Convergence Analysis for Anisotropic Monte Carlo Sampling Spectra: Supplementary Material

GURPRIT SINGH and WOJCIECH JAROSZ, Dartmouth College

ACM Reference format:
Gurprit Singh and Wojciech Jarosz. 2017. Convergence Analysis for Anisotropic Monte Carlo Sampling Spectra: Supplementary Material. *ACM Trans. Graph.* 36, 4, Article 137 (July 2017), 22 pages.
DOI: 10.1145/3072959.3073656

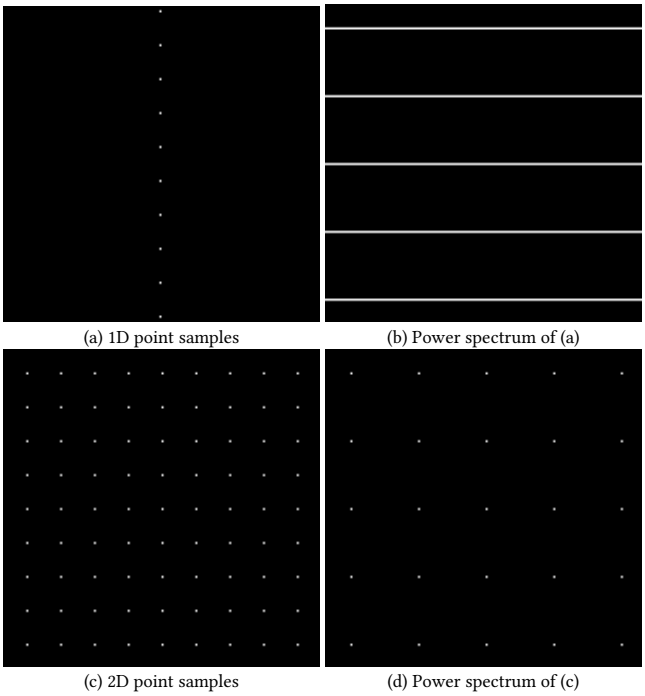
1 SAMPLING POWER SPECTRA

Monomial approximation of radial power spectra. In this section, we analyze the convergence tool proposed by Pilleboue et al. [2015] and describe the notion behind confining the low-frequency region of the expected power spectra to $\alpha \sqrt[d]{N}$ frequency. We start by analyzing the *regular* sampling power spectrum. As shown in Fig. 1, for N *regular* point samples in 1D the corresponding power spectrum in Fig. 1(b) contains replicas of the *DC line*, which is at the center of Fig. 1(b). These replicas (aliases) are N (frequency) distance away from the DC peak along the vertical axis (i.e. at $v_y = N$). Now, if we consider a 2D regular grid samples, the corresponding expected power spectrum contains only high energy peaks (no lines), and the distribution of peaks is such that the horizontal and the vertical location of the first peak (replica) from the DC component is exactly at frequency $\sqrt{N}$. This idea can be generalized to higher dimensions where the first peak is present at a frequency $\sqrt[d]{N}$ along the canonical axes. This explains the fact why the regular sampling power spectrum contains less and less peaks in the low frequency region around the DC component as we keep increasing the number of point samples.

This notion can be generalized to jitter and other stochastic (blue noise) samplers, where the first peak in the jitter and/or blue noise samplers can be obtained at a distance $\alpha \sqrt[d]{N}$ from the DC component, where $\alpha > 0$. Note that, the α variable can be used to align all different radial sampling power spectra such that in the low-frequency region they all can be categorized, in an unbiased fashion, according to their radial behavior.

Now, for our analysis, we normalize the frequency components (v_i) of the d-dimensional power spectrum with $\sqrt[d]{N}$, which allows us to write the $\mathcal{P}_S(v)$ in the normalized frequency form $\mathcal{P}_S\left(\frac{v}{\alpha \sqrt[d]{N}}\right)$. As a result, we can represent the low-frequency region of the power spectrum for different samplers in a unit region as illustrated in Fig. 2, irrespective of the number of point samples N used to compute the power spectrum. In general, higher the number of samples used to compute the sampling power spectrum, the more accurate this representation would get for a given power spectrum.

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



(a) 1D point samples (b) Power spectrum of (a)
(c) 2D point samples (d) Power spectrum of (c)

Fig. 1. Regular grid sampling pattern (first column) is shown with the corresponding expected power spectra (second column) for both 1D (top row) and 2D (bottom row) dimensions. In 1D, the first replica in the expected power spectrum of regular samples is located at a frequency N from the DC component, where N is the number of samples used to generate the spectrum. In 2D, the first replica (peak) is located at a frequency $\sqrt{N}$ from the DC as shown in (d).

Monomial radial profile. To obtain the generalized monomial *profile* for point samples, we first consider the case of jittered samples. The *unnormalized* power spectrum for jittered sampling power spectrum can be written as:

$$\langle \mathcal{P}_S(v) \rangle = \begin{cases} N^2 & \text{DC component} \\ N\left(1 - \prod_i^d \text{Sinc}(\pi v_i)^2\right) & \text{otherwise} \end{cases}, \quad (1)$$

whereas, the fully normalized power spectrum can be written as:

$$\langle \mathcal{P}_S(v) \rangle = \begin{cases} 1 & \text{DC component} \\ \frac{1}{N}\left(1 - \prod_i^d \text{Sinc}(\pi v_i)^2\right) & \text{otherwise} \end{cases}. \quad (2)$$

Note that, if we consider samplers that generate the same number of N samples at each realization, the corresponding fully normalized sampling power spectrum would have its DC component value

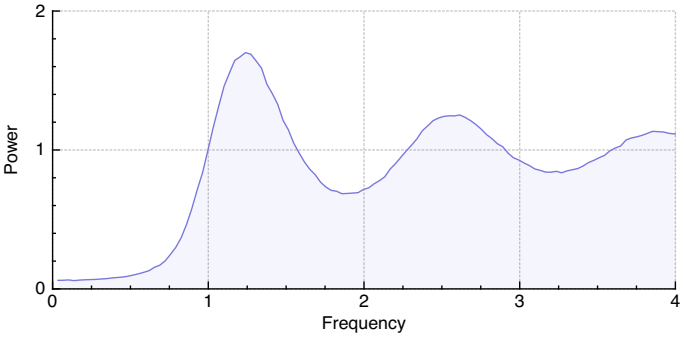


Fig. 2. The radial curve is scaled using α parameter so that the low-frequency region could be well confined within the unit radial power axes. Note that, once the α value is adjusted, for any number of samples, the corresponding radial power spectrum would not scale with N, given that we normalize the radial frequency variable ρ by $\alpha \sqrt[d]{N}$.

equal to 1, as shown in Eq. (2). Jittered sampler belongs to this class of samplers and, therefore, the variance due to jittered samples does not depend on the DC component. Therefore, we are only interested in the expected power spectrum at frequencies other than DC, which for jittered samples in d-dimensions is provided in Eq. (2) as follows:

$$\langle \mathcal{P}_S(\nu) \rangle = \frac{1}{N}\left(1 - \prod_i^d \mathrm{Sinc}(\pi \nu_i)^2\right) \tag{3}$$

where ν_i is the i-th dimension of frequency vector ν. Pilleboue et al. [2015] proposed a theoretical tool that allows to write the $\left(1 - \prod_i^d \mathrm{Sinc}(\pi\nu_i)^2\right)$ term from Eq. (3), in the following radially averaged form:

$$= \begin{cases} \dfrac{\gamma \rho^b}{\alpha^b \sqrt[d]{N}^b} & \nu < \alpha \sqrt[d]{N} \\ \gamma & \text{otherwise} \end{cases} . \tag{4}$$

where $\rho > 0$ is the radial frequency and $b = 2$ for jitter sampling pattern [Gabrielli and Torquato 2004; Torquato et al. 2006]. Substituting Eq. (4) for $\left(1 - \prod_i^d \mathrm{Sinc}(\pi\nu_i)^2\right)$ in Eq. (3), we get the generalized expected power spectrum form, which in radial frequency terms can be written as follows:

$$\left\langle \check{\mathcal{P}}_S(\rho) \right\rangle = \begin{cases} \dfrac{\gamma \rho^b}{\alpha^b \sqrt[d]{N}^b N} & \nu < \alpha \sqrt[d]{N} \\ \dfrac{\gamma}{N} & \text{otherwise} \end{cases} , \tag{5}$$

where $\check{\mathcal{P}}_S(\cdot)$ represents the radial power spectrum. Note that, the extra $\frac{1}{N}$ factor in Eq. (5) is due to the factor $1/N$ in Eq. (3). This generalized polynomial form of the expected sampling power spectrum is the same as the one proposed by Pilleboue et al. [2015], with the only exception that in this case (Eq. (5)) we are using a fully normalized power spectrum that gives an extra $\frac{1}{N}$ factor in this form.

2 VARIANCE FORMULATION IN DISCRETIZED FORM

Following the variance formulation proposed by Pilleboue et al. [2015]:

$$Var(I_N) = \int_0^\infty \rho^{d-1} \int_{\mathcal{S}^{d-1}} \langle \mathcal{P}_S(\rho\mathbf{n}) \rangle \, \mathcal{P}_f(\rho\mathbf{n}) \, \mathrm{d}\mathbf{n} \, \mathrm{d}\rho \,, \tag{6}$$

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

If we swap the order of integration as shown below:

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

the inner integral correspond to the integration over all the radial frequencies for a given direction $\mathbf{n}$, whereas, the outer integral corresponds to integration over all the directions. Lets denote the inner integral as a function of $\mathbf{n}$:

$$\mathcal{R}_0^\infty(\mathbf{n}) = \int_0^\infty \rho^{d-1} \langle \mathcal{P}_S(\rho\mathbf{n}) \rangle \, \mathcal{P}_f(\rho\mathbf{n}) \mathrm{d}\rho \tag{8}$$

By using the linearity principal of definite integrals, we can rewrite the outer integral in the variance formulation in the following form:

$$\begin{aligned} Var(I_N) = &\int_{\mathbf{0}}^{\mathbf{n}_1} \mathcal{R}_0^\infty(\mathbf{n}_1) \, \mathrm{d}\mathbf{n} + \int_{\mathbf{n}_1}^{\mathbf{n}_2} \mathcal{R}_0^\infty(\mathbf{n}_2) \, \mathrm{d}\mathbf{n} + \ldots \\ &+ \int_{\mathbf{n}_{m-1}}^{\mathbf{n}_m} \mathcal{R}_0^\infty(\mathbf{n}_m) \, \mathrm{d}\mathbf{n} \,, \end{aligned} \tag{9}$$

where we can assume $\mathbf{n}_m \equiv \mathbf{0}$ to ensure that the integral is over a closed hypersphere (In 2D, $\mathbf{n}_m = 360^\circ \equiv 0^\circ$). For $m \to \infty$, the domain of integration for each integral becomes infinitesimally small. In the context of true infinitesimal calculus [Keisler 2012], we can rewrite the above integral in the following form:

$$\begin{aligned} Var(I_N) = &\mathcal{R}_0^\infty(\mathbf{n}_1) \, \Delta\mathbf{n}_1 + \mathcal{R}_0^\infty(\mathbf{n}_2) \, \Delta\mathbf{n}_2 + \ldots \\ &+ \mathcal{R}_0^\infty(\mathbf{n}_m) \, \Delta\mathbf{n}_m \,, \end{aligned} \tag{10}$$

with $m \to \infty$, where the $\Delta\mathbf{n}_k$s doesn't need to be of the same size. This allows us to rewrite the variance formulation in the following form:

$$Var(I_N) = \lim_{m\to\infty} \sum_{k=1}^{m} \mathcal{R}_0^\infty(\mathbf{n}) \Delta\mathbf{n}_m \tag{11}$$

$$Var(I_N) = \lim_{m\to\infty} \sum_{k=1}^{m} \int_0^\infty \rho^{d-1} \langle \mathcal{P}_S(\rho\mathbf{n}) \rangle \, \mathcal{P}_f(\rho\mathbf{n}) \mathrm{d}\rho \Delta\mathbf{n}_m \tag{12}$$

which is the form proposed in Eq. (9) in the main paper.

