Monte Carlo Integration

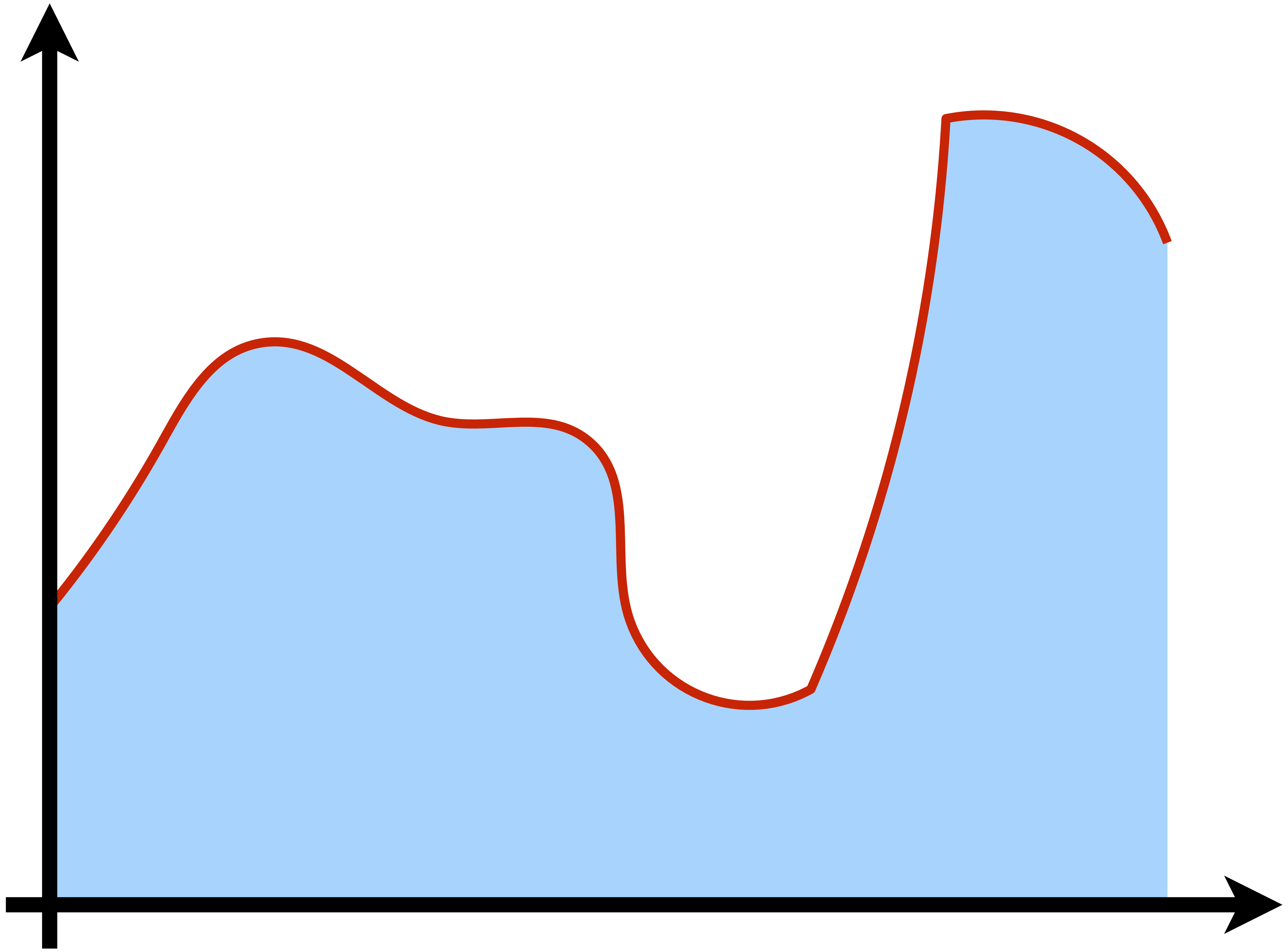
Philipp Slusallek   *Karol Myszkowski*
Gurprit Singh

A la Carte

- Numerical Integration
- Monte Carlo Integration
- Quasi Monte Carlo Integration

