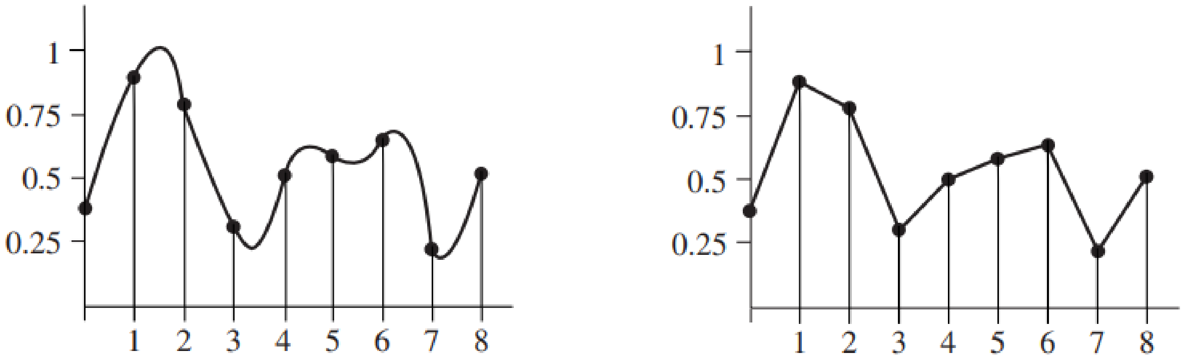
Low Discrepancy Sampling

Sampling and Reconstruction Theory

Displays use the image pixel values to construct a new image function over the display surface. This function is defined at all points on the display, not just the infinitesimal points of the digital image's pixels. This process of taking a collection of sample values and converting them back to a continuous function is called *reconstruction*. In order to compute the discrete pixel values in the digital image, it is necessary to sample the original continuously defined image function.

Because the sampling and reconstruction process involves approximation, it introduces error known as *aliasing*, which can manifest itself in many ways, including jagged edges or flickering in animations. These errors occur because the sampling process is not able to capture all of the information from the continuously defined image function.



1D sampling and reconstruction example

Ideal Sampling and Reconstruction

We will try to deduce the theory from Fourier analysis in 1D. Denote the function we want to reconstruct $f(x)$ and its Fourier transformations $F(\omega)$. The reconstructed function in spatial and frequency space are respectively $\tilde{f}(x)$ and $\tilde{F}(\omega)$.

Sampling and Reconstruction in Space

The sampling process requires us to choose a set of equally spaced sample positions and compute the function's value at those positions. Which is equivalent to multiply $f(x)$ with a shah function

$$m_T(x) = T \sum_{i=-\infty}^{\infty} \delta(x - iT)$$

where T defines the period, or sampling rate. The sampled values are thus

$$m_T(x)f(x) = T \sum_i \delta(x - iT)f(iT)$$

These sample values can be used to define a reconstructed function $\tilde{f}$ by choosing a reconstruction filter function $r(x)$ and computing the convolution

$$(m_T(x)f(x)) * r(x)$$

For reconstruction, convolution gives a weighted sum of scaled instances of the reconstruction filter centered at the sample points:

$$\tilde{f}(x) = T \sum_{i=-\infty}^{\infty} f(iT)r(x - iT)$$

Sampling and Reconstruction in Frequency

We can gain a deeper understanding of the sampling process by analyzing the sampled function in the frequency domain.

In particular, we will be able to determine the conditions under which the original function can be exactly recovered from its values at the sample locations.

Assume that the function $f(x)$ is band limited. By definition, band-limited functions have frequency space representations with compact support, such that $F(\omega) = 0$ for all $|\omega| > \omega_0$ where ω_0 is the band limit. Assume also that we have $F(\omega)$ available in our hands.