3 SUM OF BIG $O(\cdot)$ NOTATIONS

For completeness sake, we show a proof of the following statement (more details [Cormen et al. 2001]):

$$O(p(N)) + O(q(N)) = O(max(p(N), q(N))) \tag{13}$$

where,

$$O(p(N)) = C_1 p(N) \;\; \forall \;\; N \geq M_1 \,, \tag{14}$$

$$O(q(N)) = C_2 q(N) \;\; \forall \;\; N \geq M_2 \,. \tag{15}$$

where, $C_1 > 0, C_2 > 0$ and $M_1 \in \mathbf{Z}^+/0$ Now, consider $M = M_1 + M_2$ and $C = max(C_1, C_2)$. By adding Eq. (14) and (15), we get:

$$\begin{aligned} O(p(N)) + O(q(N)) &= C_1 p(N) + C_2 q(N) && (16)\\ &\leq C(p(N) + q(N)) && (17)\\ &\leq 2C\max(p(N), q(N)) && (18)\end{aligned}$$

which implies:

$$O(p(N)) + O(q(N)) = O(\max(p(N), q(N))) \quad (19)$$

□

4 DETAILED PROOF: CONVERGENCE FROM SINGLE DIRECTION

To show that the overall convergence rate of a sampling pattern can be obtained from a particular direction of its sampling power spectrum, we start by splitting the variance integral from *Eq. 9 from the main paper* in-parts as follows:

$$\begin{aligned} Var(I_N) < &\int_0^\infty \rho^{d-1} \langle \mathcal{P}_S(\rho \mathbf{n}_1)\rangle \mathcal{P}_f(\rho \mathbf{n}_1) \mathrm{d}\rho \Delta \mathbf{n}_1 \\ &+ \sum_{k=2}^{m} \left[\int_0^\infty \rho^{d-1} \langle \mathcal{P}_S(\rho \mathbf{n}_k)\rangle \mathcal{P}_f(\rho \mathbf{n}_k) \mathrm{d}\rho \right] \Delta \mathbf{n}_k, \end{aligned} \quad (20)$$

where, we have partitioned the summation by writing out one of the direction from the rest of the m directions. Following Eq. (11) from the main paper, we divide the integral over the rest of the $m-1$ directions in the low and high frequency regions while keeping the first direction as it is:

$$\begin{aligned} Var(I_N) < &\int_0^\infty \rho^{d-1} \langle \mathcal{P}_S(\rho \mathbf{n}_1)\rangle \mathcal{P}_f(\rho \mathbf{n}_1) \mathrm{d}\rho \Delta \mathbf{n}_1 \\ &+ \sum_{k=2}^{m} \left[\int_0^{\rho_k} \rho^{d-1} \langle \mathcal{P}_S(\rho \mathbf{n}_k)\rangle \mathcal{P}_f(\rho \mathbf{n}_k) \mathrm{d}\rho \right. \\ &\left. + \int_{\rho_k}^\infty \rho^{d-1} \langle \mathcal{P}_S(\rho \mathbf{n}_k)\rangle \mathcal{P}_f(\rho \mathbf{n}_k) \mathrm{d}\rho \right] \Delta \mathbf{n}_k. \end{aligned} \quad (21)$$

Lets assume that the sampling power spectrum has a constant energy over the whole radial frequencies in the differential cone $\Delta \mathbf{n}_1$. As a result, $\mathbf{n}_1$ direction can be approximated by a constant radial profile ($b_k = 0$), whereas, for all other directions we assume to have a non-zero monomial behavior ($b_k > 0$) in the low frequency region. By plugging in the constant profile ($\mathcal{P}_S(\nu) = \gamma/N$), along direction $\mathbf{n}_1$ we can rewrite the whole integral as:

$$\begin{aligned} Var(I_N) < &\int_0^\infty \rho^{d-1} \frac{\gamma}{N} \mathcal{P}_f(\rho \mathbf{n}_1) \mathrm{d}\rho \Delta \mathbf{n}_1 \\ &+ \sum_{k=2}^{\text{directions}} \left[\int_0^{\rho_k} \rho^{d-1} \langle \mathcal{P}_S(\rho \mathbf{n}_k)\rangle \mathcal{P}_f(\rho \mathbf{n}_k) \mathrm{d}\rho \right. \\ &\left. + \int_{\rho_k}^\infty \rho^{d-1} \langle \mathcal{P}_S(\rho \mathbf{n}_k)\rangle \mathcal{P}_f(\rho \mathbf{n}_k) \mathrm{d}\rho \right] \Delta \mathbf{n}_k. \end{aligned} \quad (22)$$

Here, the first integral in $\mathbf{n}_1$ direction doesn't depend on N. Therefore, irrespective of what the underlying integrand f is, the integral can be approximated by $O\left(N^{-1}\right)$ resulting in:

$$\begin{aligned} Var(I_N) < O\left(\frac{1}{N}\right) + \sum_{k=2}^{m} \Bigg[&\int_0^{\rho_k} \rho^{d-1} \langle \mathcal{P}_S(\rho \mathbf{n}_k)\rangle \mathcal{P}_f(\rho \mathbf{n}_k) \mathrm{d}\rho \\ &+ \int_{\rho_k}^\infty \rho^{d-1} \langle \mathcal{P}_S(\rho \mathbf{n}_k)\rangle \mathcal{P}_f(\rho \mathbf{n}_k) \mathrm{d}\rho \Bigg] \Delta \mathbf{n}_k. \end{aligned} \quad (23)$$

To solve the integrals over the other m directions, we consider a best and a worst case (from Eq. 10 from the main paper) for a given class of functions [Brandolini et al. 2001, 2003] and use it directly for $\mathcal{P}_f$. Similarly, for the sampling power spectrum $\mathcal{P}_S$, we insert the monomial radial profile of degree $b_k > 0$ in the above equation (Eq. 12 from the main paper). The resulting variance integral for the worst case can be rewritten in the following form:

$$\begin{aligned} Var(I_N) < O\left(\frac{1}{N}\right) + \sum_{k=2}^{m} \Bigg[&\int_0^{\rho_0} \rho^{d-1} \frac{\gamma_k}{N} \left(\frac{\rho}{\alpha_k N_k}\right)^{b_k} c_f \mathrm{d}\rho \\ &+ \int_{\rho_0}^{\rho_k} \rho^{d-1} \frac{\gamma_k}{N} \left(\frac{\rho}{\alpha_k N_k}\right)^{b_k} c_f \rho^{-d-1} \mathrm{d}\rho \\ &+ \int_{\rho_k}^\infty \rho^{d-1} \frac{\gamma_k}{N} c_f \rho^{-d-1} \mathrm{d}\rho \Bigg] \Delta \mathbf{n}_k \quad (24) \\ = O\left(\frac{1}{N}\right) + \frac{\gamma_k\, c_f}{N} \sum_{k=2}^{m} \Bigg[&\int_0^{\rho_0} \frac{\rho^{b_k+d-1}}{(\alpha_k N_k)^{b_k}} \mathrm{d}\rho \\ &+ \int_{\rho_0}^{\rho_k} \frac{\rho^{b_k-2}}{(\alpha_k N_k)^{b_k}} \mathrm{d}\rho + \int_{\rho_k}^\infty \rho^{-2} \mathrm{d}\rho \Bigg] \Delta \mathbf{n}_k. \quad (25) \end{aligned}$$

Here, $\rho_k = \alpha_k N_k$, which after plugging in above equation gives:

$$Var(I_N) < \begin{cases} \sum_{k=2}^{m} O\left(\frac{1}{N N_k{}^{b_k}}\right) + O\left(\frac{1}{N}\right) & 0 < b_k < 1 \\ \sum_{k=2}^{m} O\left(\frac{1}{N N_k}\right) + O\left(\frac{1}{N}\right) & b_k \geq 1 \end{cases}. \quad (26)$$

Since, the sum of $O(\cdot)$ notations is asymptotically dominated by the worst $O(\cdot)$ behavior (Sec. 3). We can write the overall convergence rate, using Eq. (26), in the following form:

$$\boxed{Var(I_N) < O\left(\frac{1}{N}\right)}. \quad (27)$$

This shows that, the asymptotic convergence rate of a sampling pattern, with anisotropic sampling power spectra, is dominated by the direction that has a constant radial behavior (which results in the $O(1/N)$ convergence rate). However, if none of the directions has a constant profile behavior, the overall convergence rate can be summarized from Eq. (26) in the following form:

$$\boxed{Var(I_N) < \begin{cases} O\left(\frac{1}{N N_k{}^{b_k}}\right) & \substack{0<b_k<1 \\ \text{infimum}} \\ O\left(\frac{1}{N N_k}\right) & b_k \geq 1 \end{cases}}, \quad (28)$$

where, the overall convergence rate would be dominated by the k-th direction having the minimum value of b_k, for $b_k \in (0, 1)$. Similarly,

for the best case:

$$Var(I_N) < O\left(\frac{1}{N}\right) + \sum_{k=2}^{m}\left[\int_0^{\rho_0} \rho^{d-1}\frac{\gamma_k}{N}\left(\frac{\rho}{\alpha_k N_k}\right)^{b_k} c_f \mathrm{d}\rho\right]\Delta\mathbf{n}_k \quad (29)$$

$$= O\left(\frac{1}{N}\right) + \frac{\gamma_k\, c_f}{N}\sum_{k=2}^{m}\left[\int_0^{\rho_0} \frac{\rho^{b_k+d-1}}{(\alpha_k N_k)^{b_k}}\mathrm{d}\rho\right]\Delta\mathbf{n}_k\,, \quad (30)$$

implying:

$$Var(I_N) < \sum_{k=2}^{m} O\left(\frac{1}{N {N_k}^{b_k}}\right) + O\left(\frac{1}{N}\right)\,, \quad (31)$$

where, $b_k > 0$. Again, since the sum of $O(\cdot)$ notations is asymptotically dominated by the worst $O(\cdot)$ behavior. The overall convergence rate is:

$$\boxed{Var(I_N) < O\left(\frac{1}{N}\right)} \quad (32)$$

However, from Eq. (31), if none of the directions has a constant profile behavior, the overall convergence rate would be dominated by the k-th direction having the minimum value of b_k, and can be written in the following form:

$$\boxed{Var(I_N) < O\left(\frac{1}{N {N_k}^{b_k}}\right) \forall_{\substack{b_k>0 \\ \text{infimum}}}} \quad (33)$$

□

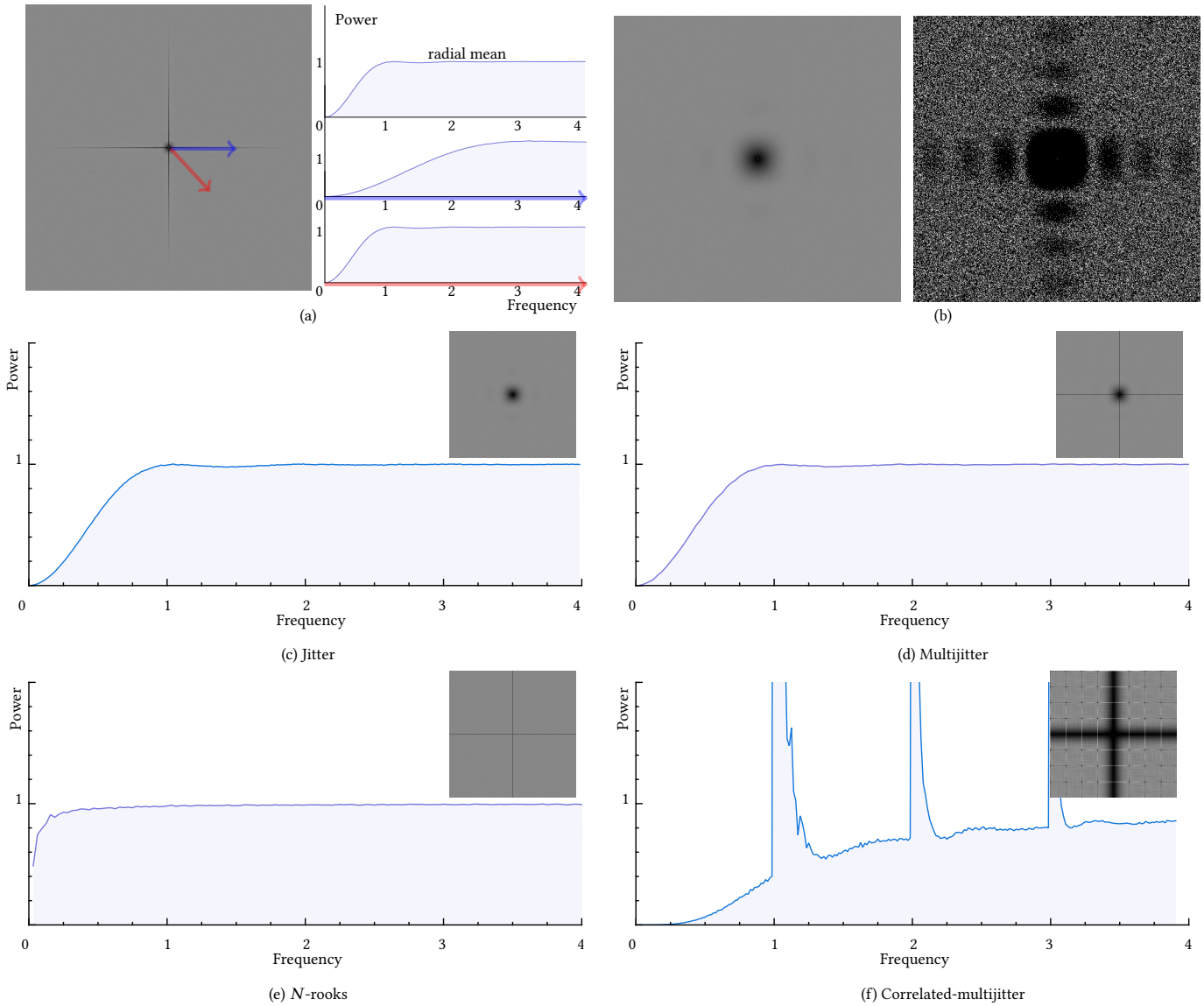


Fig. 3. **Top row**: (a) The expected power spectrum of multijittered samples with $N = 256$ samples is shown which has hairline anisotropic structures along the canonical axes. This is because along the canonical axis the radial behavior has jittered profile with $N_k = N = 256$ strata whereas, along all other directions the radial behvior has jittered profile with only $N_k = \sqrt{N} = 16$ strata (blue vs. red arrows). Radial averaging (radial mean) masks the good anisotropic properties of the sampler along the canonical axes. (b) The expected power spectrum of jittered samples is shown on the left side whereas the logarithmically scaled values of this spectrum are shown on the right to emphasize the anisotropic structures present in the jittered spectrum. **Middle and bottom row**: The expected power spectra and the corresponding expected radial mean spectra of jittered, multijittered, N-rooks and the correlated-multijittered sampling patterns is shown. As we can see, jittered and multijittered have identically looking radial mean shape despite the dark cross region in the multi-jittered spectrum.