Numerical Integration

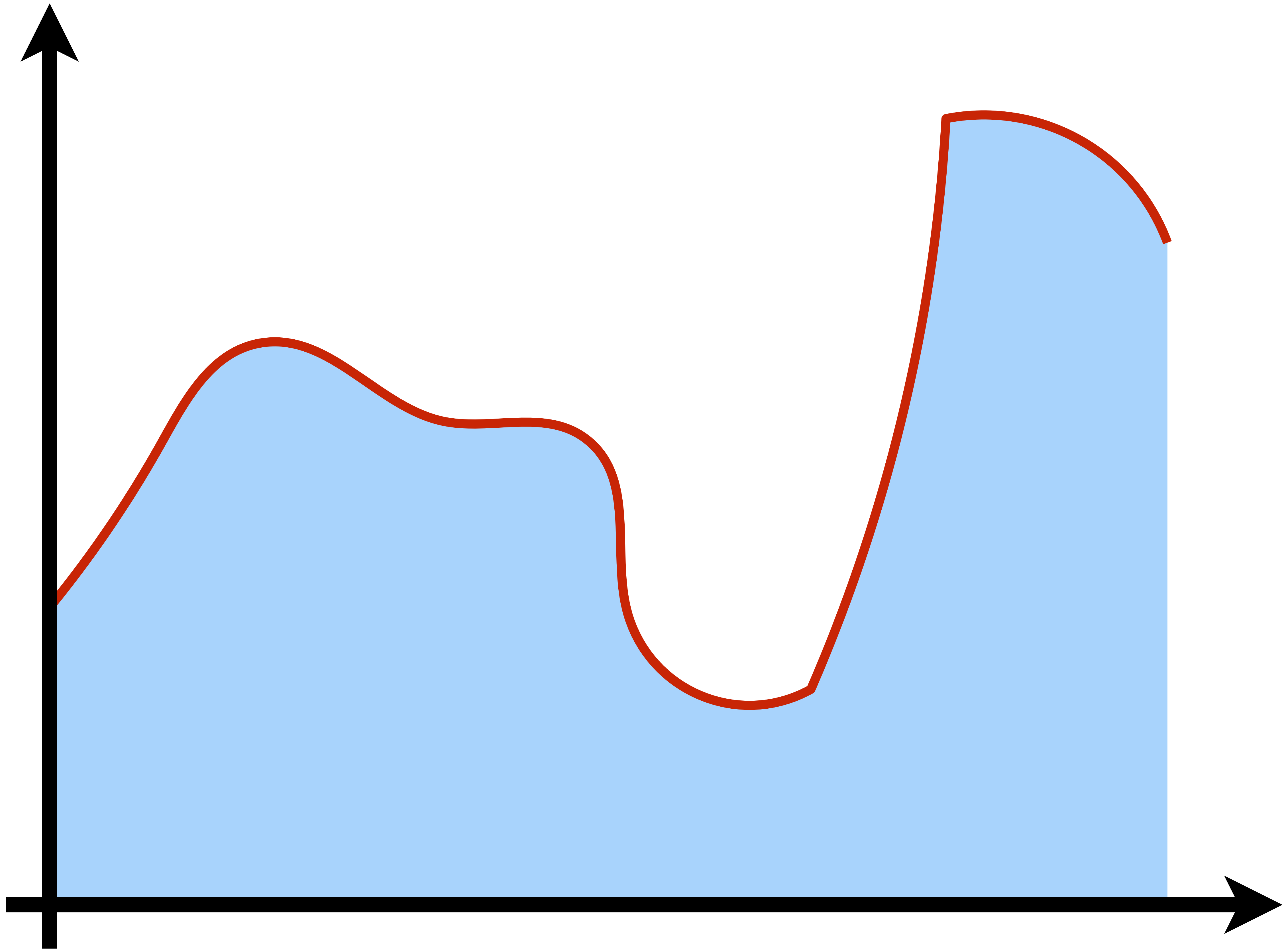
$$\int_a^b f(x)dx$$



Numerical Integration

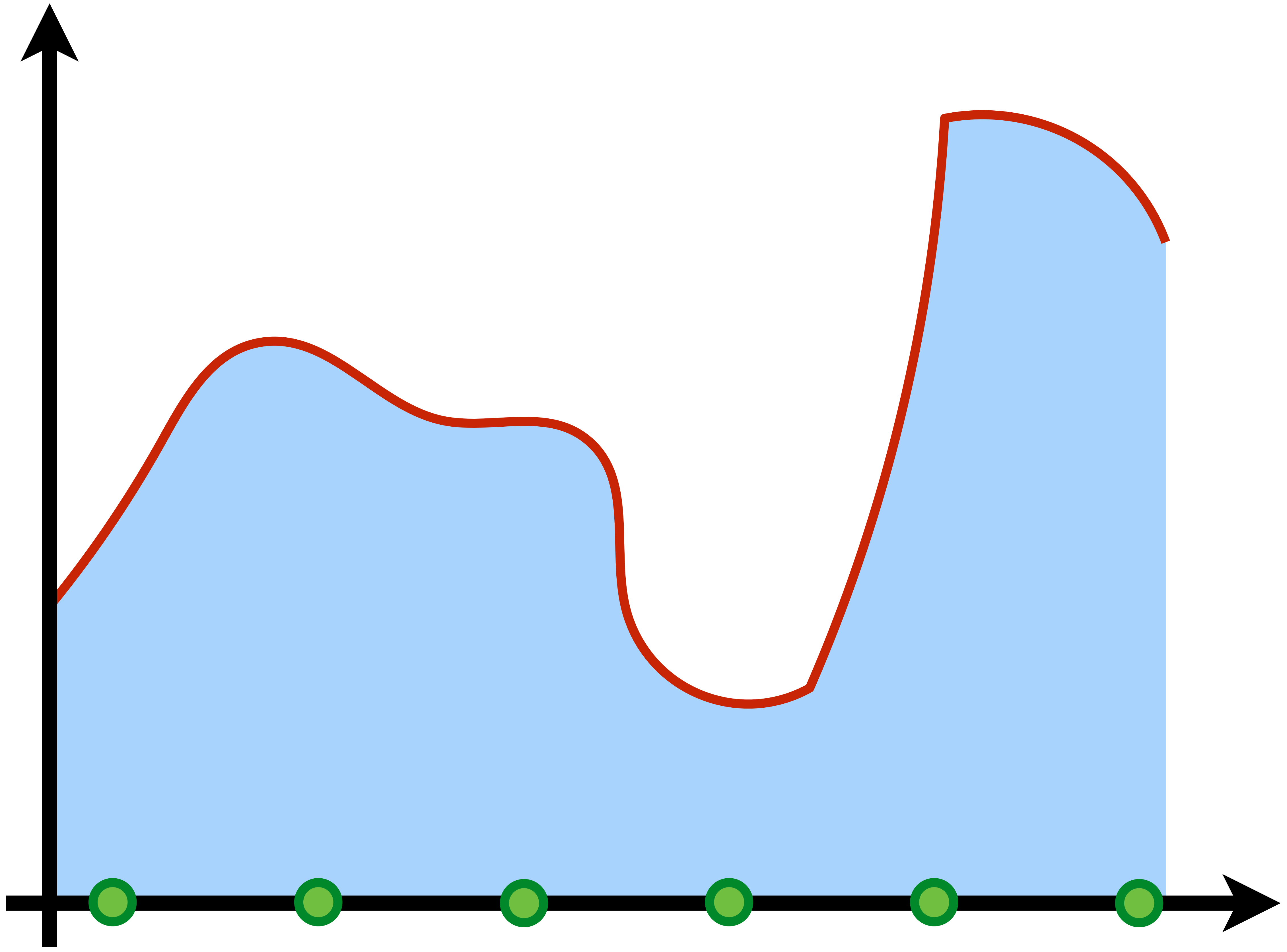
$$\int_a^b f(x)dx$$

- Analytic evaluation: accurate and fast



Numerical Integration