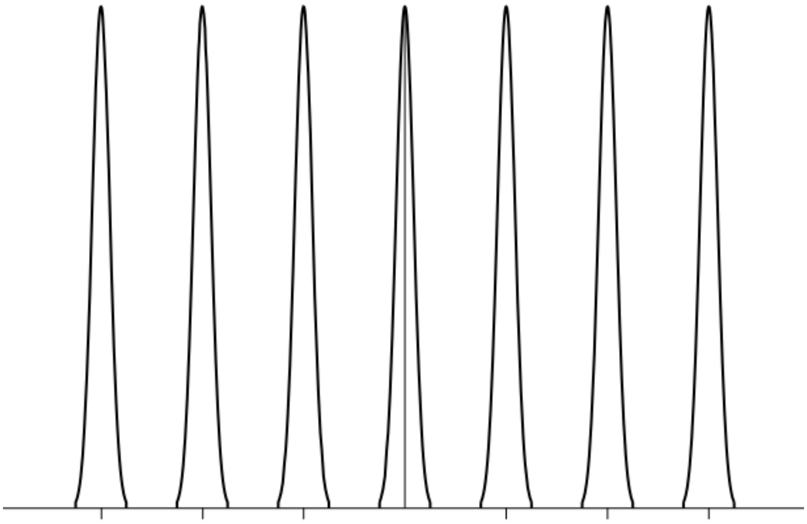
Recall an important property of Fourier transform operator $\mathcal{F}$

$$\begin{aligned} \mathcal{F}(f(x)g(x)) &= F(\omega) * G(\omega) \\ \mathcal{F}(f(x) * g(x)) &= F(\omega)G(\omega) \end{aligned}$$

and the Fourier transform of the shah function is

$$\mathcal{F}(m_T(x)) = m_{\frac{1}{T}}(\omega)$$

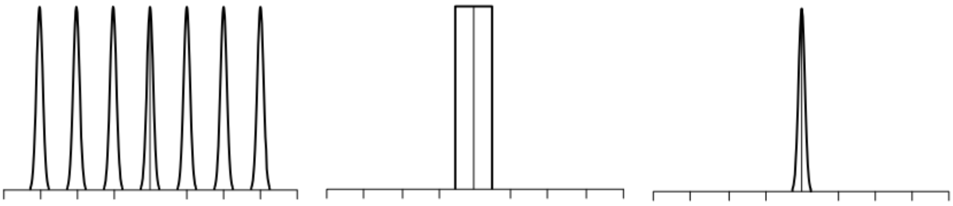
The frequency domain representation of the sampled signal is given by the convolution of $F(\omega)$ and this new shah function. Convolving a function with a delta function just yields a copy of the function, so convolving with a shah function yields an infinite sequence of copies of the original function, with spacing equal to the period of the shah.



The Convolution of F(ω) and the Shah Function. The result is infinitely many copies of F

In order to throw away all but the center copy of the spectrum, we multiply by a gate function of the appropriate width. The gate function $g_T(x)$ of width T is defined as

$$g_T(x) = \begin{cases} \frac{1}{2T} & |x| < T \\ 0 & \text{otherwise} \end{cases}$$



Multiplying a series of copies of F(ω) by the appropriate box function yields the original spectrum.

This multiplication step corresponds to convolution with the reconstruction filter in the spatial domain. This is the ideal sampling and reconstruction process:

$$\tilde{F} = \left(F(\omega) * m_{\frac{1}{T}(\omega)}\right) g_{T(\omega)} = F(\omega)$$

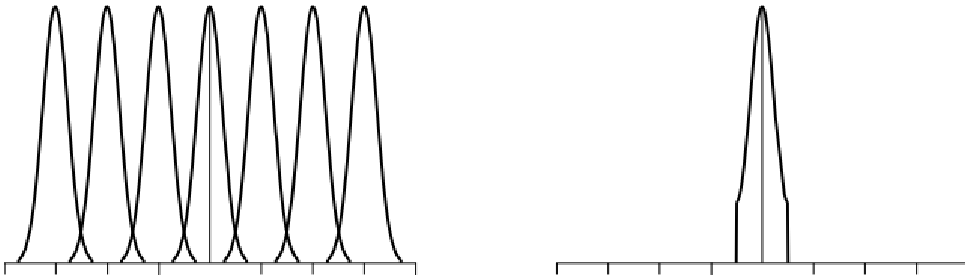
Take the inverse Fourier transform of $\tilde{F}$ we get the ideal sampling and reconstruction process in spatial domain:

$$\tilde{f} = (f(x) m_T(x)) * \mathrm{sinc}(x) = \sum_{i=-\infty}^{\infty} \mathrm{sinc}(x-i) f(i)$$

Unfortunately, because the sinc function has infinite extent, it is necessary to use all of the sample values $f(i)$ to compute any particular value of $\tilde{f}(x)$ in the spatial domain. Filters with finite spatial extent are preferable for practical implementations even though they don't reconstruct the original function perfectly.

