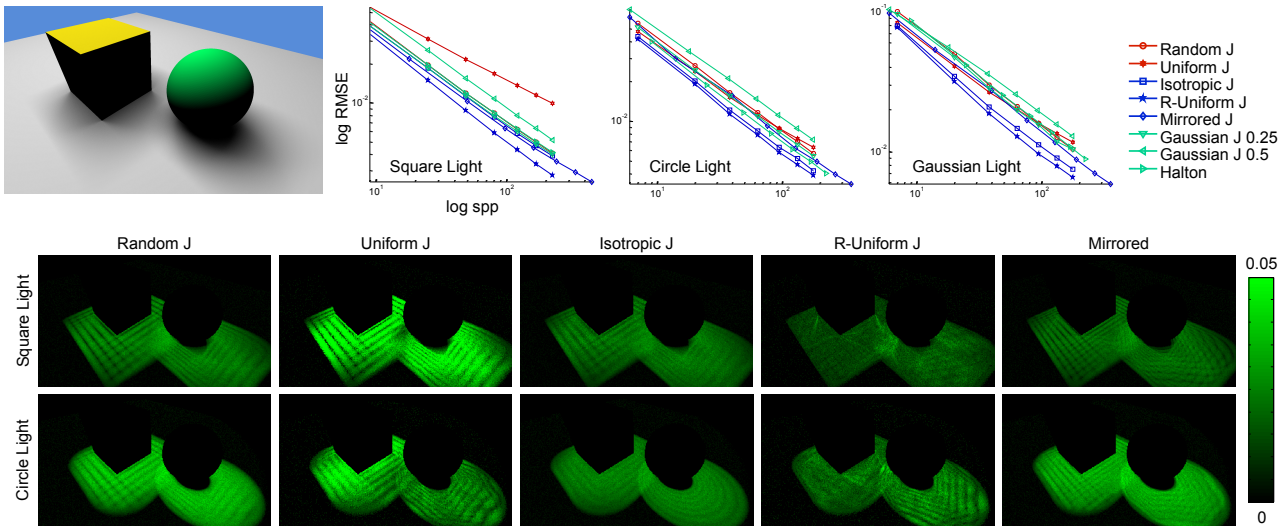

Fig. 13. Error plots (top) and distribution (bottom) for a scene. The error distribution is illustrated for 49 (50 for mirrored jitter sampling) spp for square, and 38 (40 for mirrored jitter sampling) spp for circle and Gaussian lights.

hence the error of uniform jittering, while the opposite is true for disparate slopes. This is further illustrated in Figure 9 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 lines, uniform jittering performs better for the circle light, as previously observed [Ramamoorthi et al. 2012].

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 negative correlations introduced by the changing slopes in the strata. We illustrate this in Figure 12, where different rotation angles are applied, and also in Figures 10, and 13 with the optimal angle under the name r-uniform. Hence, even for square light sources, uniform jittering can be the best alternative if the right angle of rotation can be estimated (please see the supplementary images for more examples with uniform and r-uniform jitter sampling, in Figures 9, 12, 13 uniform jittering refers to the one with rotation for maximum variance, in other figures the rotation is random).

4.2.4 *Isotropic Jitter Sampling.* As elaborated in Section 3.4, isotropic jitter sampling tries to reduce the variance of uniform jitter sampling by introducing a random rotation. By denoting the autocorrelation of the function $f(\mathbf{x},\mathbf{y}) = v(\mathbf{x},\mathbf{y})l(\mathbf{y})$ (with respect to $\mathbf{y}$) with $a_f(\mathbf{x},\mathbf{h})$, and the overall autocorrelation of the scene with $a_f(\mathbf{h}) = \int_R a_f(\mathbf{x},\mathbf{h})d\mathbf{x}$, the total change in the error can be written as:

$$\int_R \left(e^{UNI}(\mathbf{x}) - e^{ISO}(\mathbf{x})\right) d\mathbf{x} = \frac{1}{n^2}\sum_{i\neq j} \frac{a_f(\mathbf{h}_{ij}) - \frac{1}{|\partial V_d|||\mathbf{h}_{ij}||^{d-1}} \int a_f(\mathbf{h})\delta(||\mathbf{h}|| - ||\mathbf{h}_{ij}||)d\mathbf{h}}{a_{\Pi_V}(\mathbf{h}_{ij})}. \quad (24)$$

Thus, 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 9, left) but also for circle lights and long lines of low curvature (Figure 9, circle light - square occluder). For these cases, uniform jitter sampling has higher error, while isotropic sampling performs significantly better.

For 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 (please see Appendix B for an analysis of flat occluders and receivers for other sampling methods). For this case, the integrand has the form $f(\mathbf{x},\mathbf{y}) = v(\tau\mathbf{x} + (1-\tau)\mathbf{y})l(\mathbf{y})$. Setting $\tau' = 1-\tau$ for brevity, the overall autocorrelation $a_f(\mathbf{h})$ becomes

$$\begin{aligned} &\int_R \int v(\tau\mathbf{x} + \tau'\mathbf{y})l(\mathbf{y})v(\tau\mathbf{x} + \tau'(\mathbf{y}-\mathbf{h}))l(\mathbf{y}-\mathbf{h})d\mathbf{y}d\mathbf{x} \\ &= \int l(\mathbf{y})l(\mathbf{y}-\mathbf{h})d\mathbf{y}\frac{1}{\tau}\int_{R'} v(\mathbf{x})v(\mathbf{x}-\tau'\mathbf{h})d\mathbf{x} \\ &= \frac{1}{\tau}a_l(\mathbf{h})a_v(\tau'\mathbf{h}), \end{aligned} \quad (25)$$

where a_l is the autocorrelation of l, and a_v is that of v when the integral is taken over R', a scaled, translated, and symmetrized version of R, assuming a sufficiently large R such that the whole soft shadow can be seen in R'. For a disk occluder and circle light, both a_l and a_v are radially symmetric, and thus we get the same error for both uniform and isotropic jittering, as each term of Equation 24 will be zero. Such a case is illustrated in Figure 9, right, third row (circle light - disk occluder pair).