$$\int_a^b f(x)dx$$

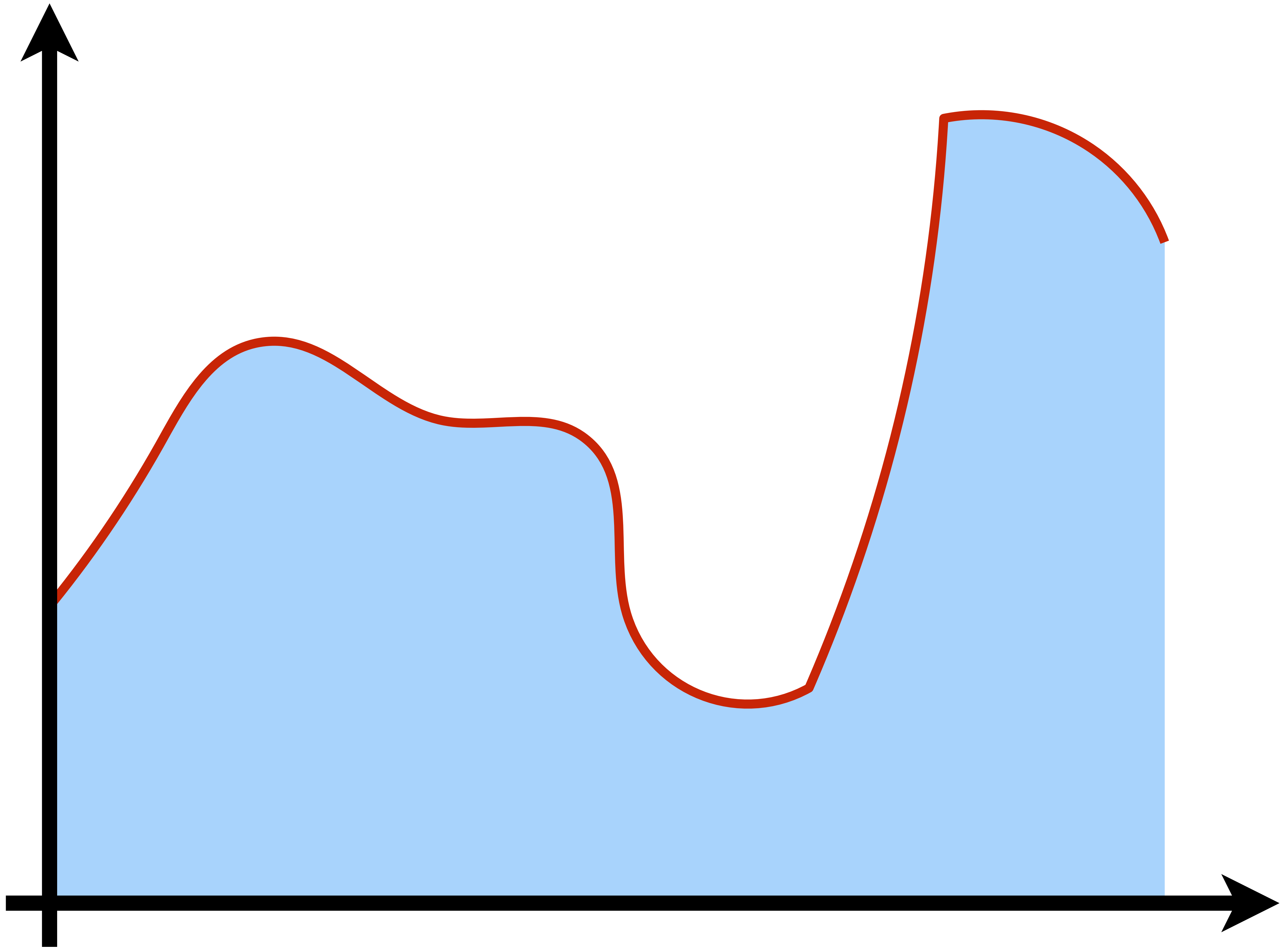


- Numerical evaluations:
  - Provide only approximate solutions,
  - Rate of convergence is important
  - Often involves evaluations only at selected locations