Aliasing

One of the most serious practical problems with the ideal sampling and reconstruction approach is the assumption that the signal is band limited.



> *Sampling Theorem*
> As long as the frequency of uniform sample points ω_s is greater than twice the maximum frequency present in the signal ω_0, it is possible to reconstruct the original signal perfectly from the samples. This minimum sampling frequency is called the Nyquist frequency.

Unfortunately, few of the interesting functions in computer graphics are band limited. In particular, any function containing a discontinuity cannot be band limited, and therefore we cannot perfectly sample and reconstruct it.

Source of Aliasing

- Geometry
  When projected onto the image plane, an object's boundary introduces a step function—the image function's value instantaneously jumps from one value to another.
- Very small objects
  If the geometry is small enough that it falls between samples on the image plane, it can unpredictably disappear and reappear over multiple frames of an animation
- Texture and materials
  Shading aliasing can be caused by texture maps that haven't been filtered correctly or from small highlights on shiny surfaces.

The key insight about aliasing in rendered images is that we can never remove all of its sources, so we must develop techniques to mitigate its impact on the quality of the final image.

Anti-aliasing

Nonuniform Sampling

Although the image functions that we will be sampling are known to have infinite-frequency components and thus can't be perfectly reconstructed from point samples, it is possible to reduce the visual impact of aliasing by varying the spacing between samples in a nonuniform way.

For a fixed sampling rate that isn't sufficient to capture the function, both uniform and nonuniform sampling produce incorrect reconstructed signals. However, nonuniform sampling tends to turn the regular aliasing artifacts into noise, which is less distracting to the human visual system. In frequency space, the copies of the sampled signal end up being randomly shifted as well, so that when reconstruction is performed the result is random error rather than coherent aliasing.

Adaptive Sampling

If we can identify the regions of the signal with frequencies higher than the Nyquist limit, we can take additional samples in those regions without needing to incur the computational expense of increasing the sampling frequency everywhere. It can be difficult to get this approach to work well in practice, because finding all of the places where supersampling is needed is difficult. Most techniques for doing so are based on examining adjacent sample values and finding places where there is a significant change in value between the two; the assumption is that the signal has high frequencies in that region

Prefiltering

Another approach to eliminating aliasing that sampling theory offers is to filter (i.e., blur) the original function so that no high frequencies remain that can't be captured accurately at the sampling rate being used.

Evaluating Sampling Patterns: Discrepancy

Given a renderer and a candidate algorithm for placing samples, one way to evaluate the algorithm's effectiveness is to use that sampling pattern to render an image and to compute the error in the image compared to a reference image rendered with a large number of samples. We will use this approach to compare sampling algorithms later in this chapter, though it only tells us how well the algorithm did for one specific scene, and it doesn't give us a sense of the quality of the sample points without going through the rendering process.

Outside of Fourier analysis, mathematicians have developed a concept called *discrepancy* that can be used to evaluate the quality of a pattern of n-dimensional sample positions. Patterns that are well distributed (in a manner to be formalized shortly) have low discrepancy values, and thus the sample pattern generation problem can be considered to be one of finding a suitable low-discrepancy pattern of points.

> *Definition: Discrepancy*:
> Given a sequence of sample points $P = x_1, .., x_N$ and a family of shapes B that are subsets of $[0,1)^n$, the discrepancy of P w.r.t. B is
>
> $$D_N(B,P) = \sup_{b \in B} \left| \frac{\#\{x_i \in b\}}{N} - V(b) \right|$$
>
> where $\#\{x_i \in b\}$ is the number of points in b and $V(b)$ is the volume of b.

When the set of shapes B is the set of boxes with a corner at the origin, this value is called the star discrepancy $D_N^*(P)$.

> *Koksma-Hlawka inequality*
> Discrepancy gives an upper bound on the estimation error:
>
> $$\left| \frac{1}{n} \sum_{i=1}^{n} f(x_i) - \int f(u)\, du \right| \leq V(f) D^*(x_1, \ldots, x_n)$$
>
> where $V(f)$ is the total variation of integrand.