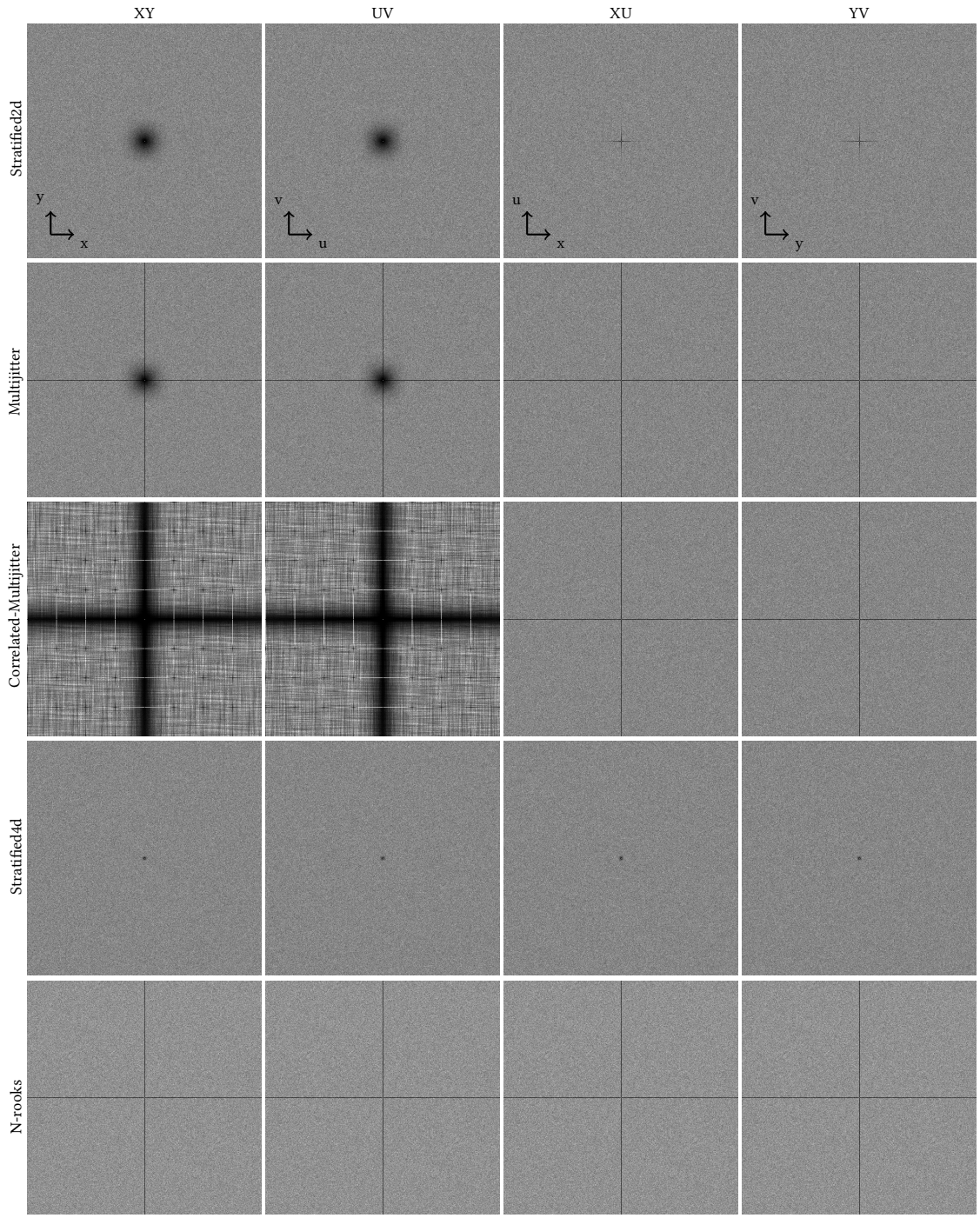


Fig. 4. The expected power spectra is shown for different samplers which is computed using $N = 4096$ samples for randomly shuffled jittered (Stratified2d, which is also known as uncorrelated jitter) sampler, randomly shuffled multijittered, randomly shuffled correlated-multijittered and Stratified4d (jittering directly in 4D). The uncorrelated jittered sampling (which we refer to as random shuffling) involves sampling the original 2D subspaces (XY & UV) with jittered samples and then connecting these two subspaces with random shuffling to form a 4D sampler. The same task can be performed for multjittered and correlated-multijittered samples. In the mixed dimensions (XU, and YV) multijittered and correlated-multijittered exhibits the Latin-hypercube behaviour in their power spectra. Latin-hypercube sampling power spectrum is also shown in the bottom row.

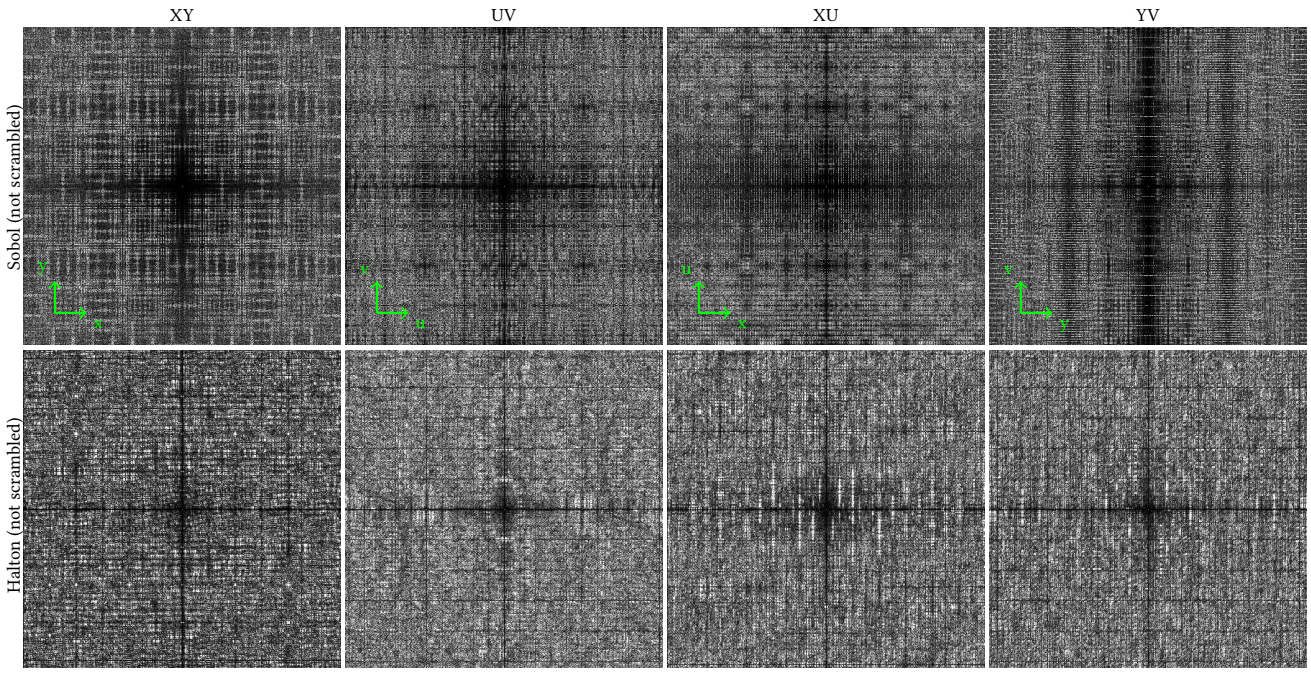


Fig. 5. The expected power spectra for Sobol and Halton samplers is shown for the first four dimensions along different 2D projections (XY, UV, XU, YV).

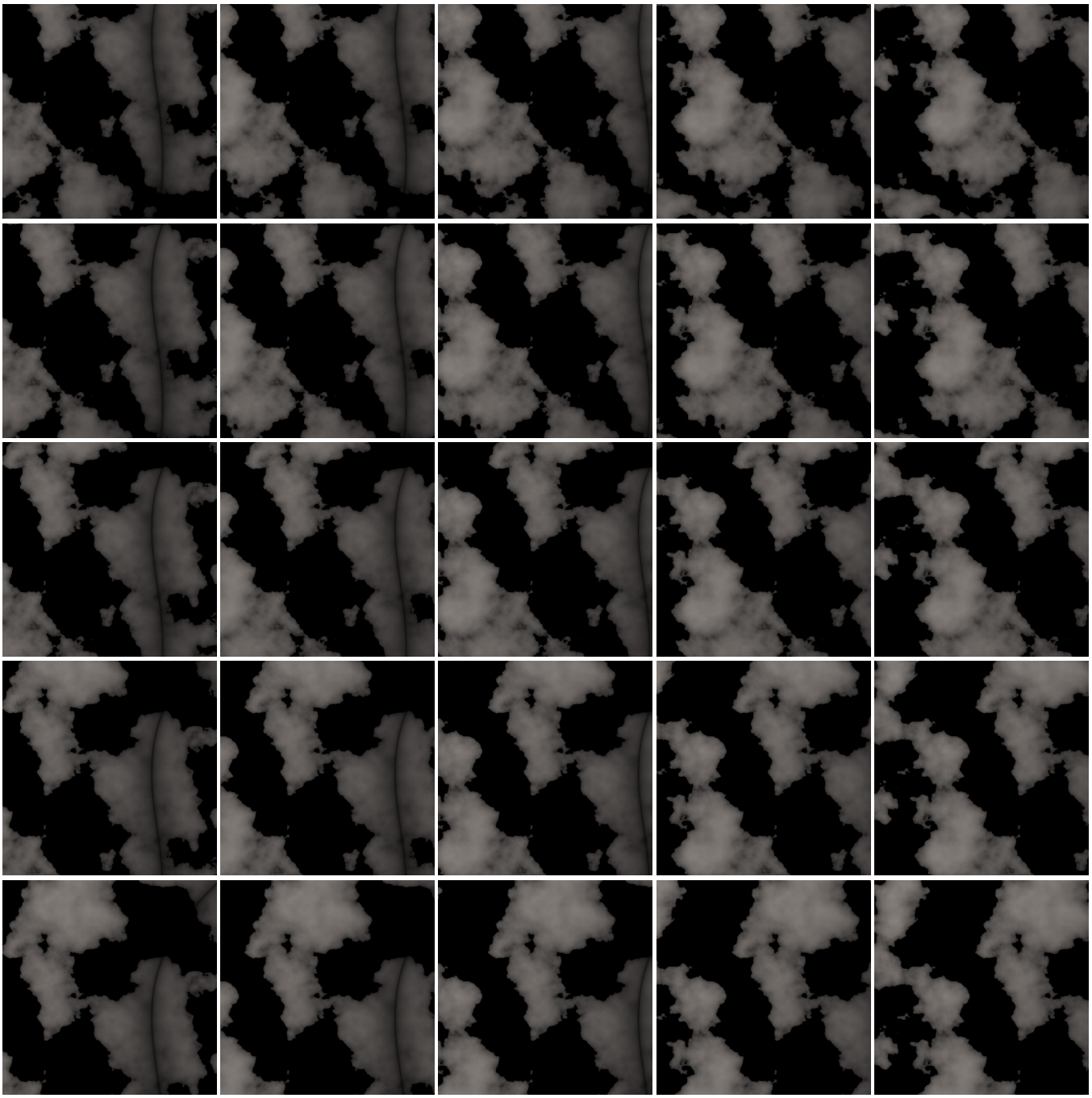


Fig. 6. **Pixel 2:** Here we illustrate the 4D light field for **Pixel 2** of Fig. 7 from the main paper. The object behind Pixel 2 is rendered from different (u, v) locations on the lens. These rendered images are tiled together with the center tile showing the image rendered from the center of the lens $(u = 0, v = 0)$.