As explained in Section 3.4 and in the last section, by rotating the uniformly jittered pattern (r-uniform), errors lower than isotropic jittering can be obtained (Figure 12), since the error of isotropic sampling is an average over all rotation angles. Unlike all other jittering methods considered, getting the right angles, however, requires knowledge on the light types and scene geometry. Isotropic

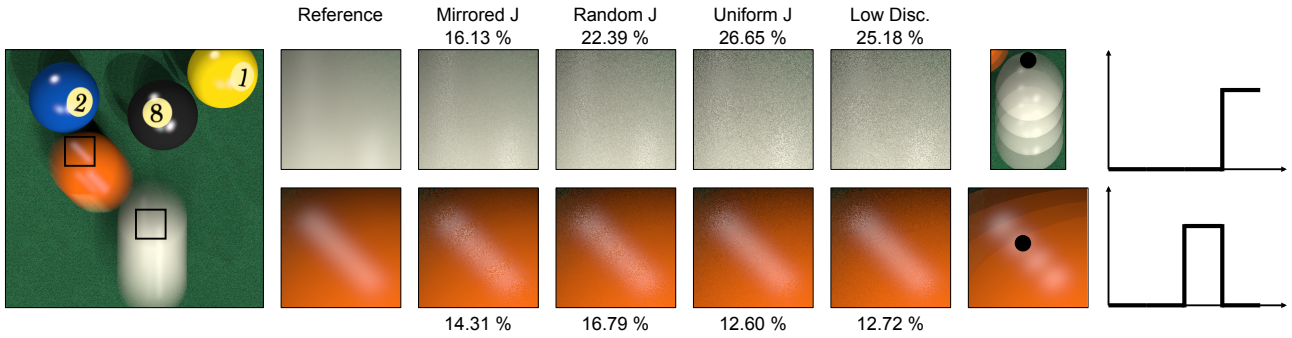

Fig. 14. Motion blur computed by different sampling strategies (4 time samples, Low Disc.: the low discrepancy sampler of PBRT [Pharr and Humphreys 2010]). In the zoomed-in area of the white ball's trajectory, the white ball appears and stays till the end of the shutter interval. This is schematically depicted in the graph in the top-right for the black point (the rightmost images are rendered using a regular sampling for illustrating the trajectory). Due to this single discontinuity, mirrored jitter sampling works the best. On the other hand, the highlight on the orange ball generates two discontinuities, resulting in uniform jitter sampling generating the best result. (The RMS errors are reported for each inset. For the orange ball, they are for the region the highlight is present.)

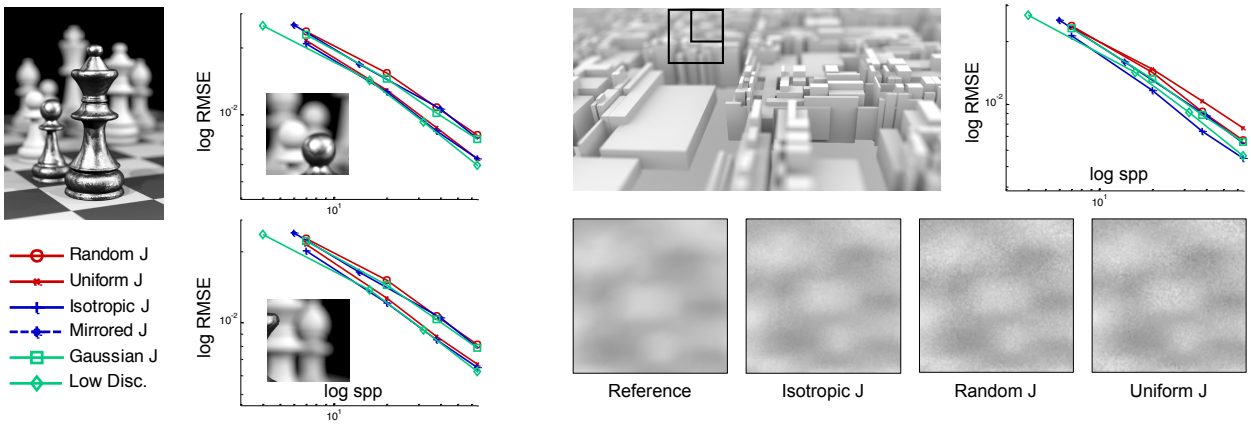

Fig. 15. RMSE is plotted against the samples per pixel (spp) for the chess (left) and microcity (right) scenes. For both cases, isotropic jittering performs consistently well, while uniform jittering has the lowest and highest error for the chess and microcity scene, respectively, due to different correlations for each scene. The long edges in the microcity scene introduce extra correlations, degrading the performance of uniform jittering. (The zoomed-in views are computed using 25 spp. For the microcity scene, the insets are from the area covered by the smaller square, while the RMSE plot is computed for the area in the bigger square. For Gaussian jittering, $\sigma = 0.25/\sqrt{n}$ for n spp.)

jittering is thus an unadaptive alternative to capture such angles by random sampling.