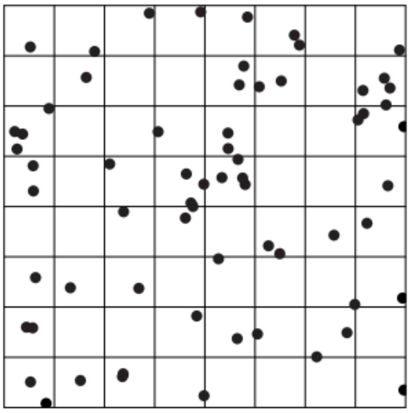
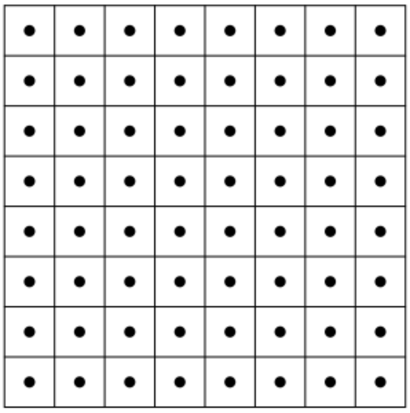
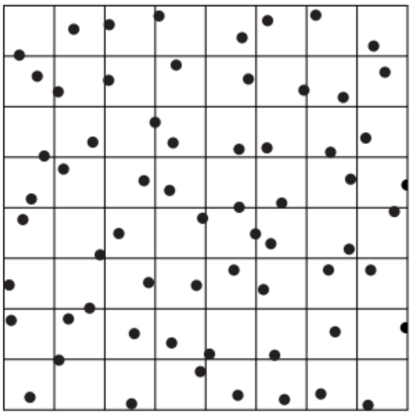
Low Discrepancy Sampling Methods

Stratified(Jittered) Sampling

The stratified sampling subdivides pixel areas into rectangular regions and generates a single sample inside each region. These regions are commonly called *strata*.

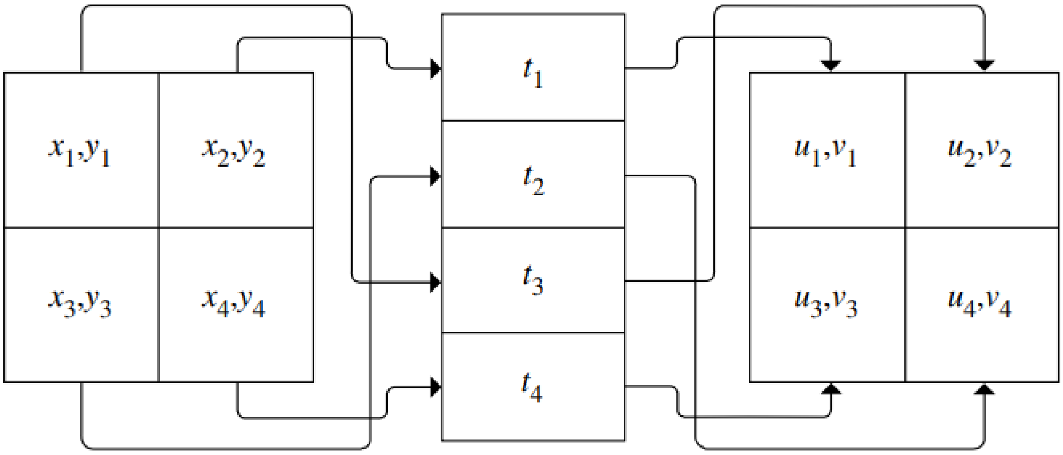
The key idea behind stratification is that by subdividing the sampling domain into nonoverlapping regions and taking a single sample from each one, we are less likely to miss important features of the image entirely, since the samples are guaranteed not to all be close together. Put another way, it does us no good if many samples are taken from nearby points in the sample space, since each new sample doesn't add much new information about the behavior of the image function.

The stratified sampler places each sample at a random point inside each stratum by jittering the center point of the stratum by a random amount up to half the stratum's width and height. The nonuniformity that results from this jittering helps turn aliasing into noise.

Random	Uniformly Stratified	Stratified Jittered
		

Direct application of stratification to high-dimensional sampling quickly leads to an intractable number of samples(curse of dimensionality).