Numerical Integration

$$\int_a^b f(x)dx$$

- Numerical quadrature: designed for 1D integrals
- Cubature/Quadratures: for higher dimensions



Numerical Integration

- Hybrid methods: First transform the integral analytically for simpler numerical handling

Numerical Integration

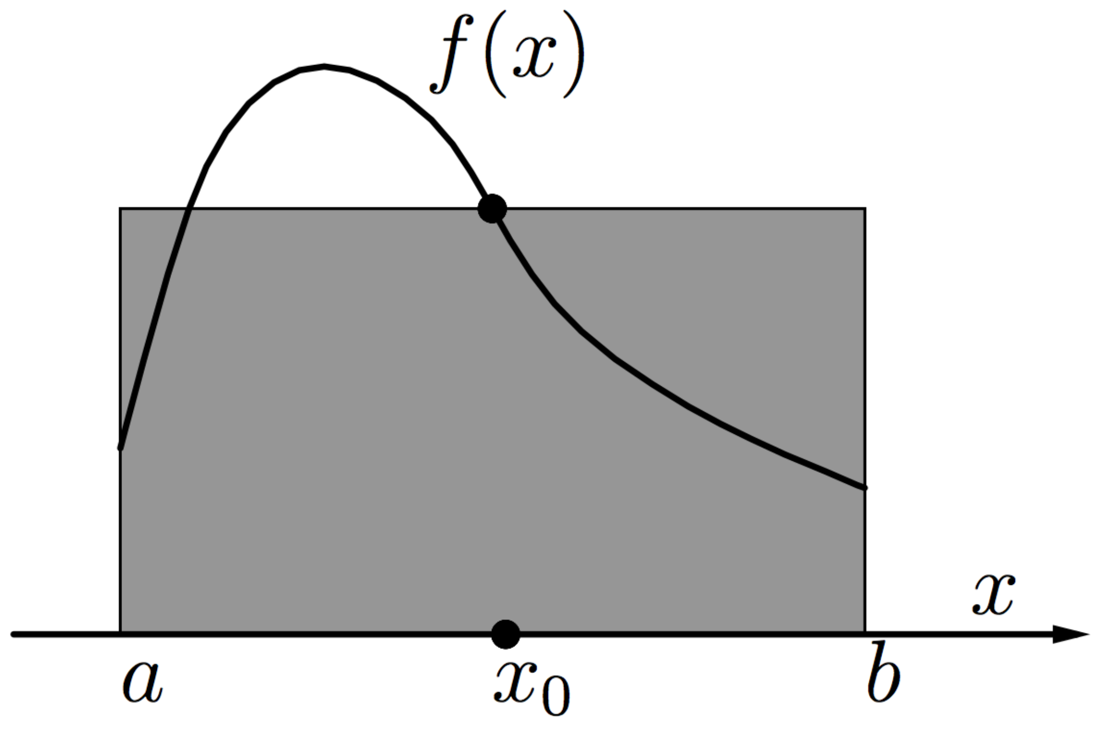
- A number of solutions are developed for the numeric solution of integrals
- Most prominent are the Quadrature rules, where the weights w_i and the sample positions x_i are determined in advance