5. MOTION BLUR AND DEPTH OF FIELD

Motion blur and depth-of-field are two important types of effects that require estimation of integrals involving discontinuities. In this section, we provide an analysis on how the estimation accuracy is affected by different sampling strategies. All scenes for this section were rendered using PBRT [Pharr and Humphreys 2010].

Motion blur We start with motion blur, which is based on taking a 1 dimensional integral in time and hence easy to analyze. Note that uniform and isotropic jittering is the same for this case. In Figure 14, we analyze a classical example of motion blur with four time samples. For each zoom-in area, we also show schematically how the intensity coming from a typical point (the black dot in the rightmost images) in that area changes in the graphs on the right. For the point in the trajectory of the white ball, the intensity changes from green (background) to white (ball), and stays the same. Hence, we have a single discontinuity. In this case, mirrored jittered sampling clearly performs better than the others since the discontinuity will generate the intra-stratum negative correlation it needs. On the other hand, for a point on the trajectory of the highlight on the orange ball, there will be two discontinuities: the highlight will appear and disappear in different time instances. Due to the negative correlation between the strata corresponding to the discontinuities, uniform jittering results in the lowest error, although low discrepancy sampling performs almost as well.

Depth-of-field We illustrate our experiments with the depth-of-field effect for two different scenes in Figure 15. The first one is a chess scene that consists of pieces with mostly round shapes, while the second scene depicts a microcity with a lot of long straight edges. As shown in the variance graphs for the chess scene, uniform jittering, isotropic jittering, and the low discrepancy sampler of PBRT [Pharr and Humphreys 2010] perform very similarly, while the RMSEs for the other sampling algorithms are higher. However, for the microcity scene, uniform jittering has the highest error, while isotropic jittering still has the lowest. The correlations due to the presence of long edges in the scene does not affect isotropic jittering as expected.

APPENDIX

A. ERROR OF JITTERED POINT PATTERNS

A.1 Error of Uniform Jitter Sampling and Regular Sampling in the Spectral Domain

We derive in the paper (Appendix A.3) that the variance, and hence mean squared error (MSE) of uniform jittering in the spectral domain is given by

$$e^{UNI} = \sum_{\omega_k \neq 0} |F(\omega_k)|^2, \tag{26}$$

where ω_k's are the points of a (infinite) regular grid of spacing $n^{1/d}$ centered at the origin in the spectral domain, and F is the Fourier transform of f.

MSE of Regular Sampling Assuming that the integration points $\mathbf{x}_i + \mathbf{t}$ lie on a regular grid translated by $\mathbf{t}$ with respect to the origin, the MSE of regular sampling is given by

$$e^{REG} = \left(\int f(\mathbf{x}) d\mathbf{x} - \frac{1}{n} \sum_{i=1}^{n} f(\mathbf{x}_i + \mathbf{t}) \right)^2. \tag{27}$$

The first term is equal to $F(\mathbf{0})$. We would like to write the second term in the frequency domain as well. Note that the translations $\mathbf{t}$ with $||\mathbf{t}||$ smaller than the spacing between the grid points $\mathbf{x}_i$ covers the set of all translations, since the grid is periodic and infinite. Thus, regardless of the translation, there will always be n points in the support V of the function f. We can then equivalently write the second term as a sum over an infinite grid. We denote the set of points of the infinite grid centered at the origin in the spatial domain with $\mathbf{x}_i$, and those in the frequency domain with ω_k. We can then rewrite the sum in the frequency domain as

$$\begin{aligned}
&\frac{1}{n} \sum_{i=1}^{n} f(\mathbf{x}_i + \mathbf{t}) \\
&= \frac{1}{n} \sum_{\mathbf{x}_i} f(\mathbf{x}_i + \mathbf{t}) \\
&= \int f(\mathbf{x}) \frac{1}{n} \sum_{\mathbf{x}_i} \delta(\mathbf{x} - (\mathbf{x}_i + \mathbf{t})) d\mathbf{x} \\
&= \int F(\omega) \left(\sum_{\omega_k} \delta(\omega - \omega_k) e^{-i2\pi\omega^T \mathbf{t}} \right)^* d\omega \\
&= \sum_{\omega_k} F(\omega_k) e^{i2\pi\omega_k^T \mathbf{t}},
\end{aligned} \tag{28}$$

where $(\cdot)^*$ denotes the conjugate. Since f is a real function, $F(-\omega) = F^*(\omega)$. Hence, by summing the terms for the each pair ω_k and $-\omega_k$, we can get

$$\begin{aligned}
&F(\omega_k) e^{i2\pi\omega_k^T \mathbf{t}} + F(-\omega_k) e^{-i2\pi\omega_k^T \mathbf{t}} \\
&= |F(\omega_k)| e^{i(\phi(\omega_k) + 2\pi\omega_k^T \mathbf{t})} + |F(\omega_k)| e^{-i(\phi(\omega_k) + 2\pi\omega_k^T \mathbf{t})} \\
&= 2|F(\omega_k)| cos(\phi(\omega_k) + 2\pi\omega_k^T \mathbf{t}),
\end{aligned} \tag{29}$$