We can reap most of the benefits of stratification without paying the price in excessive total sampling by computing lower dimensional stratified patterns for subsets of the domain's dimensions and then randomly associating samples from each set of dimensions. (This process is sometimes called padding.)



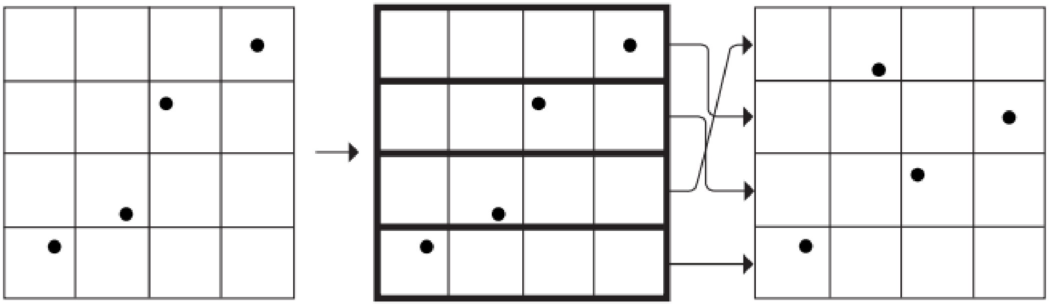
(x,y) are pixel positions, t is ray generation time and (u,v) are lens positions

In the above example, we independently generate four 2D stratified image samples, four 1D stratified time samples, and four 2D stratified lens samples. Then we randomly associate a time and lens sample value with each image sample.

Latin Hypercube(N-Rooks) Sampling

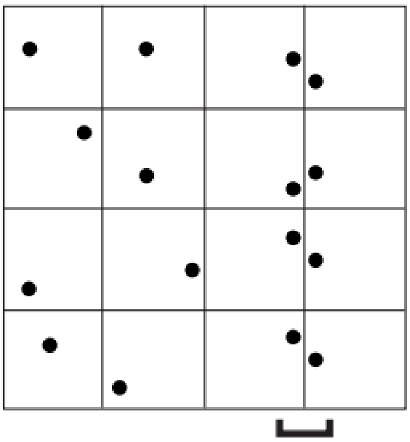
N-Rooks sampling can generate any number of samples in any number of dimensions with a reasonably good distribution.

LHS uniformly divides each dimension's axis into n regions and generates a jittered sample in each of the n regions along the diagonal. These samples are then randomly shuffled in each dimension, creating a pattern with good distribution. This can be done by generating random samples in the cells along the diagonal and then randomly permuting their coordinates.



Shuffle along y-axis is not showed

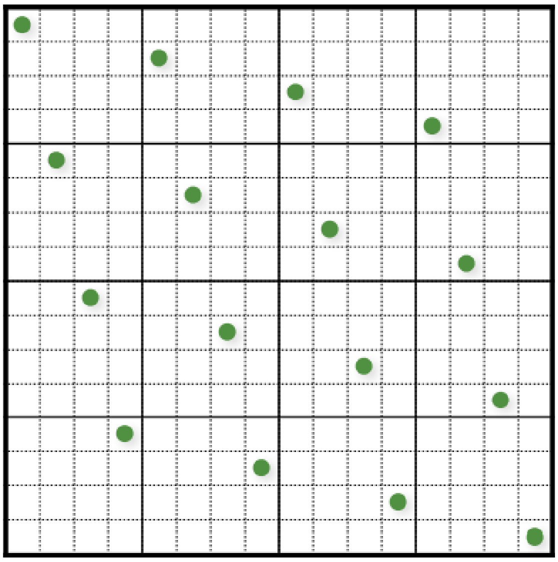
An advantage of LHS is that it minimizes clumping of the samples when they are projected onto any of the axes of the sampling dimensions. This property is in contrast to stratified sampling, where $2n$ of the $n \times n$ samples in a 2D pattern may project to essentially the same point on each of the axes.