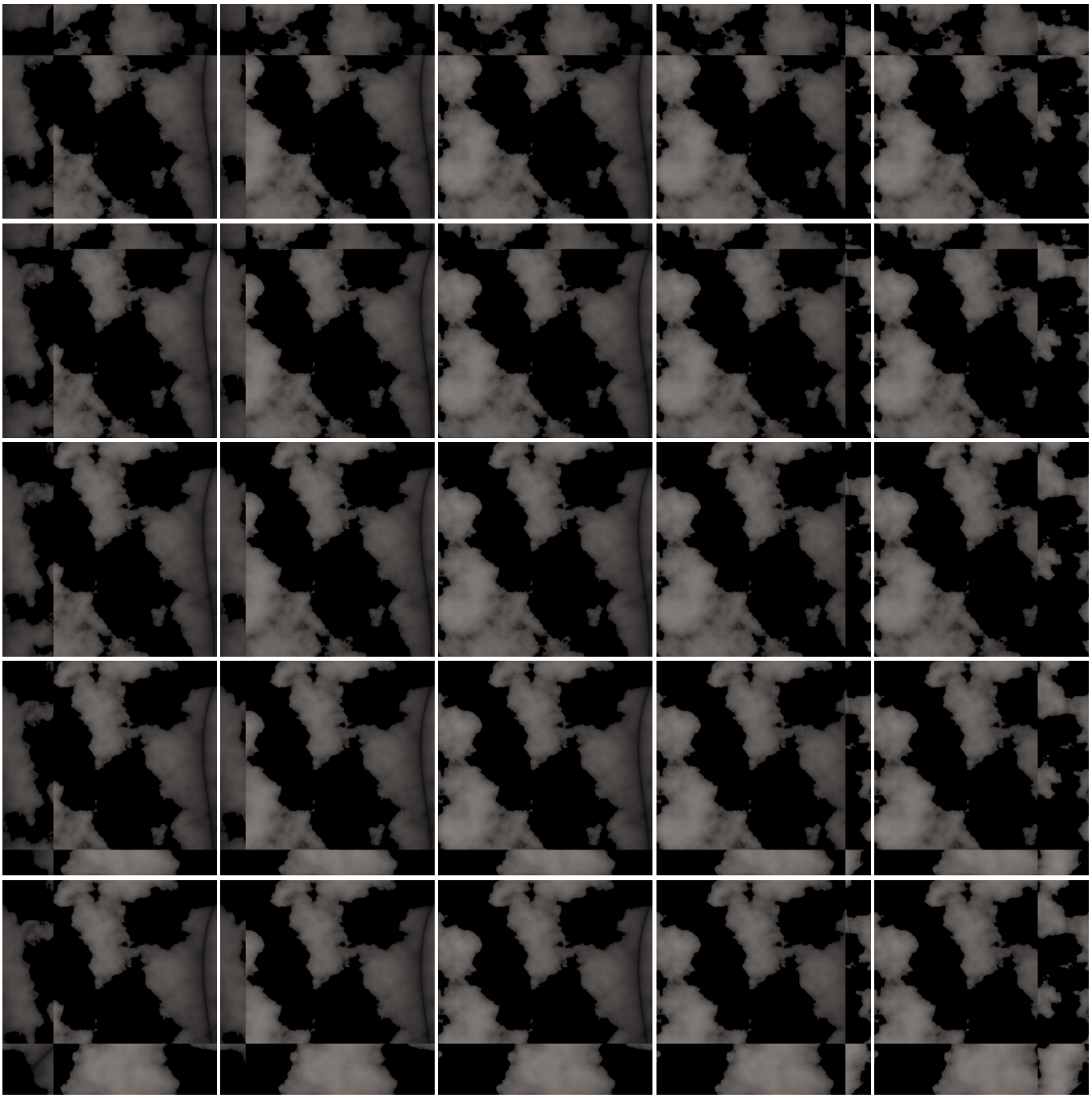


Fig. 7. **Pixel 2:** Here we illustrate the inversely sheared 4D light field for **Pixel 2** of Fig. 7 from the main paper, after tororidal wrapping. The original light field is shown in Fig. 6. The object behind Pixel 2 is rendered from different (u, v) locations on the lens. These rendered images are tiled together with the center tile showing the image rendered from the center of the lens ($u = 0$, $v = 0$), which is not affected by the shear.

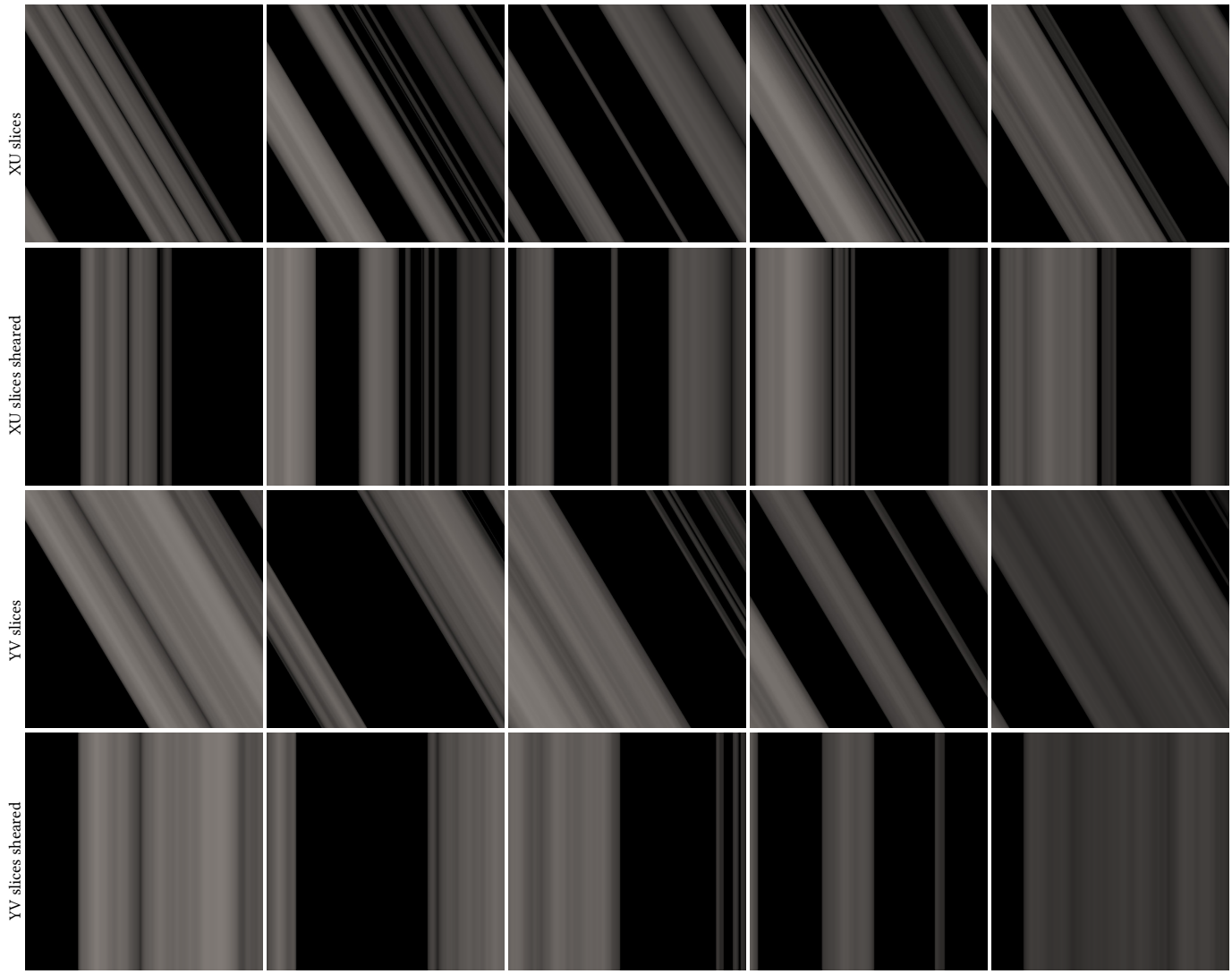


Fig. 8. **Pixel 2:** Slices in the mixed (XU and YV) dimensions are shown before and after inverse shearing the light field. XY is the image plane and UV is the lens space.

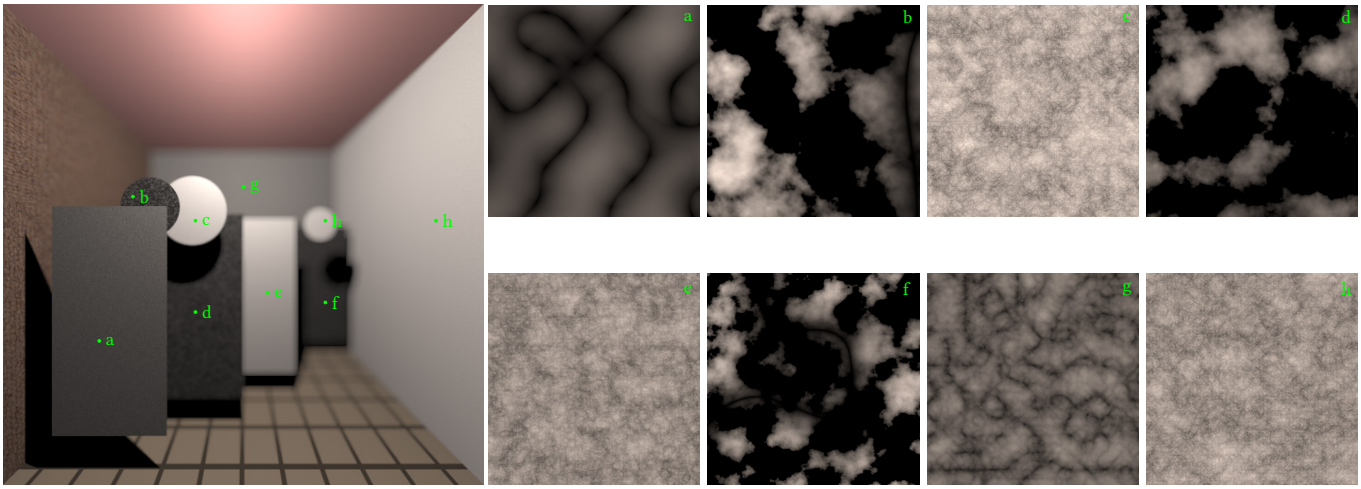


Fig. 9. The Cornell Box scene is rendered with no defocus blur to illustrate the high frequency texture underneath each pixel of each object (plane or disk) that is generated using perlin noise. This scene is used in the main paper to investigate the impact of shearing the samples on the variance convergence rate for the depth-of-field effect.

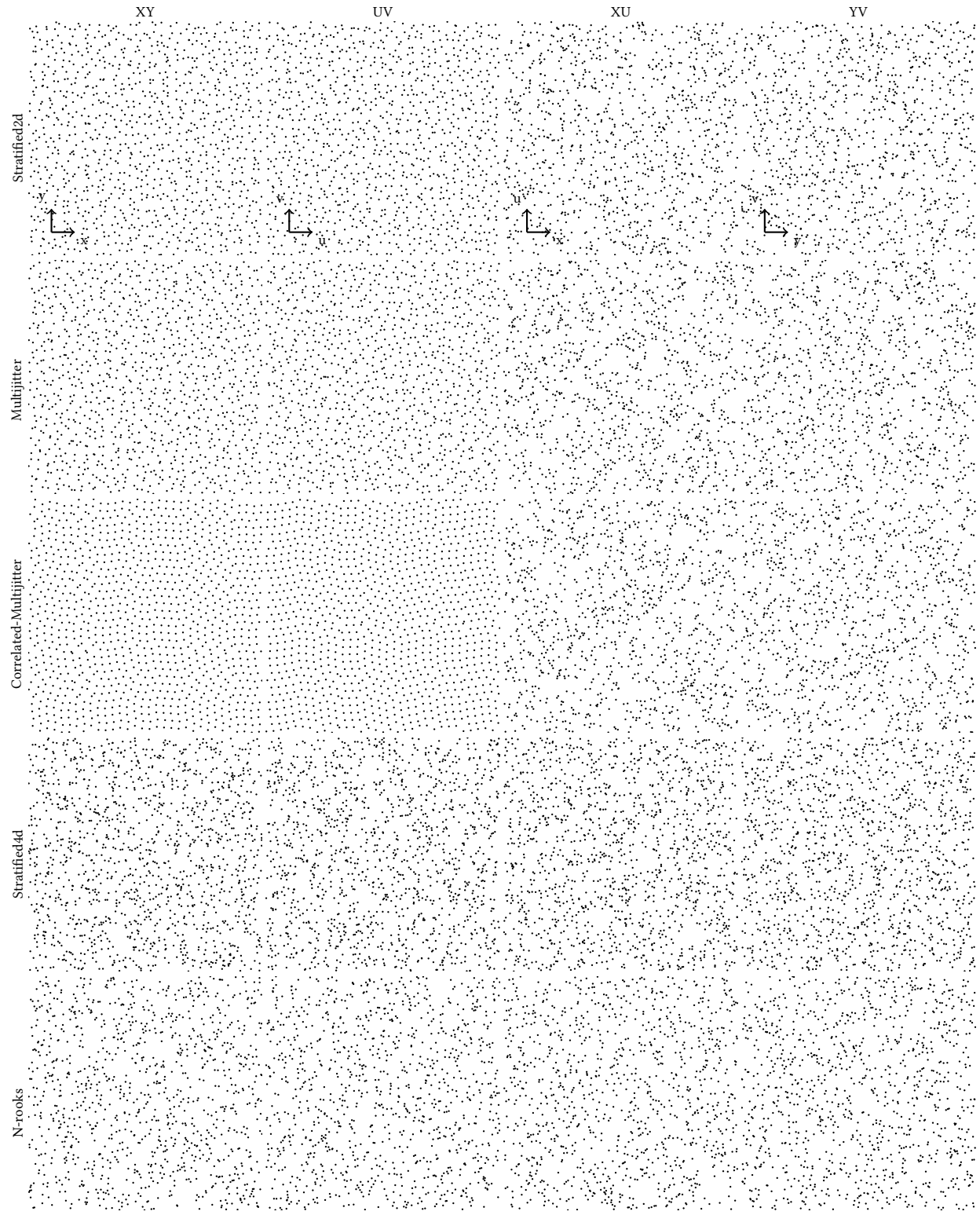


Fig. 10. Point sets generated with $N = 1024$ samples for randomly shuffled jittered (Stratified2d, aka uncorrelated jittered), randomly shuffled multijittered, randomly shuffled correlated-multijittered samplers. Stratified4d (jittering directly in 4D) is shown with $N = 1296$ samples. The corresponding expected power spectra for each projection is shown in Fig. 4.