$$\int_a^b f(x)dx = \sum_{i=1}^{N} w_i f(x_i)$$

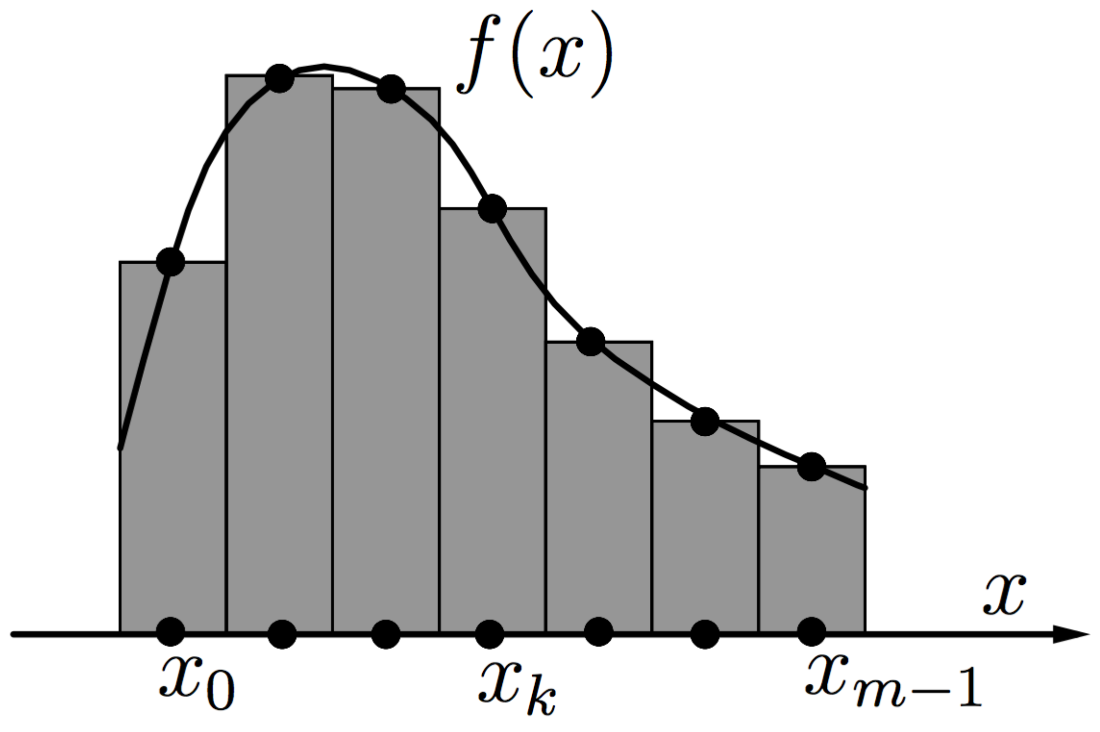
Quadrature rules

- Newton-Cots formula:
  - Midpoint rule (1 sample), Trapezoid rule (2 samples), Simpson rule (3 samples)...