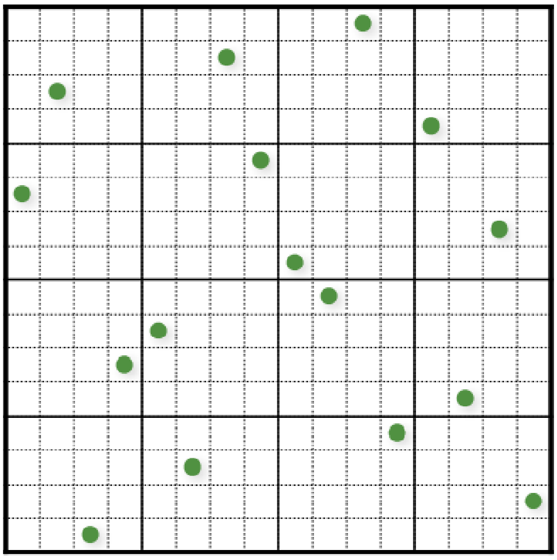
Point clumped in stratified sampling in one dimension

Multi-jittered Sampling

1. Systematically place a single sample in each cell of the upper-level grid while maintaining the n-Rooks condition.



2. Shuffle the samples in the sub-grid while still maintaining the n-Rooks condition and jittered condition.



	Stratification of low-dim projections	Stratification of Full domain
Stratified Sampling	**No**	Yes
N-Rooks Sampling	Yes	**No**
Multi-jittered Sampling	Yes	Yes

Hammersley and Halton Sequences

The Halton and Hammersley sequences are two closely related low-discrepancy point sets. Both are based on a construction called the *radical inverse* which is based on the fact that a positive integer value a can be expressed in a base b with a sequence of digits $d_m(a)\dots d_2(a)d_1(a)$ uniquely determined by

$$a = \sum_{i=1}^{m} d_i(a) b^{i-1}$$

where all digits $d_i(a)$ are between 0 and $b-1$.

The radical inverse function Φ_b in base b converts a nonnegative integer a to a fractional value in $[0, 1)$ by reflecting these digits about the radix point:

$$\Phi_b(a) = 0.d_1(a)d_2(a)\dots d_m(a)$$

To generate points in an n-dimensional Halton sequence, we use the radical inverse base b, with a different base for each dimension of the pattern. The bases used must all be relatively prime to each other, so a natural choice is to use the first n prime numbers $(p_1, \dots, p_n)$:

$$x_a = (\Phi_2(a), \Phi_3(a), \Phi_5(a), \dots, \Phi_{p_n}(a))$$

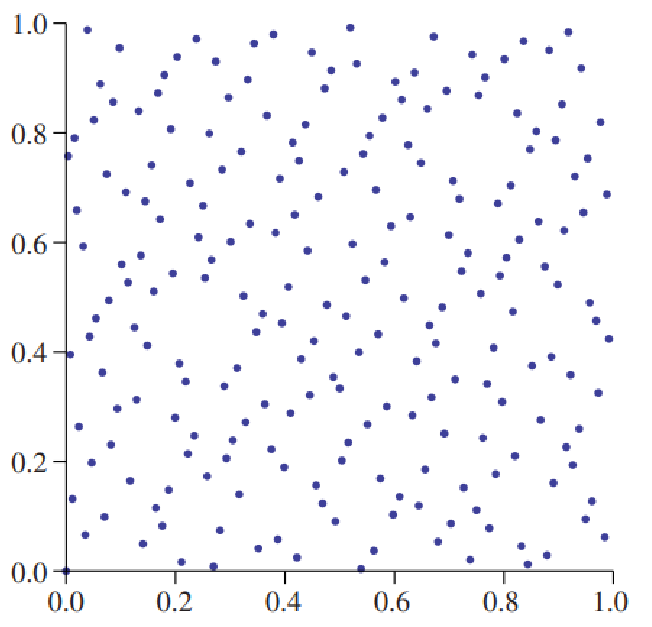
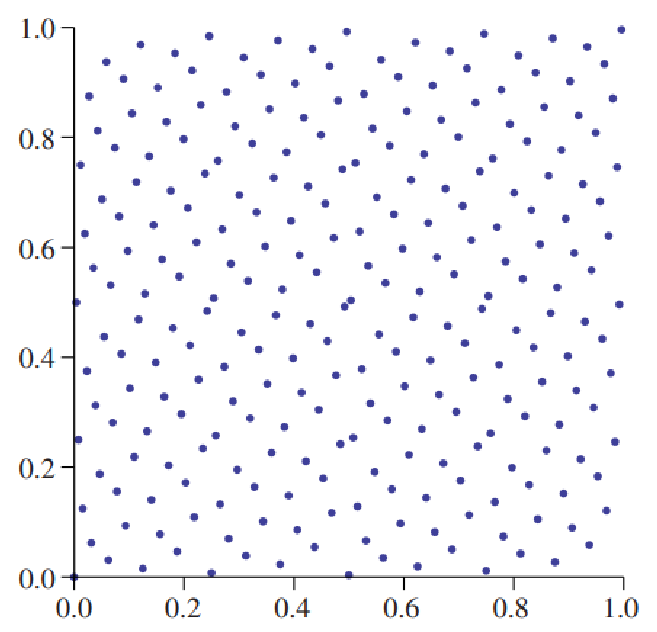
The discrepancy of an n-dimensional Halton sequence is