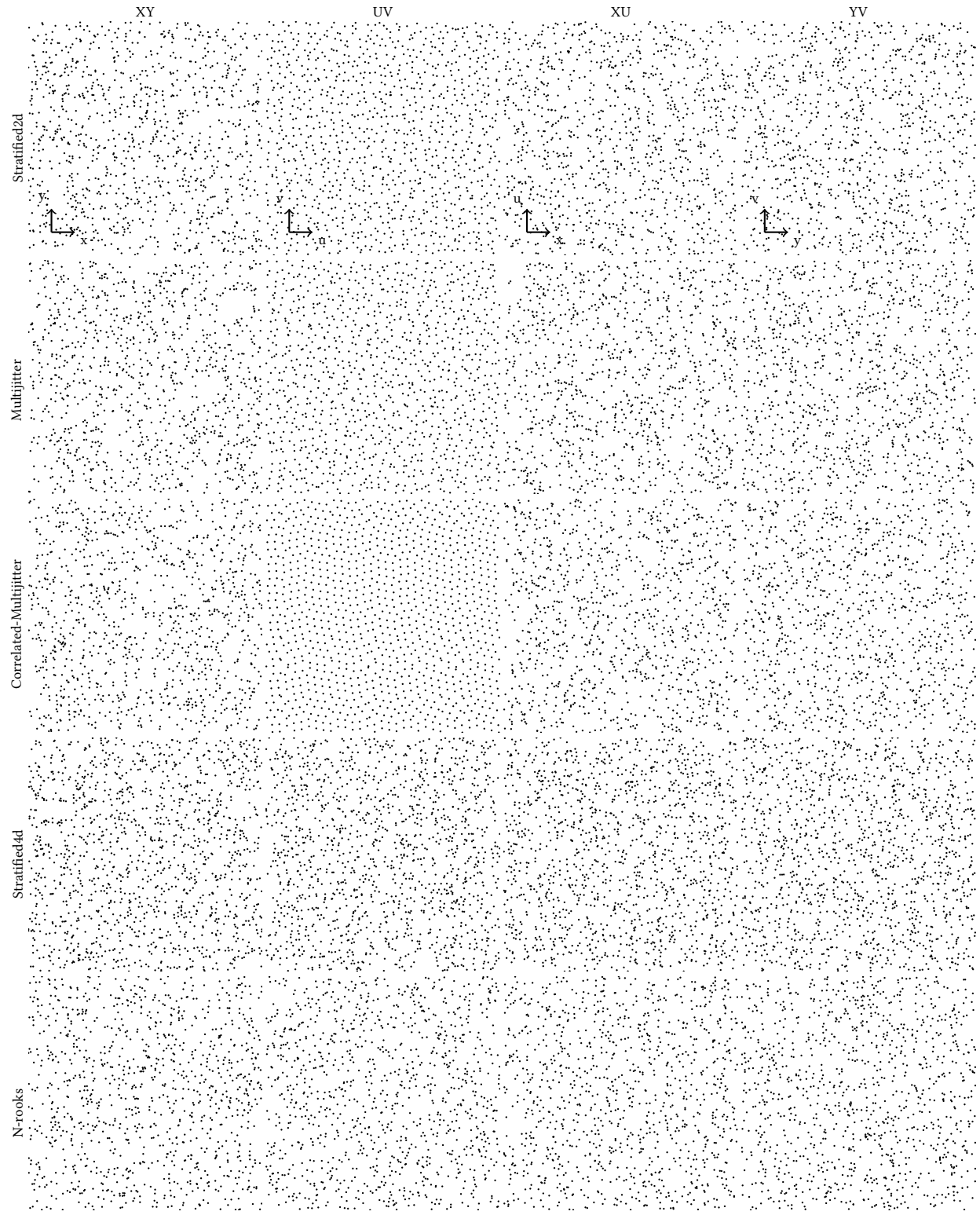


Fig. 11. The XY coordinates shown in Fig. 10 are sheared w.r.t. the UV coordinates. As a result, the XY projection samples look less uniform due to the shearing of the underlying stratification. The UV projection is unaffected since no transformation is applied on this projection.

REFERENCES

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

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

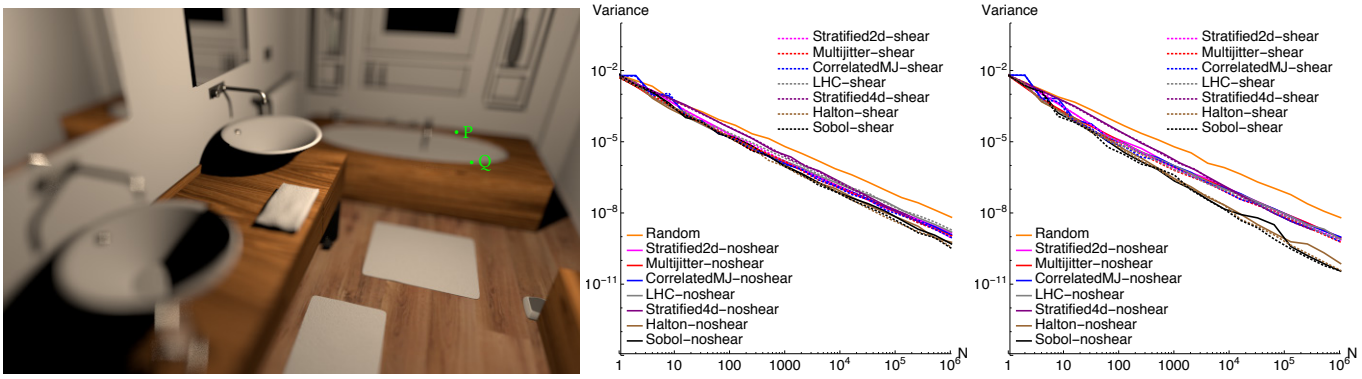
Thomas H. Cormen, Clifford Stein, Ronald L. Rivest, and Charles E. Leiserson. 2001. *Introduction to Algorithms* (2nd ed.). McGraw-Hill Higher Education.

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

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

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

Salvatore Torquato, Obiama U. Uche, and Frank H. Stillinger. 2006. Random sequential addition of hard spheres in high Euclidean dimensions. *Physical Review E* 74, 6 (2006).



(a) (b) Pixel P (c) Pixel Q

Fig. 12. **Bathroom scene:**. We render a bathroom scene with a point light source and a finite size lens to ensure 4D underlying integral (due to depth of field only) in each of the pixel. In the variance plots in (b) and (c) we do not observe any major improvements with existing samplers even after shearing because all the pixels in this scene have varying depth. This results in a wedge shaped frequency content for all of the pixels. Since existing samplers have only hairline anisotropy, the alignment between these low energy hairline structures w.r.t. the high energy wedge spectrum of the integrand (after shearing) is not enough to get any variance or convergence improvement.

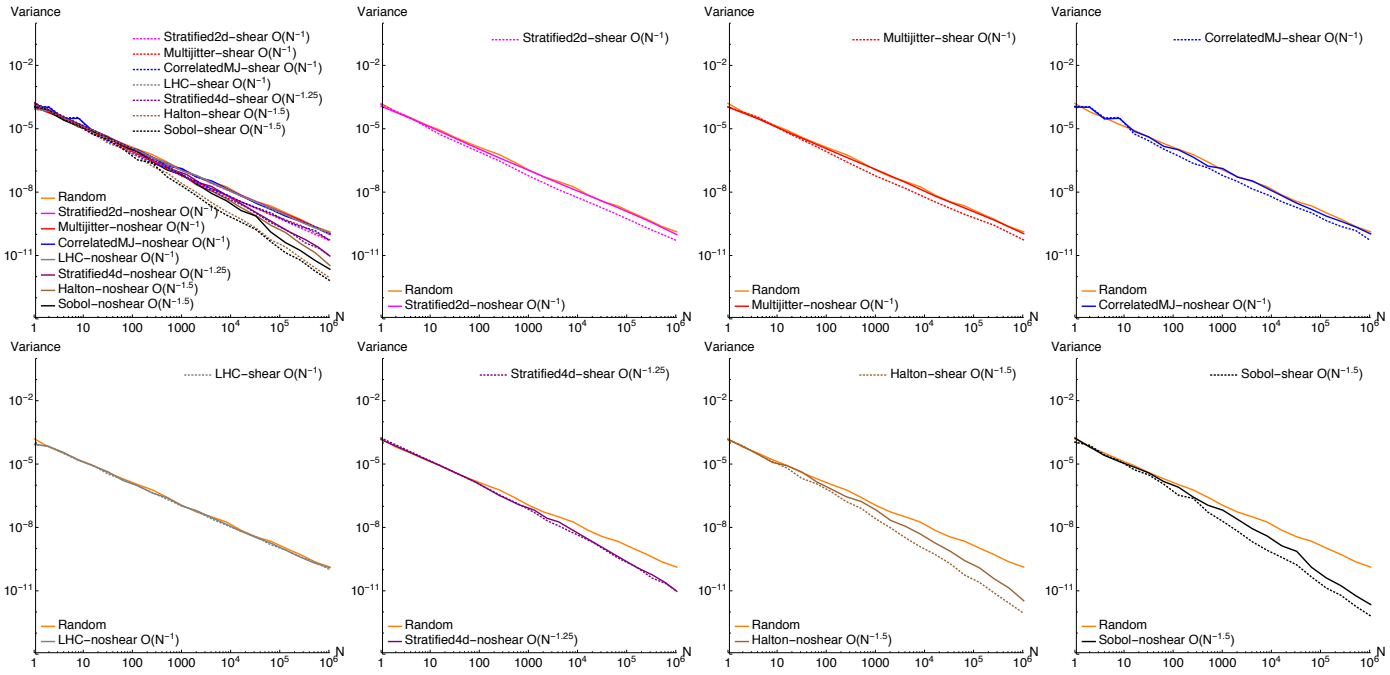


Fig. 13. **Pixel 2:** Here we are illustarting the improvements in variance and convergence rate after shearing the samples for the Cornell Box scene which is rendered with defocus blur (see Fig .7 of the main paper). Random sampler plot is shown for reference.