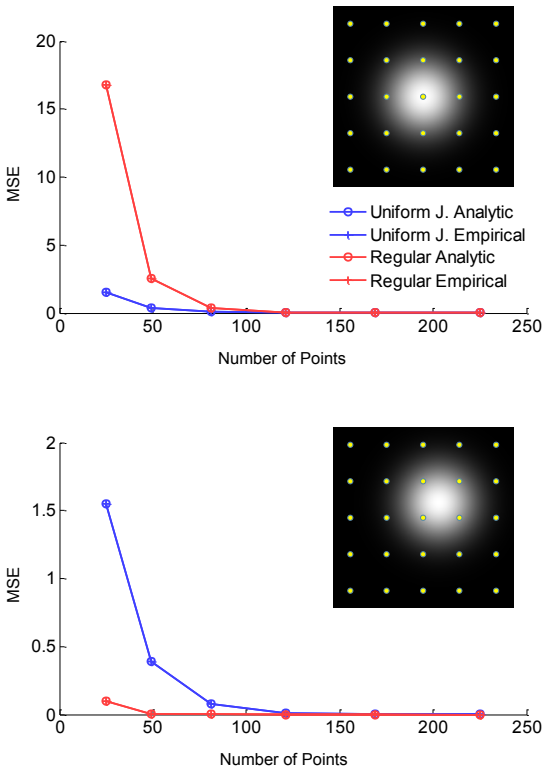

Fig. 16. When the Gaussian is symmetric around one of the points in the grid, uniform jittering results in significantly lower MSE than regular sampling (top). However, regular sampling can slightly (note the scaling of the error axes) decrease the MSE when the mean of the Gaussian is not aligned with a point in the grid (bottom).

where $\phi(\omega)$ is the phase of $F(\omega)$. Thus, we can write the sum in Equation 28 as

$$\begin{aligned}
&\sum_{\omega_k} F(\omega_k) e^{i2\pi\omega_k^T \mathbf{t}} \\
&= \sum_{\omega_k \neq \mathbf{0}} |F(\omega_k)| cos(\phi(\omega_k) + 2\pi\omega_k^T \mathbf{t}) + F(\mathbf{0}).
\end{aligned} \tag{30}$$

Plugging this into Equation 27, we can get the final form of the MSE for regular sampling as

$$e^{REG} = \left(\sum_{\omega_k \neq 0} |F(\omega_k)| \cos(\phi(\omega_k) + 2\pi\omega_k^T \mathbf{t}) \right)^2. \tag{31}$$

Analysis for isotropic Gaussians Let us assume that the integrand is an isotropic Gaussian in two dimensions:

$$\begin{aligned}
f(\mathbf{x}) &= e^{-||\mathbf{x}-\mu||^2/\sigma^2} / (\pi\sigma^2) \\
F(\omega) &= e^{-\pi^2\sigma^2||\omega||^2} e^{-i2\pi\mu^T\omega},
\end{aligned} \tag{32}$$

with Fourier transform F. The MSE formulae for uniform jittering and regular sampling can be derived as:

$$e^{UNI} = \sum_{\omega_k \neq 0} e^{-2\pi^2\sigma^2||\omega_k||^2}. \tag{33}$$

$$e^{REG} = \left(\sum_{\omega_k \neq 0} e^{-\pi^2 \sigma^2 ||\omega_k||^2} \cos(2\pi \omega_k^T (\mathbf{t} - \mu)) \right)^2. \quad (34)$$

From these two expressions, it is easy to see that when $\mu = \mathbf{t}$, that is, when the grid is aligned with the mean of the Gaussian (i.e. the Gaussian is symmetric with respect to the grid), $e^{UNI} < e^{REG}$. However, as $\mu - \mathbf{t}$ changes, regular grid can perform better. We show empirical and theoretical MSE curves for these two different cases in Figure 16. We set $\sigma = 0.05$ and μ is aligned with one of the points in the regular grid on the top, while falls in-between two sampling points on the bottom. We see a significant reduction in MSE with uniform jittering when μ is aligned with a point in the regular grid. The theoretical MSE's agree perfectly with the empirical ones (computed with $50,000$ different realizations of uniform jittering).

A.2 Per-Stratum Error of Mirrored Jitter Sampling

Assuming a binary integrand $f(\mathbf{x})$, we would like to express the error (variance) term for a single stratum as a function of the area of the support of f (i.e. where $f(\mathbf{x}) = 1$) in that stratum. In order to derive the per-stratum error, we first rewrite the variance expression for mirrored jitter sampling in Equation 13 by replacing the number of points n with the number of strata m, and noting that $s_i^2 = s_i$ since f is a binary function:

$$\begin{aligned} var(\hat{I}) = \frac{1}{m^2} \sum_{i=1}^{m} \frac{1}{2} m \int s_i(\mathbf{x}) s_i^-(\mathbf{x}) d\mathbf{x} \\ + \frac{1}{2} m \int s_i(\mathbf{x}) d\mathbf{x} - \left(m \int s_i(\mathbf{x}) d\mathbf{x} \right)^2. \end{aligned} \quad (35)$$

We are interested in the term for a single stratum S_i. Dropping the subscript for simplicity, we would like to express

$$e := \frac{1}{2} m \int s(\mathbf{x}) s^-(\mathbf{x}) d\mathbf{x} + \frac{1}{2} m \int s(\mathbf{x}) d\mathbf{x} - \left(m \int s(\mathbf{x}) d\mathbf{x} \right)^2, \quad (36)$$

as a function of the area of the support of s. Since f is a binary function, the normalized area of the support of s is given by $a := m \int s(\mathbf{x}) d\mathbf{x} \in [0, 1]$. Denoting $b := m \int s(\mathbf{x}) s^-(\mathbf{x}) d\mathbf{x}$, we can thus write $e = (1/2)b + (1/2)a - a^2$. The goal is then to express b in terms of a. Note that b is in fact the intersection of the support of s and that of its mirrored version s^- around the center of the stratum.