$$D_N^*(x_a) = O\left(\frac{(\log N)^n}{N}\right)$$

which is asymptotically optimal.

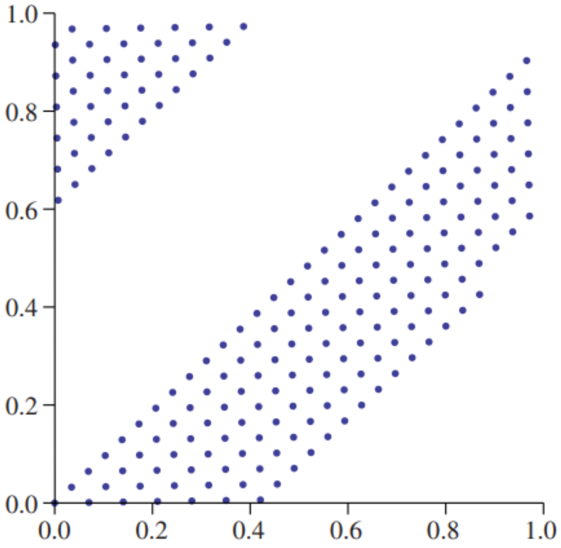
If the number of samples N is fixed, the Hammersley point set can be used, giving slightly lower discrepancy. Hammersley point sets are defined by

$$x_a = \left(\frac{a}{N}, \Phi_{b_1}(a), \Phi_{b_2}(a), \dots, \Phi_{b_n}(a)\right)$$

Halton(216 points)	Hammersley (256 points)
	

Scrambled Halton and Hammersley Sequences

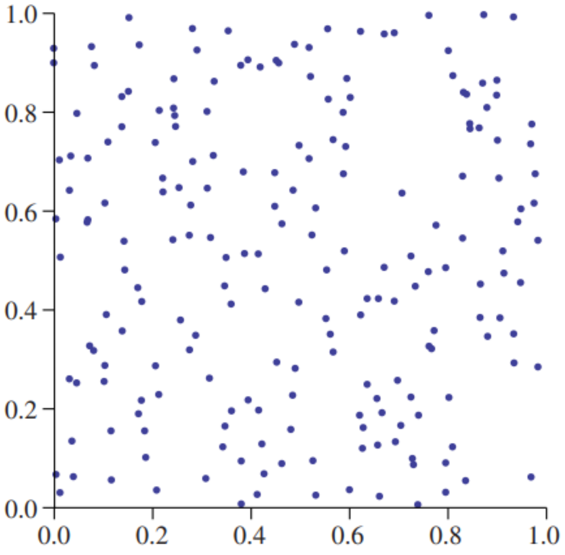
The Hammersley and Halton sequences have the shortcoming that as the base b increases, sample values can exhibit surprisingly regular patterns.