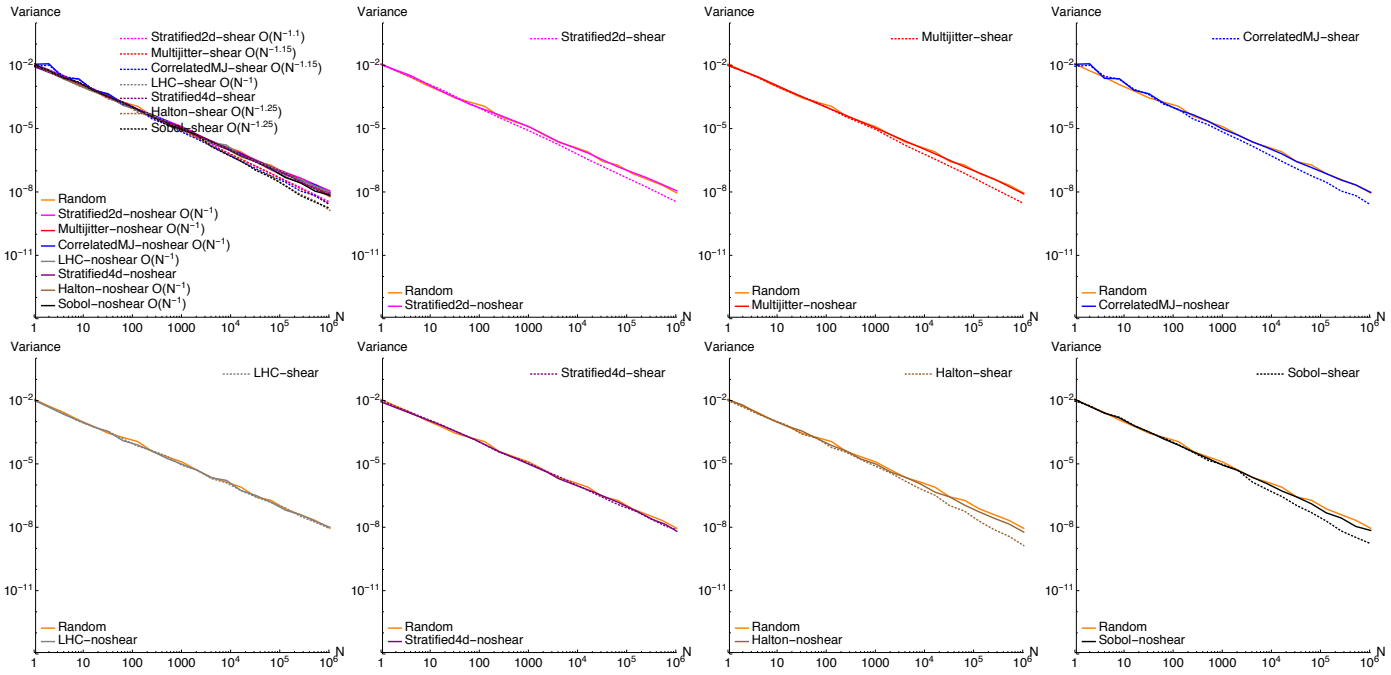


Fig. 14. **Pixel 3:** Here we are illustarting the improvements in variance and convergence rate after shearing the samples for the Cornell Box scene which is rendered with defocus blur (see Fig .7 of the main paper). Random sampler plot is shown for reference.

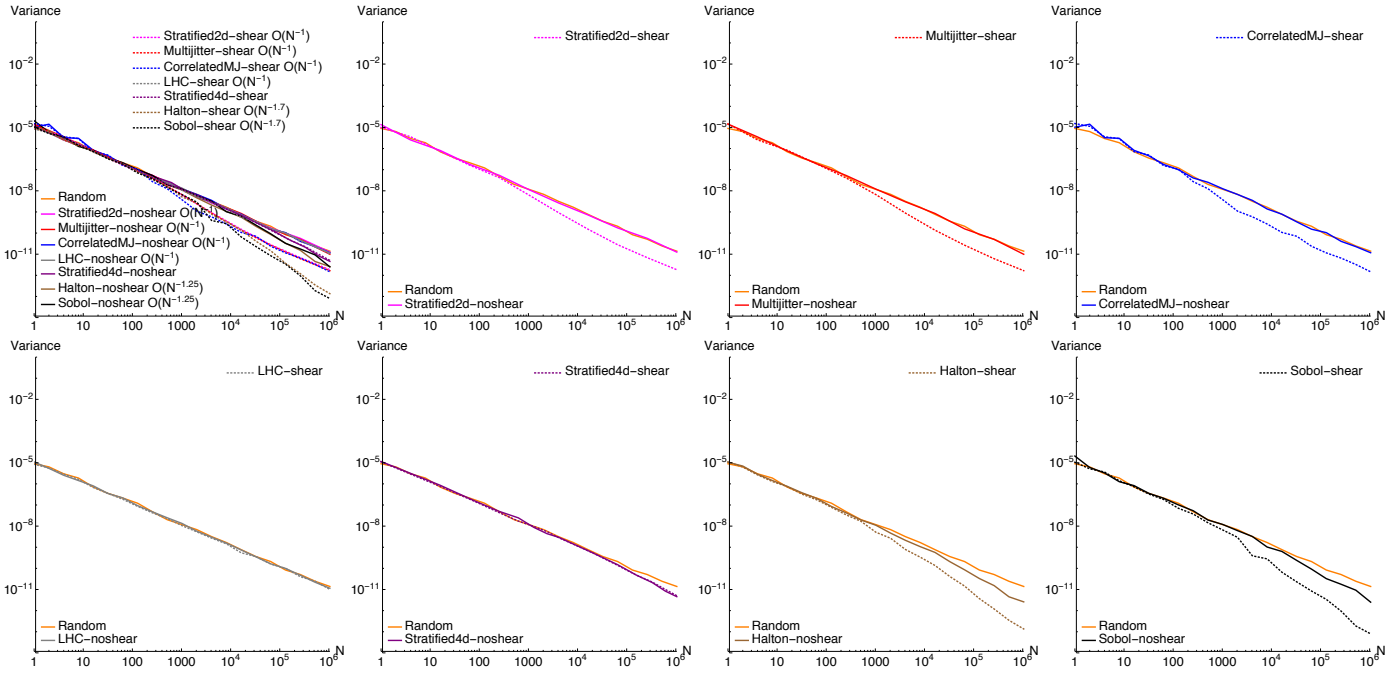


Fig. 15. **Pixel 4:** Here we are illustarting the improvements in variance and convergence rate after shearing the samples for the Cornell Box scene which is rendered with defocus blur (see Fig .7 of the main paper). Random sampler plot is shown for reference.

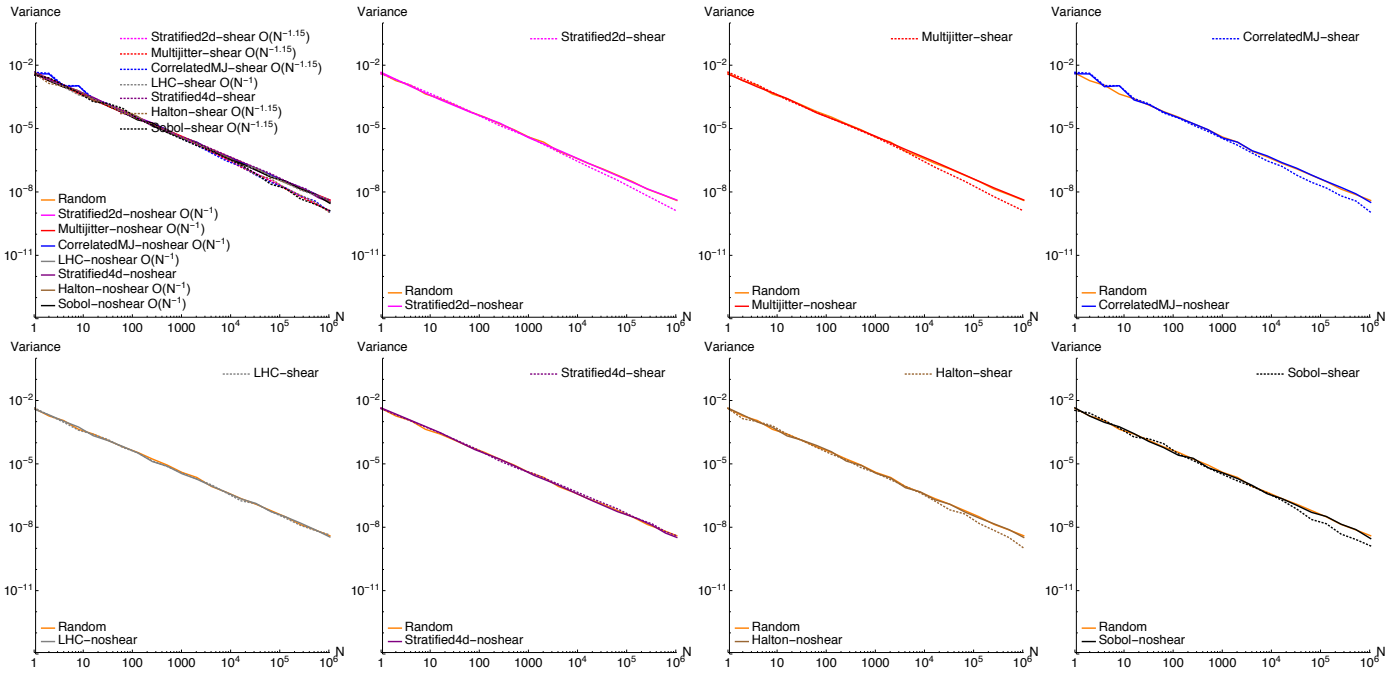


Fig. 16. **Pixel 5:** Here we are illustarting the improvements in variance and convergence rate after shearing the samples for the Cornell Box scene which is rendered with defocus blur (see Fig .7 of the main paper). Random sampler plot is shown for reference.

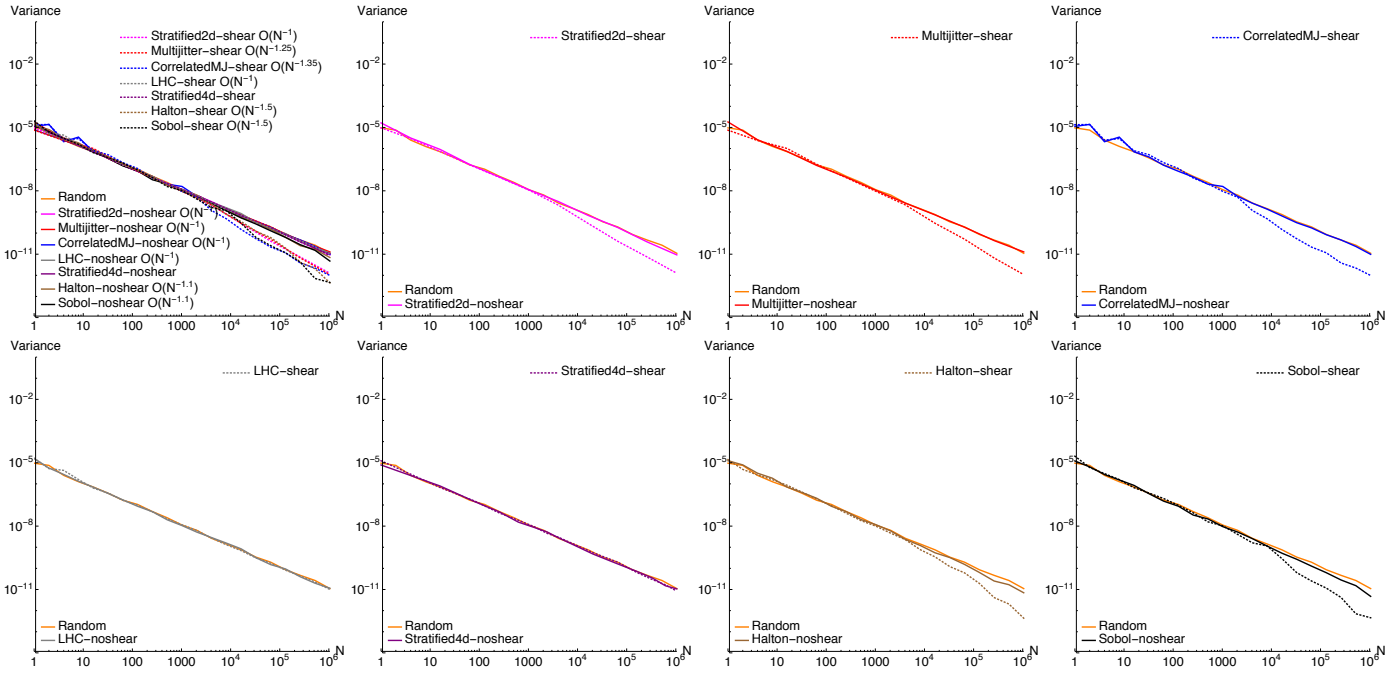


Fig. 17. **Pixel 6:** Here we are illustarting the improvements in variance and convergence rate after shearing the samples for the Cornell Box scene which is rendered with defocus blur (see Fig .7 of the main paper). Random sampler plot is shown for reference.