We assume that the integrand is sufficiently smooth with respect to the size of the strata such that the boundary of the support of f is locally linear in each stratum. Under this assumption, we illustrate the possible cases in Figure 17, top, where the boundary of the support intersects the boundaries of the stratum at two points. We also show the intersection area, i.e. b, in Figure 17, bottom. For the first two cases there is no intersection, and hence $b = 0$ and $e = (1/2)a - a^2$. For these cases, $a \in [0, 1/2]$. For the last two cases, $a \in [1/2, 1]$, and there is an intersection given by $b = 1 - 2(1 - a) = 2a - 1$, resulting in $e = (3/2)a - (1/2) - a^2$.

B. FLAT OCCLUDERS AND RECEIVERS

For flat occluders and receivers parallel to the light plane of an area light, the visibility function can be written as $v(\mathbf{x}, \mathbf{y}) = v(\tau \mathbf{x} + (1 - \tau)\mathbf{y})$, where $\tau = d_2/d_1$ with d_1 the distance from the receiver to the light plane, and d_2 the distance from the occluder to the light plane [Egan et al. 2011; Ramamoorthi et al. 2012; Mehta et al. 2012]. Hence, the function v just translates over the light plane

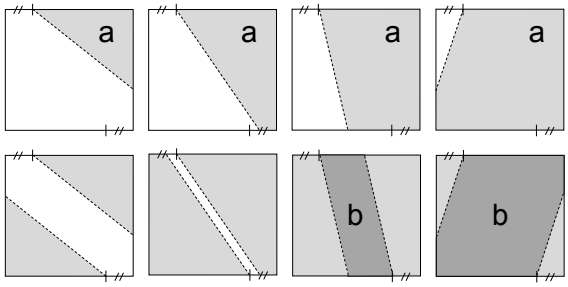

Fig. 17. (Top) There are four different cases where the boundary (dashed lines) of the support of f (the gray shaded regions) can intersect the boundaries of the stratum. (Bottom) For each case, the intersection of the support of s and that of its mirrored version s^- is depicted as the darker shaded region.

as the pixel location $\mathbf{x}$ changes. For this case, we will show that expressions for variance involving the autocorrelation of the visibility function v can be derived. We will then utilize these to analytically compare random, uniform, and mirrored jittering.

For the simplicity of the derivations, without loss of generality, we work with a visibility of the form $v(\mathbf{x}+\mathbf{y})$. This implies that we pre-scale the original visibility function $v((1-\tau)[\tau/(1-\tau)\mathbf{x}+\mathbf{y}])$ by $1/(1-\tau)$, and ignore the constant $\tau/(1-\tau)$ that will appear in front of all the terms below due to the integration with respect to $\mathbf{x}$ on the image plane, as it will not affect the error comparisons. Then, the integrand becomes $f(\mathbf{x}, \mathbf{y}) = v(\mathbf{x} + \mathbf{y}) l(\mathbf{y})$. We would like to compute the total error given by $e = \int_R var(\hat{I}(\mathbf{x})) d\mathbf{x}$ for different sampling patterns.

First, we rewrite the errors (or variances) for different sampling patterns as below:

$$e^{RND}(\mathbf{x}) = \frac{1}{n^2} \sum_{i=1}^{n} n I_{ii}(\mathbf{x}) - [n I_i(\mathbf{x})]^2, \quad (37)$$

$$e^{UNI}(\mathbf{x}) = \frac{1}{n^2} \sum_{i,j=1}^{n} n I_{ij}(\mathbf{x}) - n^2 I_i(\mathbf{x}) I_j(\mathbf{x}), \quad (38)$$

$$e^{MIR}(\mathbf{x}) = \frac{1}{n^2} \sum_{i=1}^{n} \frac{1}{2} n I_{ii^-}(\mathbf{x}) + \frac{1}{2} n I_{ii}(\mathbf{x}) - [n I_i(\mathbf{x})]^2, \quad (39)$$

where $e^{RND}(\mathbf{x})$, $e^{UNI}(\mathbf{x})$, $e^{MIR}(\mathbf{x})$ denote the error for random, uniform, and mirrored jitter sampling, respectively. Note that for mirrored jitter sampling, there are n strata and hence $2n$ points. Here,

$$\begin{aligned} I_i(\mathbf{x}) &= \int s_i(\mathbf{x}, \mathbf{y}) d\mathbf{y} \\ I_{ij}(\mathbf{x}) &= \int s_i(\mathbf{x}, \mathbf{y}) s_j(\mathbf{x}, \mathbf{y}) d\mathbf{y} \\ I_{ii^-}(\mathbf{x}) &= \int s_i(\mathbf{x}, \mathbf{y}) s_i^-(\mathbf{x}, \mathbf{y}) d\mathbf{y}, \end{aligned} \quad (40)$$

with $s_i(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}, \mathbf{y} + \mathbf{o}_i) \Pi_S(\mathbf{y})$ denoting the part of the integrand $f(\mathbf{x}, \mathbf{y})$ in stratum S_i with center $\mathbf{o}_i$ on the light plane (for a given pixel location $\mathbf{x}$), Π_S is 1 inside S and 0 otherwise, S is the square of length $1/\sqrt{n}$ centered around the origin, and $s_i^-(\mathbf{x}, \mathbf{y}) = s_i(\mathbf{x}, -\mathbf{y})$. To simplify the expressions below further, we assume a square light, such that $s_i(\mathbf{x}, \mathbf{y}) = v(\mathbf{o}_i + \mathbf{y} + \mathbf{x}) \Pi_S(\mathbf{y})$, and define a normalized $\Pi(\mathbf{y})$ given by $\Pi(\mathbf{y}) := n \Pi_S(\mathbf{y})$ such that $\int \Pi(\mathbf{y}) d\mathbf{y} = 1$.