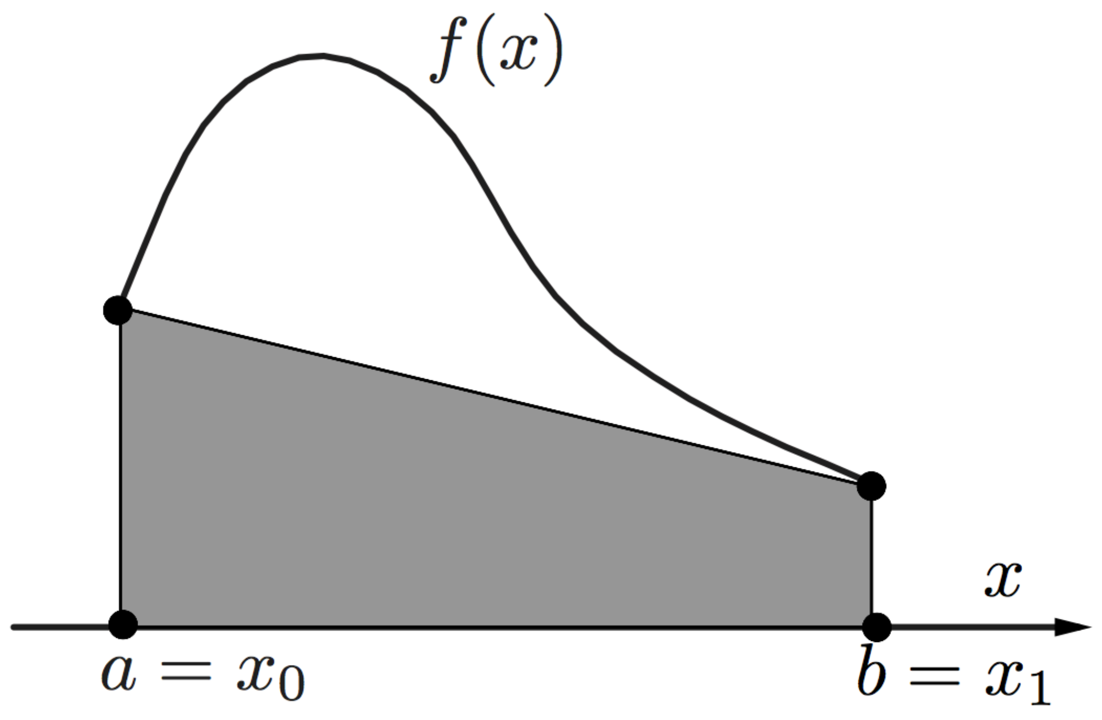
Image courtesy



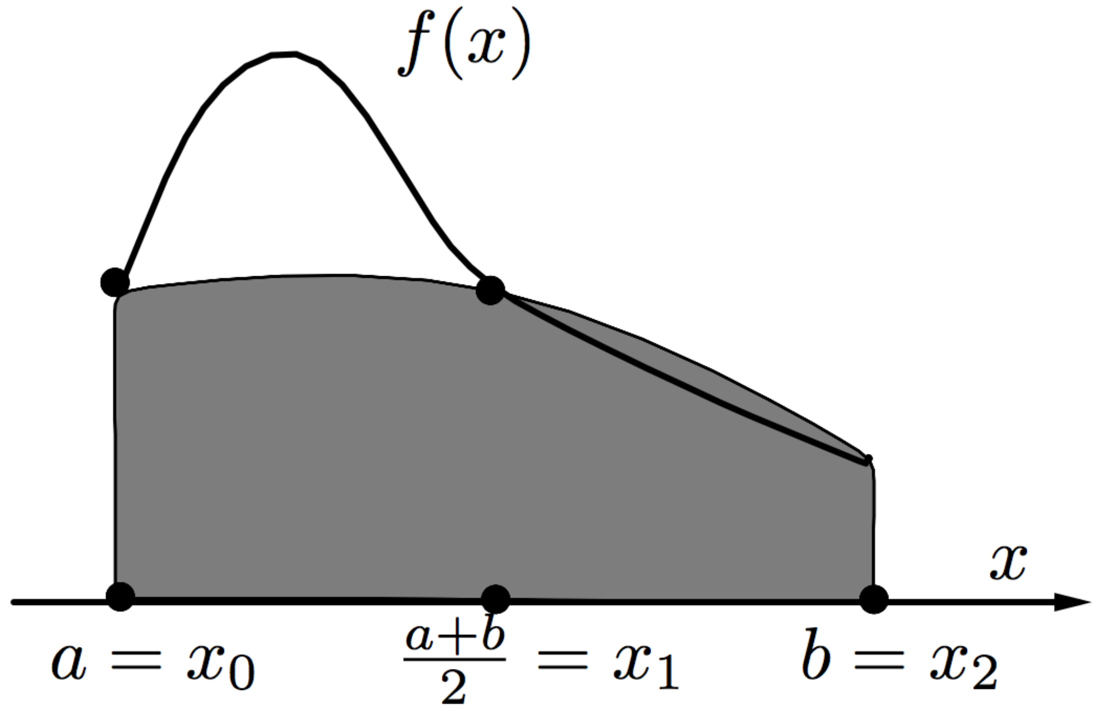
Midpoint formula



Composite midpoint formula



Trapezoidal formula



Cavalieri-Simpson formula

Quadrature rules

- Newton-Cots formula:
    - Midpoint rule (1 sample), Trapezoid rule (2 samples), Simpson rule (3 samples)...
        - Samples are nesting (for powers of 2)
        - Approximates the integral as sum of weighted, equidistant samples

Quadrature rules

- Gauss quadratures:
  - An n-point Gauss quadrature is constructed to yield exact results for polynomials of degree 2n-1 or less.