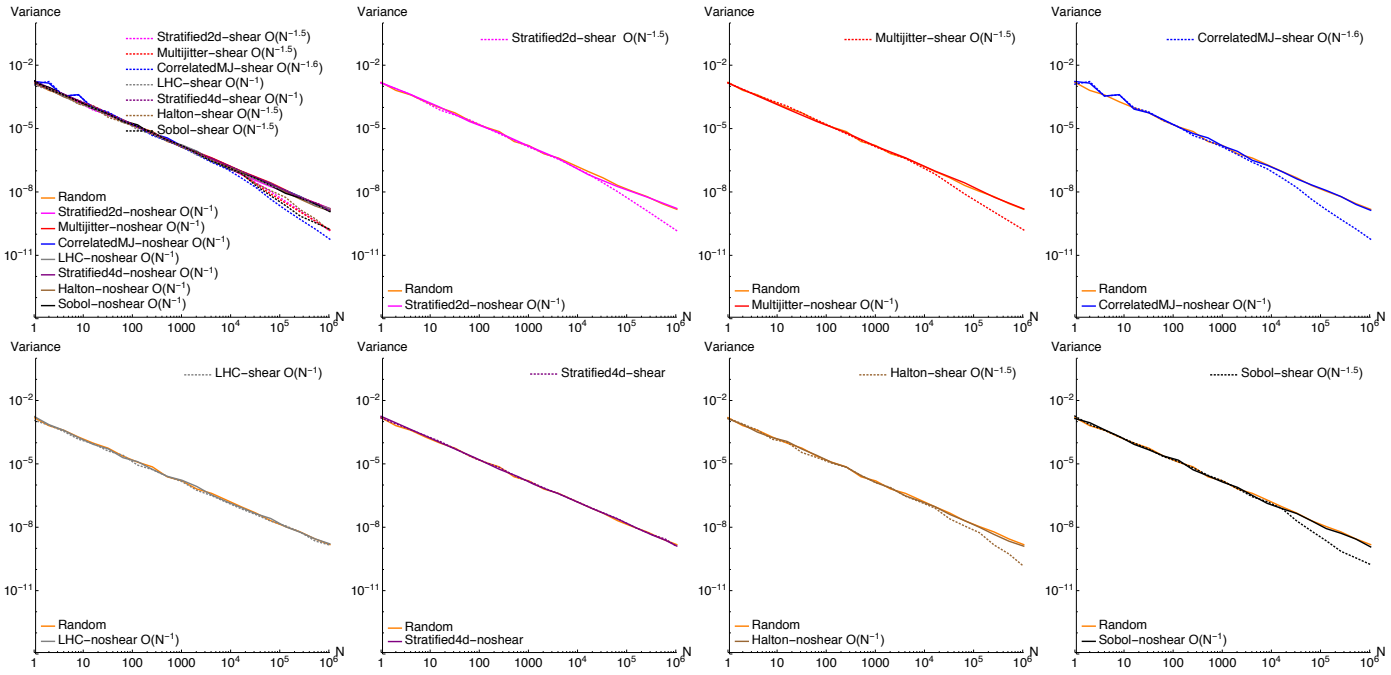


Fig. 18. **Pixel 7:** Here we are illustarting the improvements in variance and convergence rate after shearing the samples for the Cornell Box scene which is rendered with defocus blur (see Fig .7 of the main paper). Random sampler plot is shown for reference.

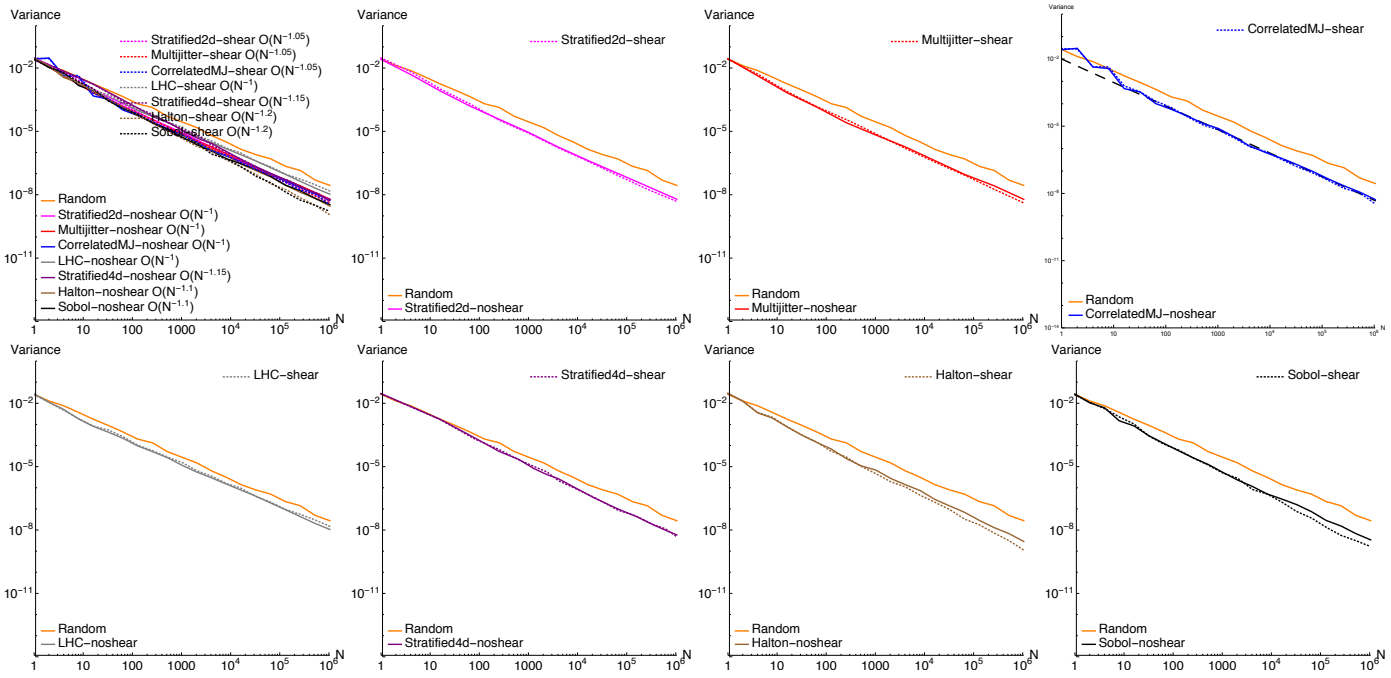


Fig. 19. **Pixel 8:** Here we are illustarting the improvements in variance and convergence rate after shearing the samples for the Cornell Box scene which is rendered with defocus blur (see Fig .7 of the main paper). Random sampler plot is shown for reference.

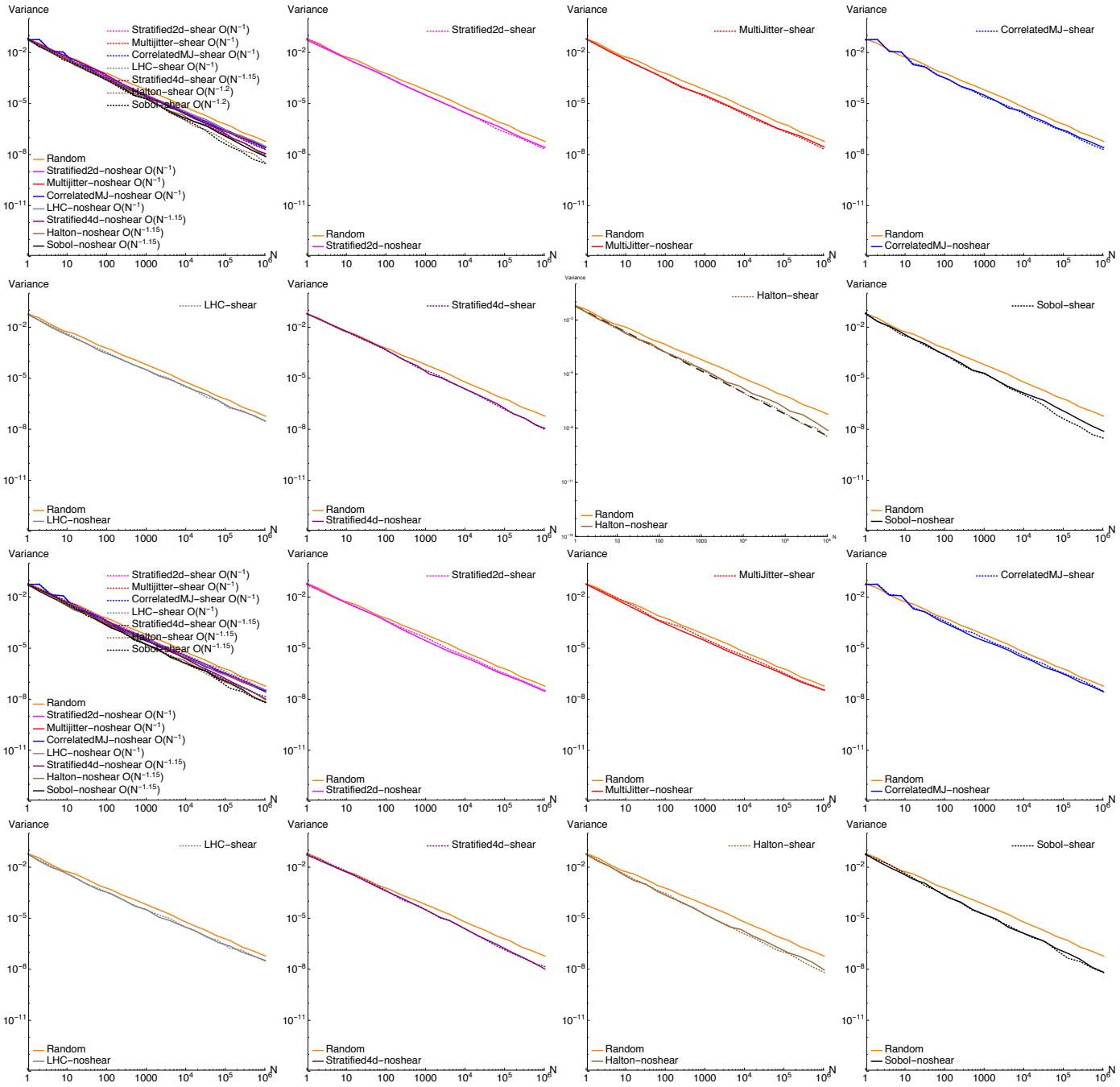


Fig. 20. **Pixel 9:** Here we show the variance plots for Pixel 9 (with one occluder) for the Cornell Box scene shown in Fig. 11 of the main paper. **Top two rows** show convergence improvement when shearing is performed with respect to the minimum depth. **Bottom two rows** show convergence plots when shearing is performed w.r.t the maximum depth in the pixel.

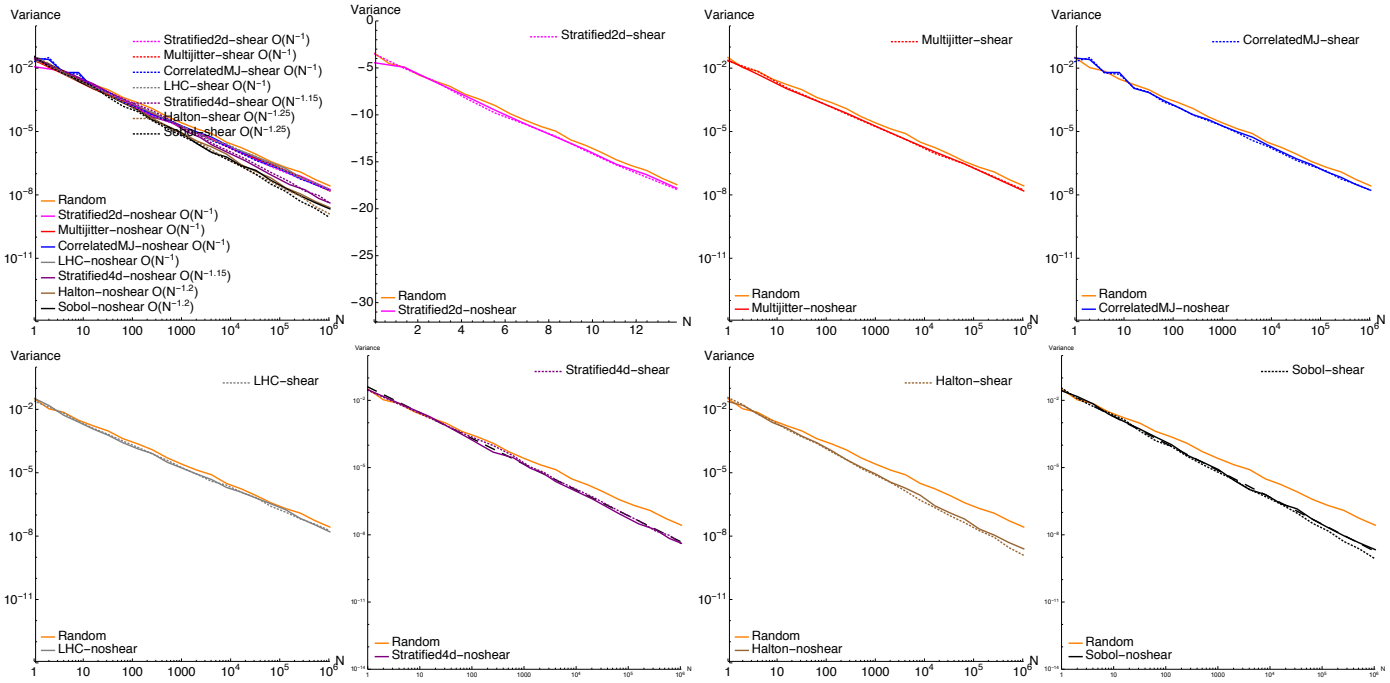


Fig. 21. **Pixel 10:** Here we are illustarting the improvements in variance and convergence rate after shearing the samples for the Cornell Box scene which is rendered with defocus blur (see Fig .7 of the main paper). Random sampler plot is shown for reference.