Halton Sample Values without Scrambling

This issue can be addressed with scrambled Halton and Hammersley sequences, where a permutation is applied to the digits when computing the radical inverse:

$$\Psi_b(a) = 0.p(d_1(a))p(d_2(a))\dots p(d_m(a))$$

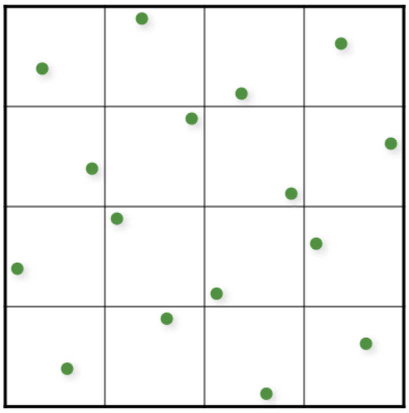
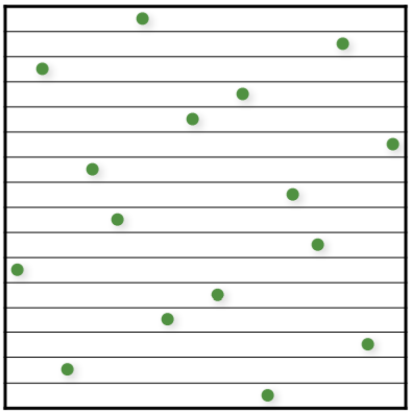
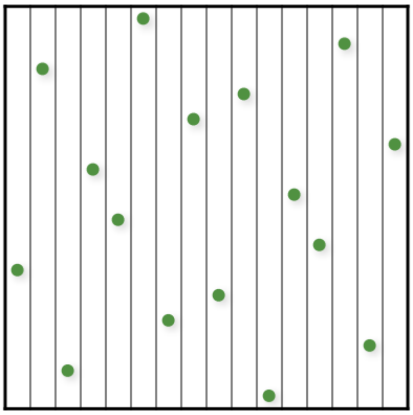
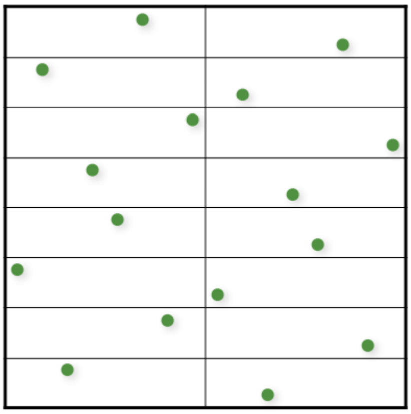
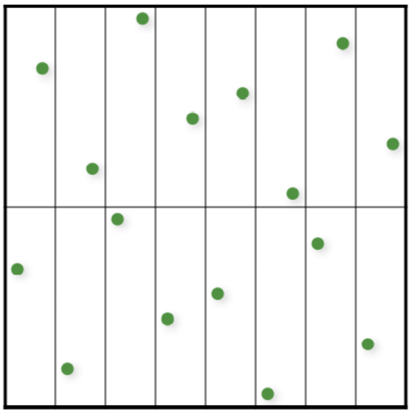


Halton Sample Values with Scrambling

(0,2)-Sequences

$(0,2)$-sequence uses the first two dimensions of a low-discrepancy sequence derived by Sobol. The first 16 samples in a $(0,2)$-sequence satisfy

One sample in each $(\frac{1}{4}, \frac{1}{4})$ box	One sample in each $(\frac{1}{16}, 1)$ and $(1, \frac{1}{16})$ box	One sample in each $(\frac{1}{2}, \frac{1}{8})$ and $(\frac{1}{8}, \frac{1}{2})$ box

One sample in each $(\frac{1}{4},\frac{1}{4})$ box	One sample in each $(\frac{1}{16},1)$ and $(1,\frac{1}{16})$ box	One sample in each $(\frac{1}{2},\frac{1}{8})$ and $(\frac{1}{8},\frac{1}{2})$ box
	 	 

In general, any sequence of length $2^{l_1+l_2}$ (where l_i is a nonnegative integer) from a $(0,2)$-sequence satisfies the general stratification constraint: one element in one *elementary intervals*. The set of elementary intervals in two dimensions, base 2, is defined as

$$E=\left\{\left[\frac{a_1}{2^{l_1}},\frac{a_1+1}{2l_1}\right)\times\left[\frac{a_2}{2^{l_2}},\frac{a_2+1}{2^{l_2}}\right)\right\}$$

where the integer $a_i=0,1,2,4\ldots,2^{l_i}-1$ and l_1 and l_2 are variables.

References

CS 563 Advanced Topics in Computer Graphics by Wadii Bellamine(wpi.edu)
Sampling and Reconstruction (pbr-book.org)