  - Extends freedom by allowing choice of sample locations
  - It doesn't nest (but nesting alternatives do exist)

Quadrature rules

- Gauss quadratures:
    - An n-point Gauss quadrature is constructed to yield exact results for polynomials of degree 2n-1 or less.
    - Extends freedom by allowing choice of sample locations
    - It doesn't nest (but nesting alternatives do exist)

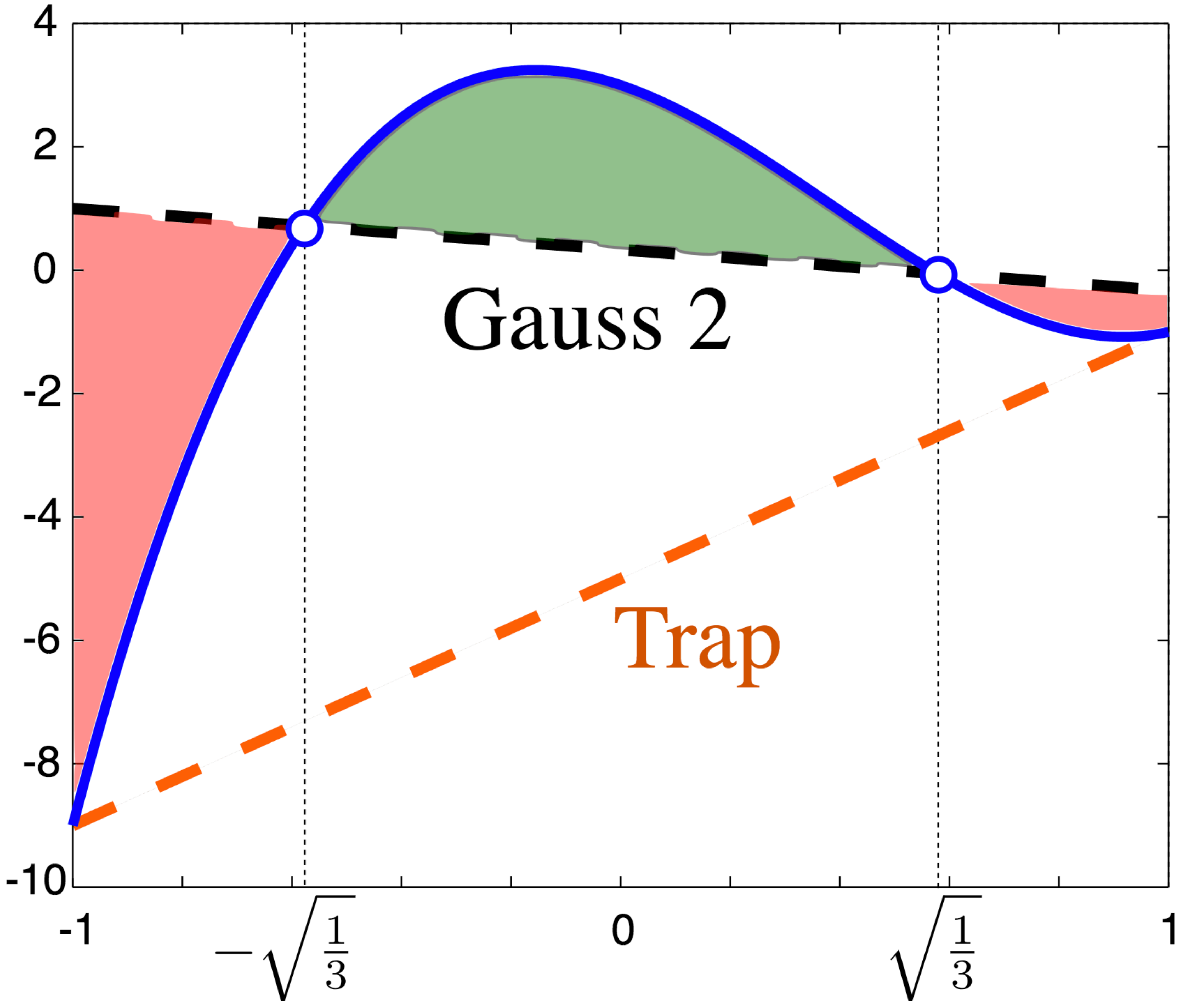


Image from Wikipedia

Quadrature rules

Newton-Cots formula*

Gauss quadratures*

Both approaches achieve convergence of the order $\mathcal{O}(N^{-r})$, given N samples and a smooth integrand that has r-continuous derivatives