We can compute the total error for each sampling pattern by taking the integrals of the terms $nI_{ij}(\mathbf{x})$, $n^2 I_i(\mathbf{x})I_j(\mathbf{x})$, and $nI_{ii^-}(\mathbf{x})$ with respect to $\mathbf{x}$ on the image plane. We start with nI_{ij},

$$\begin{aligned}
&\int_R nI_{ij}(\mathbf{x})d\mathbf{x} \\
&= \int_R \int v(\mathbf{y}+\mathbf{o}_i+\mathbf{x})v(\mathbf{y}+\mathbf{o}_j+\mathbf{x})\Pi(\mathbf{y})d\mathbf{y}d\mathbf{x} \\
&= \int \int_R v(\mathbf{y}+\mathbf{o}_i+\mathbf{x})v(\mathbf{y}+\mathbf{o}_j+\mathbf{x})d\mathbf{x}\Pi(\mathbf{y})d\mathbf{y} \\
&= \int \int_{R+\mathbf{y}+\mathbf{o}_i} v(\mathbf{x})v(\mathbf{x}-\mathbf{h}_{ij})d\mathbf{x}\Pi(\mathbf{y})d\mathbf{y} \\
&= \int \Pi(\mathbf{y})d\mathbf{y} \int_R v(\mathbf{x})v(\mathbf{x}-\mathbf{h}_{ij})d\mathbf{x} \\
&= a_v(\mathbf{h}_{ij}).
\end{aligned} \tag{41}$$

Here, $\mathbf{h}_{ij} = \mathbf{o}_i - \mathbf{o}_j$, the autocorrelation a_v is computed over R, $R+\mathbf{y}+\mathbf{o}_i$ denotes translation of the region R by the vector $\mathbf{y}+\mathbf{o}_i$, and we assume that the region R in the pixel space is large enough such that the translation $\mathbf{y}+\mathbf{o}_i$ to R does not change the result of the integral when computing the autocorrelation of the visibility function over R. In practice, this is approximately satisfied for most scenes, except when significant parts of the soft shadows are not rendered. Similarly, for $n^2 I_i I_j$,

$$\begin{aligned}
&\int_R n^2 I_i(\mathbf{x})I_j(\mathbf{x})d\mathbf{x} \\
&= \int_R \int \int v(\mathbf{y}+\mathbf{o}_i+\mathbf{x})v(\mathbf{y}'+\mathbf{o}_j+\mathbf{x})\Pi(\mathbf{y})\Pi(\mathbf{y}')d\mathbf{y}d\mathbf{y}'d\mathbf{x} \\
&= \int \int \Pi(\mathbf{y})\Pi(\mathbf{y}') \int_R v(\mathbf{y}+\mathbf{o}_i+\mathbf{x})v(\mathbf{y}'+\mathbf{o}_j+\mathbf{x})d\mathbf{x}d\mathbf{y}d\mathbf{y}' \\
&= \int \int \Pi(\mathbf{y})\Pi(\mathbf{y}')a_v(\mathbf{y}-\mathbf{y}'+\mathbf{h}_{ij}))d\mathbf{y}d\mathbf{y}' \\
&= [\Pi * \Pi * a_v](\mathbf{h}_{ij}).
\end{aligned} \tag{42}$$

Finally, we can calculate the integral of nI_{ii^-} as,

$$\begin{aligned}
&\int_R nI_{ii^-}(\mathbf{x})d\mathbf{x} \\
&= \int_R \int v(\mathbf{y}+\mathbf{o}_i+\mathbf{x})v(-\mathbf{y}+\mathbf{o}_i+\mathbf{x})\Pi(\mathbf{y})d\mathbf{y}d\mathbf{x} \\
&= \int \Pi(\mathbf{y}) \int_R v(\mathbf{y}+\mathbf{o}_i+\mathbf{x})v(-\mathbf{y}+\mathbf{o}_i+\mathbf{x})d\mathbf{x}d\mathbf{y} \\
&= \int \Pi(\mathbf{y})a_v(2\mathbf{y})d\mathbf{y} \\
&= [\Pi_2 * a_v](\mathbf{0}),
\end{aligned} \tag{43}$$

where we defined $\Pi_2(\mathbf{y}) := 1/2\Pi(\mathbf{y}/2)$, such that $\int \Pi_2(\mathbf{y})d\mathbf{y} = 1$. To get the final expressions for variances, we can sum these terms as follows:

$$e^{RND} = \frac{1}{n}\left(a_v(\mathbf{0}) - [\Pi * \Pi * a_v](\mathbf{0})\right) \tag{44}$$

$$e^{UNI} = \frac{1}{n^2}\sum_{i,j=1}^{n} a_v(\mathbf{h}_{ij}) - [\Pi * \Pi * a_v](\mathbf{h}_{ij}) \tag{45}$$

$$e^{MIR} = \frac{1}{2n}\left([\Pi_2 * a_v](\mathbf{0}) + a_v(\mathbf{0}) - 2[\Pi * \Pi * a_v](\mathbf{0})\right) \tag{46}$$

We see that the error difference depends on the decay of $a_v(\mathbf{y})$ at $\mathbf{y} = \mathbf{0}$ for random and mirrored jittering, and at $\mathbf{y} = \mathbf{h}_{ij}$ for uniform jittering, since the convolutions take local averages of a_v.