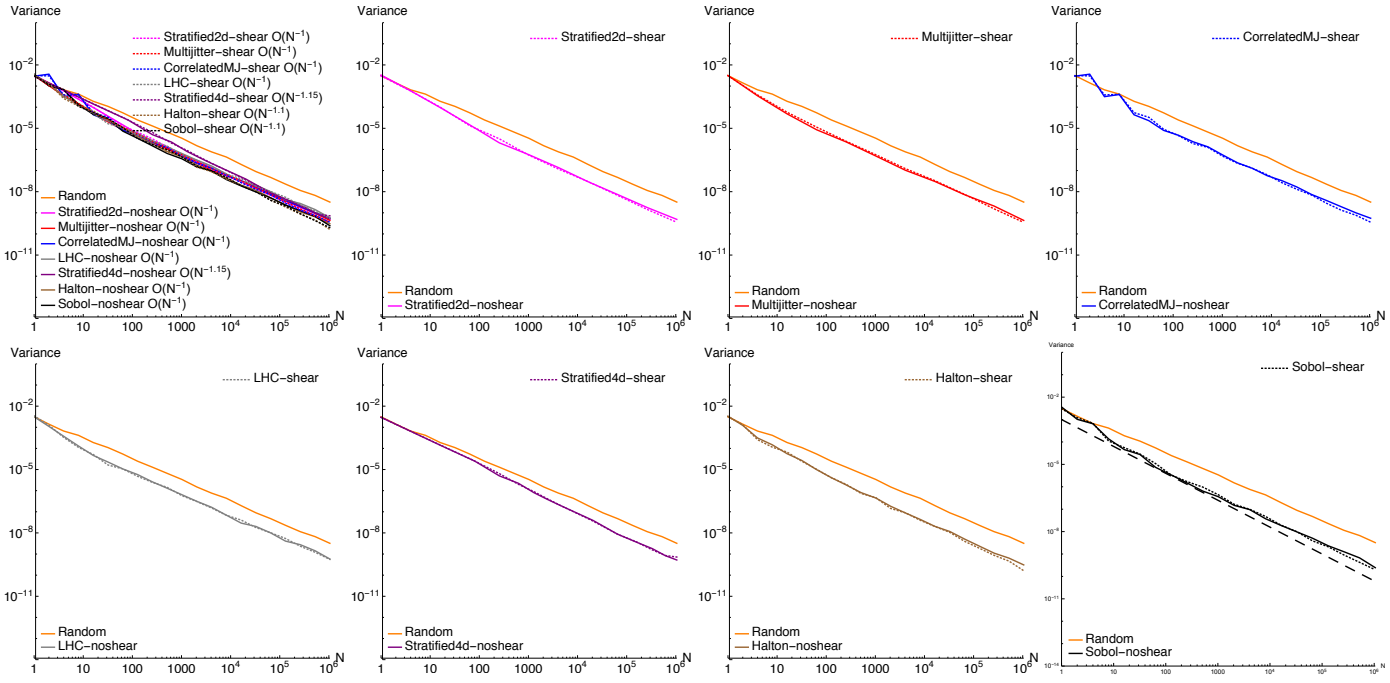


Fig. 22. **Pixel 11:** Here we are illustarting the improvements in variance and convergence rate after shearing the samples for the Cornell Box scene which is rendered with defocus blur (see Fig .7 of the main paper). Random sampler plot is shown for reference.

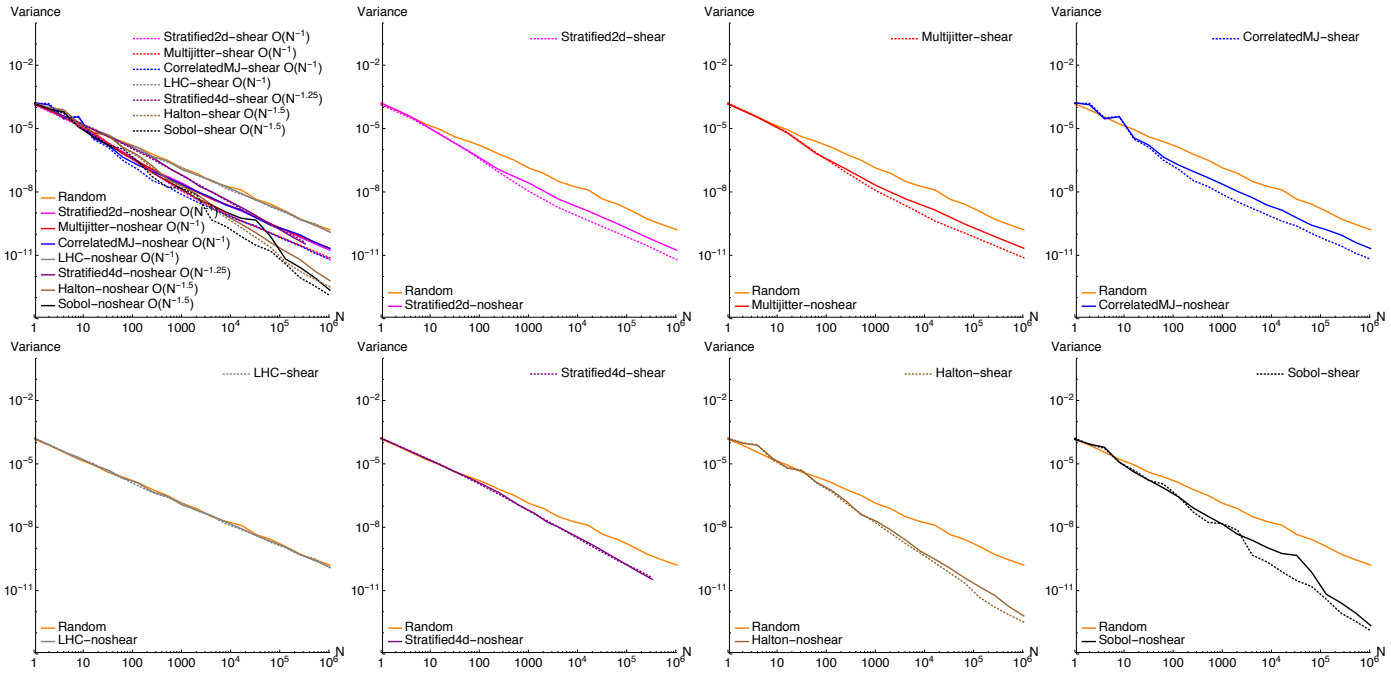


Fig. 23. **Pixel 12:** Here we are illustarting the improvements in variance and convergence rate after shearing the samples for the Cornell Box scene which is rendered with defocus blur (see Fig .7 of the main paper). Random sampler plot is shown for reference.

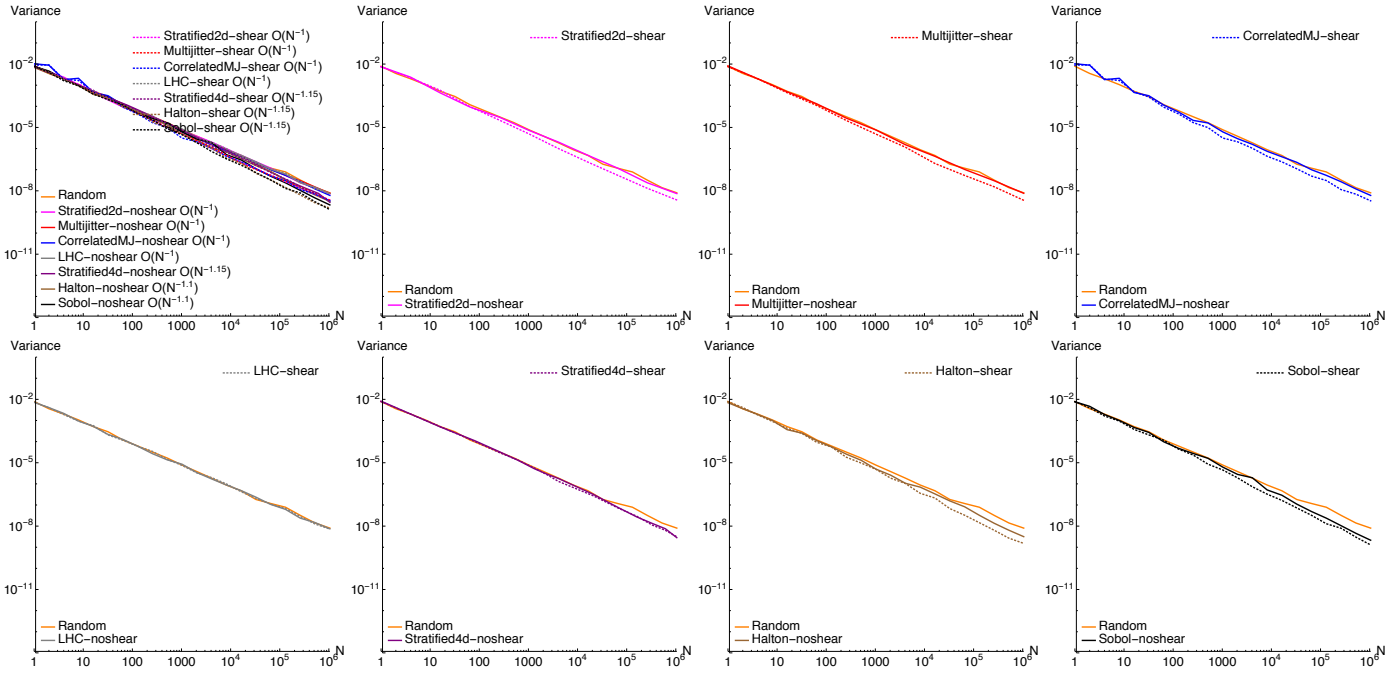


Fig. 24. **Pixel 13:** Here we are illustarting the improvements in variance and convergence rate after shearing the samples for the Cornell Box scene which is rendered with defocus blur (see Fig .7 of the main paper). Random sampler plot is shown for reference.

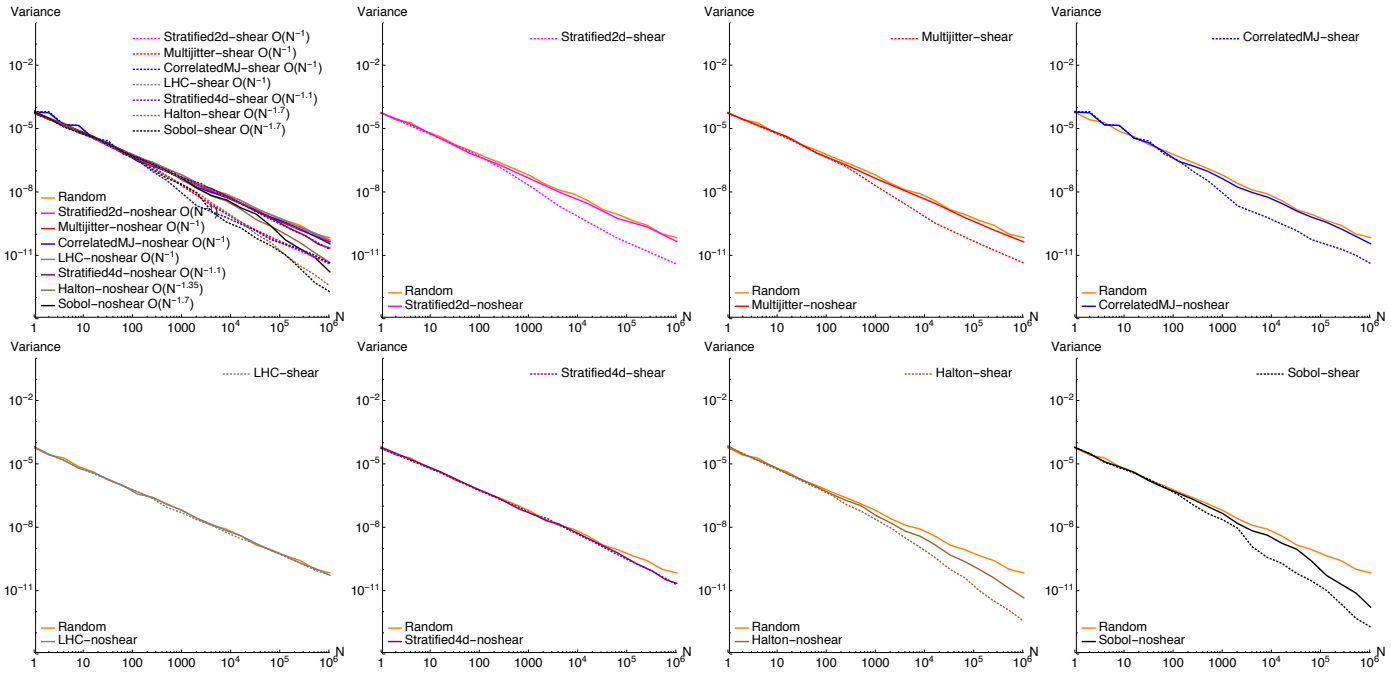


Fig. 25. **Pixel 14:** Here we are illustarting the improvements in variance and convergence rate after shearing the samples for the Cornell Box scene which is rendered with defocus blur (see Fig .7 of the main paper). Random sampler plot is shown for reference.