C. DERIVATION OF VARIANCE FOR JITTERED IMPORTANCE SAMPLING AND RANDOM JITTERING

In Equation 11, we show that variance of jittered importance sampling, where the weight function is inversely proportional to density $w(\mathbf{x}) = 1/\lambda(\mathbf{x})$ in uncorrelated jittering, is given as follows:

$$var(\hat{I}) = \sum_{i=1}^{n} \int_{S_i} f^2(\mathbf{x})/k_i(\mathbf{x})d\mathbf{x} - \left(\int_{S_i} f(\mathbf{x})d\mathbf{x}\right)^2. \tag{47}$$

However, the expression in Figure 2 of the work by Subr et al. [2014], and correspondingly the derivation in Equation 10 of the supplementary material for that work differs from this expression. Here, we show that there is a normalization difference in that expression and actually proceeding as in the work of Subr et al. [2014] also leads to the same expression as in Equation 47.

Following the notation in the mentioned work [Subr et al. 2014] and the associated supplementary material, we define $g(\mathbf{x})$ to be the probability density function used for importance sampling. This function is defined in a domain $\mathcal{D}$, which is partitioned into disjoint strata $\mathcal{D}_i$ of equal volume. Then the proposed estimator is defined as $\hat{I} = \sum_{i=1}^{n} \beta_i f(\mathbf{x}_i)/g(\mathbf{x}_i)$ with $\beta_i = \int_{D_i} g(\mathbf{x})d\mathbf{x}$ and $\mathbf{x}_i$ is the point in $\mathcal{D}_i$ with probability density proportional to a function $g(\mathbf{x})$.

Hence, in each stratum $\mathcal{D}_i$, the point $\mathbf{x}_i$ is distributed according to the probability density $g(\mathbf{x})/\beta_i = g(\mathbf{x})/\int_{D_i} g(\mathbf{x})d\mathbf{x}$ and the resulting estimator $f(\mathbf{x}_i)/[g(\mathbf{x}_i)/\int_{D_i} g(\mathbf{x})d\mathbf{x}]$ based on importance sampling is computed. Since the point in each stratum is picked independently, the expected value of the resulting estimator is simply $\mathbb{E}\hat{I} = \sum_{i=1}^{n} \mathbb{E}\beta_i f(\mathbf{x}_i)/g(\mathbf{x}_i) = \sum_{i=1}^{n} \int_{\mathcal{D}_i} f(\mathbf{x})d\mathbf{x} = \int_{\mathcal{D}} f(\mathbf{x})d\mathbf{x} = I$. Defining $I_i := \int_{\mathcal{D}_i} f(\mathbf{x})d\mathbf{x}$, the variance can be similarly computed as follows:

$$\begin{aligned}
var(\hat{I}) &= var\left(\sum_{i=1}^{n} \beta_i f(\mathbf{x}_i)/g(\mathbf{x}_i)\right) \\
&= \sum_{i=1}^{n} var(\beta_i f(\mathbf{x}_i)/g(\mathbf{x}_i)) \\
&= \sum_{i=1}^{n} \int_{\mathcal{D}_i} \beta_i f^2(\mathbf{x})/g(\mathbf{x})d\mathbf{x} - I_i^2.
\end{aligned} \tag{48}$$

By substituting our notation for the probability distribution in each stratum $k_i(\mathbf{x}) = g(\mathbf{x})/\beta_i$, and S_i for $\mathcal{D}_i$, we see that this expression is identical to our expression in Equation 47.

In Equation 10 of the supplementary material of the work by Subr et al. [2014], the $g(\mathbf{x})$, which is unnormalized when restricted to $\mathcal{D}_i$, is used instead of the normalized $g(\mathbf{x})/\beta_i$ as the probability density function for the stratum $\mathcal{D}_i$, resulting in extra normalization terms. For standard random jitter sampling, the following are set $\beta_i = 1/n$, $g(\mathbf{x}) = 1$, resulting in the expression in Equation 12.

D. FROM SPECTRAL TO SPATIAL VARIANCE FORMULA FOR STATIONARY PROCESSES

Pilleboue et al. [2015] have recently derived the variance formula for stationary (called as homogenous in the mentioned paper) processes with a constant number of points residing in a toroidal do-

main. The formula states the variance in terms of the Fourier spectrums of the sampling pattern and the integrand. In this section, we show that the derived formula in Pilleboue et al. [2015] leads to the same form as we state in Equation 5 under the assumptions in that work [Pilleboue et al. 2015].

We start from Pilleboue et al. [2015]'s integral estimator and the resulting variance formula, which was derived under the assumptions that the number of points n is constant and hence $\lambda = n$, the estimator is with $w = 1/\lambda$, and the points live in a d-dimensional toroidal domain $\mathcal{T}^d$. Then, denoting the power spectrums of the sampling pattern and integrand as $P_s = |\mathscr{F}(\sum_{i=1}^{n} \delta(\mathbf{x} - \mathbf{x}_i))|^2$ and $P_f = |\mathscr{F}(f)|^2$, the spectral formula as given in Equation E-10 of the mentioned work [Pilleboue et al. 2015] is

$$var(\hat{I}) = \frac{1}{n} \int_{\omega \neq \mathbf{0}} \frac{1}{n} \mathbb{E} P_s(\omega) P_f(\omega) d\omega. \tag{49}$$

The $P_s(\omega)$ can also be written as $P_s(\omega) = \sum_{i=1}^{n} \sum_{j=1}^{n} e^{-i2\pi\omega^T(\mathbf{x}_i - \mathbf{x}_j)}$. Hence, for $\omega = \mathbf{0}$, $\mathbb{E}P_s(\mathbf{0}) = \mathbb{E}\sum_{i=1}^{n}\sum_{j=1}^{n} 1 = n^2$. This means that the integrand in Equation 49 at $\omega = \mathbf{0}$ is $P_f(\mathbf{0})$. Since $P_f = \mathscr{F}(a_f)$, we get $P_f(\mathbf{0}) = \int a_f(\mathbf{x}) d\mathbf{x} = \left(\int f(\mathbf{x}) d\mathbf{x}\right)^2 = I^2$. Hence, Equation 49 can also be written as

$$var(\hat{I}) = \frac{1}{n^2} \int \mathbb{E} P_s(\omega) P_f(\omega) d\omega - I^2. \tag{50}$$

The expected value of the power spectrum of the sampling pattern can be derived as follows

$$\begin{aligned}
\mathbb{E}P_s(\omega) &= \mathbb{E} \sum_{i=1}^{n} \sum_{j=1}^{n} e^{-i2\pi\omega^T(\mathbf{x}_i - \mathbf{x}_j)} \\
&= \mathbb{E} \sum_{i \neq j} e^{-i2\pi\omega^T(\mathbf{x}_i - \mathbf{x}_j)} + \mathbb{E} \sum_{i=1}^{n} 1 \\
&= \int_{\mathcal{T}^d \times \mathcal{T}^d} e^{-i2\pi\omega^T(\mathbf{x}-\mathbf{y})} n^2 g(\mathbf{x} - \mathbf{y}) d\mathbf{x} d\mathbf{y} + n \\
&= n^2 \int e^{-i2\pi\omega^T \mathbf{h}} g(\mathbf{h}) d\mathbf{h} + n \\
&= n^2 G(\omega) + n,
\end{aligned} \tag{51}$$

where $G = \mathscr{F}(g)$, and the fourth equality results from integrating over the unit toroidal domain $\mathcal{T}^d$. The variance in Equation 50 can thus be written in the following form

$$\begin{aligned}
var(\hat{I}) &= \frac{1}{n^2} \int (n^2 G(\omega) + n) P_f(\omega) d\omega - I^2 \\
&= \int G(\omega) P_f(\omega) d\omega + \frac{1}{n} \int P_f(\omega) d\omega - I^2 \\
&= \int g(\mathbf{h}) a_f(\mathbf{h}) d\mathbf{h} + \frac{1}{n} a_f(\mathbf{0}) - I^2 \\
&= \int g(\mathbf{h}) a_f(\mathbf{h}) d\mathbf{h} + \frac{1}{n} \int f^2(\mathbf{x}) d\mathbf{x} - I^2,
\end{aligned} \tag{52}$$

which is the same expression as in Equation 5 with the stated assumptions.

REFERENCES

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.

CHEN, J., GE, X., WEI, L.-Y., WANG, B., WANG, Y., WANG, H., FEI, Y., QIAN, K.-L., YONG, J.-H., AND WANG, W. 2013. Bilateral blue noise sampling. *ACM Trans. Graph. 32,* 6 (Nov.), 216:1–216:11.

EGAN, K., HECHT, F., DURAND, F., AND RAMAMOORTHI, R. 2011. Frequency analysis and sheared filtering for shadow light fields of complex occluders. *ACM Trans. Graph. 30,* 2 (Apr.), 9:1–9:13.

FATTAL, R. 2011. Blue-noise point sampling using kernel density model. *ACM Trans. Graph. 30,* 4 (July), 48:1–48:12.

HACHISUKA, T., JAROSZ, W., WEISTROFFER, R. P., DALE, K., HUMPHREYS, G., ZWICKER, M., AND JENSEN, H. W. 2008. Multidimensional adaptive sampling and reconstruction for ray tracing. *ACM Trans. Graph. 27,* 3 (Aug.), 33:1–33:10.

HAHN, U., JENSEN, E. B. V., VAN LIESHOUT, M.-C., AND NIELSEN, L. S. 2003. Inhomogeneous spatial point processes by location-dependent scaling. *Adv. in Appl. Probab. 35,* 2, 295–550.

ILLIAN, J., PENTTINEN, A., STOYAN, H., AND STOYAN, D., Eds. 2008. *Statistical Analysis and Modelling of Spatial Point Patterns*. John Wiley and Sons, Ltd.

LI, H., WEI, L.-Y., SANDER, P. V., AND FU, C.-W. 2010. Anisotropic blue noise sampling. *ACM Trans. Graph. 29,* 6 (Dec.), 167:1–167:12.

MEHTA, S. U., WANG, B., AND RAMAMOORTHI, R. 2012. Axis-aligned filtering for interactive sampled soft shadows. *ACM Trans. Graph. 31,* 6 (Nov.), 163:1–163:10.

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.

OVERBECK, R. S., DONNER, C., AND RAMAMOORTHI, R. 2009. Adaptive wavelet rendering. *ACM Trans. Graph. 28,* 5 (Dec.), 140:1–140:12.

OWEN, A. B. 2013. *Monte Carlo Theory, Methods and Examples*. To be published.

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.

RAMAMOORTHI, 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.

SCHMALTZ, C., GWOSDEK, P., BRUHN, A., AND WEICKERT, J. 2010. Electrostatic halftoning. *Comput. Graph. Forum 29,* 8, 2313–2327.

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. 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