***Interested students may refer to this link for more details.**

Numerical Integration: sD case

$$\int_a^b \dots \int_a^b f(x_1, \dots, x_s) dx_1 \dots dx_s = \sum_{i_1=1}^{N} \dots \sum_{i_s=1}^{N} w_{i_1} \dots w_{i_s} f(x_{i_1}, \dots, x_{i_s})$$

- Curse of dimensionality: requires N^s samples for s-dimensional integral
- Convergence drops to $\mathcal{O}(N^{-r/s})$
- Rules must be adapted to non-square domains (typical in rendering)

Monte Carlo Integration

- Independent of the dimensions
- Independent of the underlying topology of the domain
- Variance converges at $O(N^{-1})$ irrespective of the dimensions (N is the sample count)

Integral as Expected Value

$$\int_{Q^s} f(x) d\mu_s(x) = \int_{[0,1)^s} f(x) dx = \int_{[0,1)^s} \frac{f(x)}{p(x)} p(x) dx$$

$p(x)$: is an arbitrary probability density function over the domain

UNIVERSITÄT DES SAARLANDES

Integral as Expected Value

$$\int_{[0,1)^s} f(x)dx = \int_{[0,1)^s} \frac{f(x)}{p(x)} p(x)dx$$

$p(x)$: is an arbitrary probability density function over the domain

Integral as Expected Value

$$\int_{[0,1)^s} f(x)dx = \int_{[0,1)^s} \frac{f(x)}{p(x)} p(x)dx$$

$p(x)$: is an arbitrary probability density function over the domain

$$= \int_{[0,1)^s} \left(\frac{f(x)}{p(x)}\right) p(x)dx$$

$$= E\left[\frac{f(x)}{p(x)}\right]$$

$$E[g(x)] = \int_Q g(x)p(x)dx$$

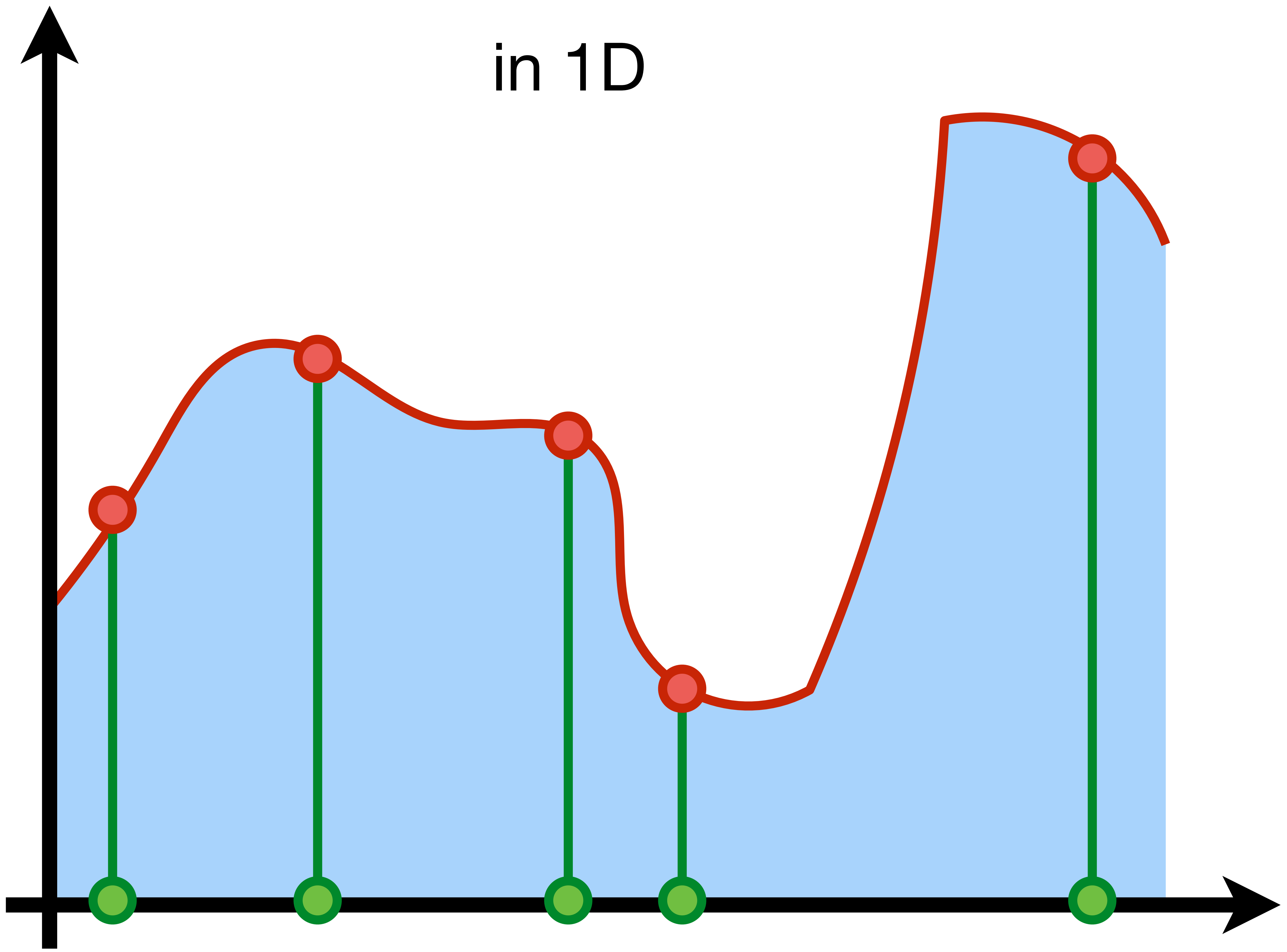
Integral as Expected Value

$$\int_{[0,1)^s} f(x)dx = E\left[\frac{f(x)}{p(x)}\right]$$

We are interested in the numerical computation of this expected value, leading to the highly important concept of **Monte Carlo Estimator**

Monte Carlo Estimator

$$\mathbf{I} = \int_0^1 f(x)dx$$

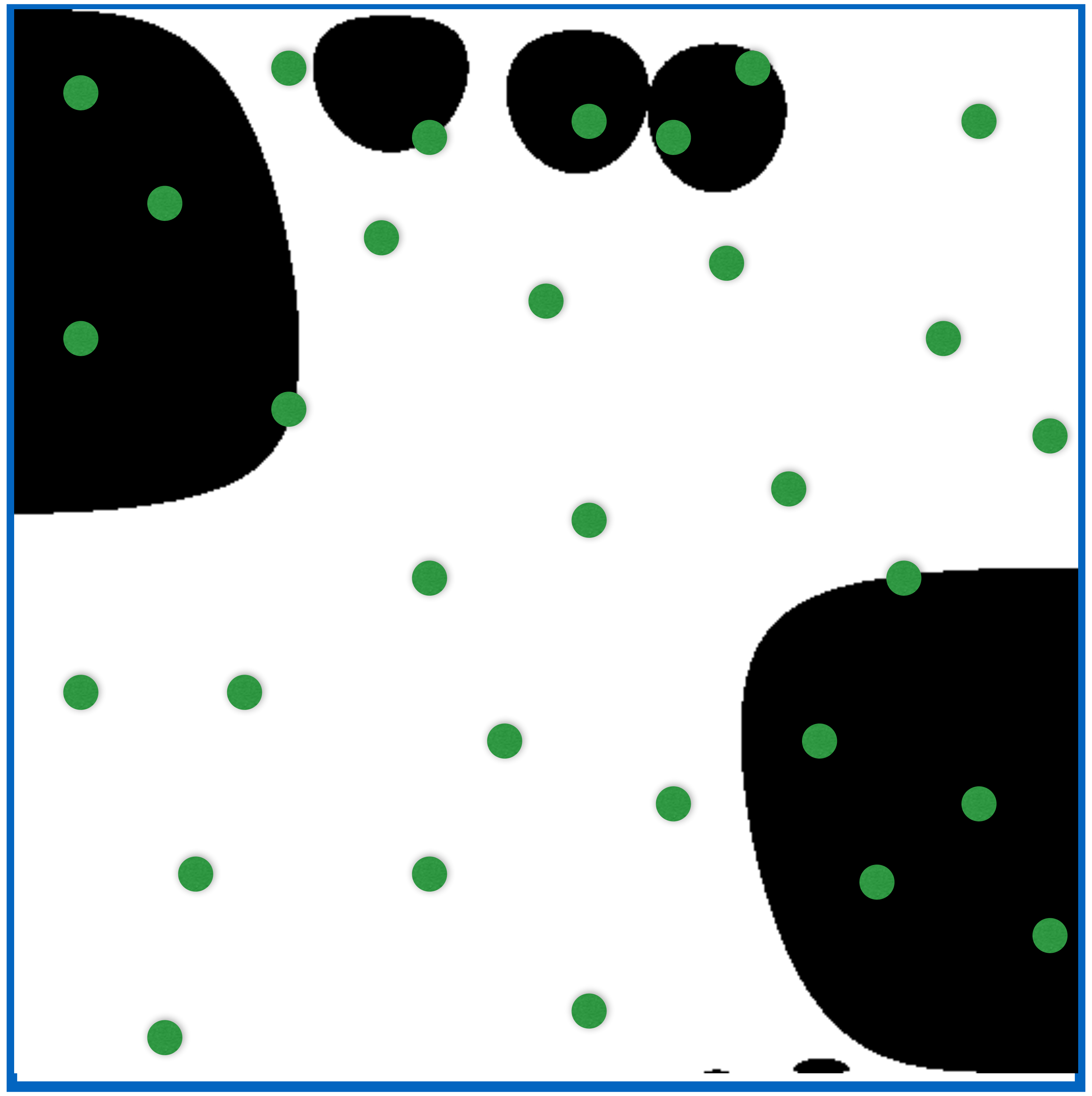


$p(x)$: is the probability density function from which samples are drawn

Monte Carlo Estimator

$$\mathbf{I} = \int_0^1 f(x)dx \approx \frac{1}{N}\sum_{i=1}^{N}\frac{f(x_i)}{p(x_i)}$$

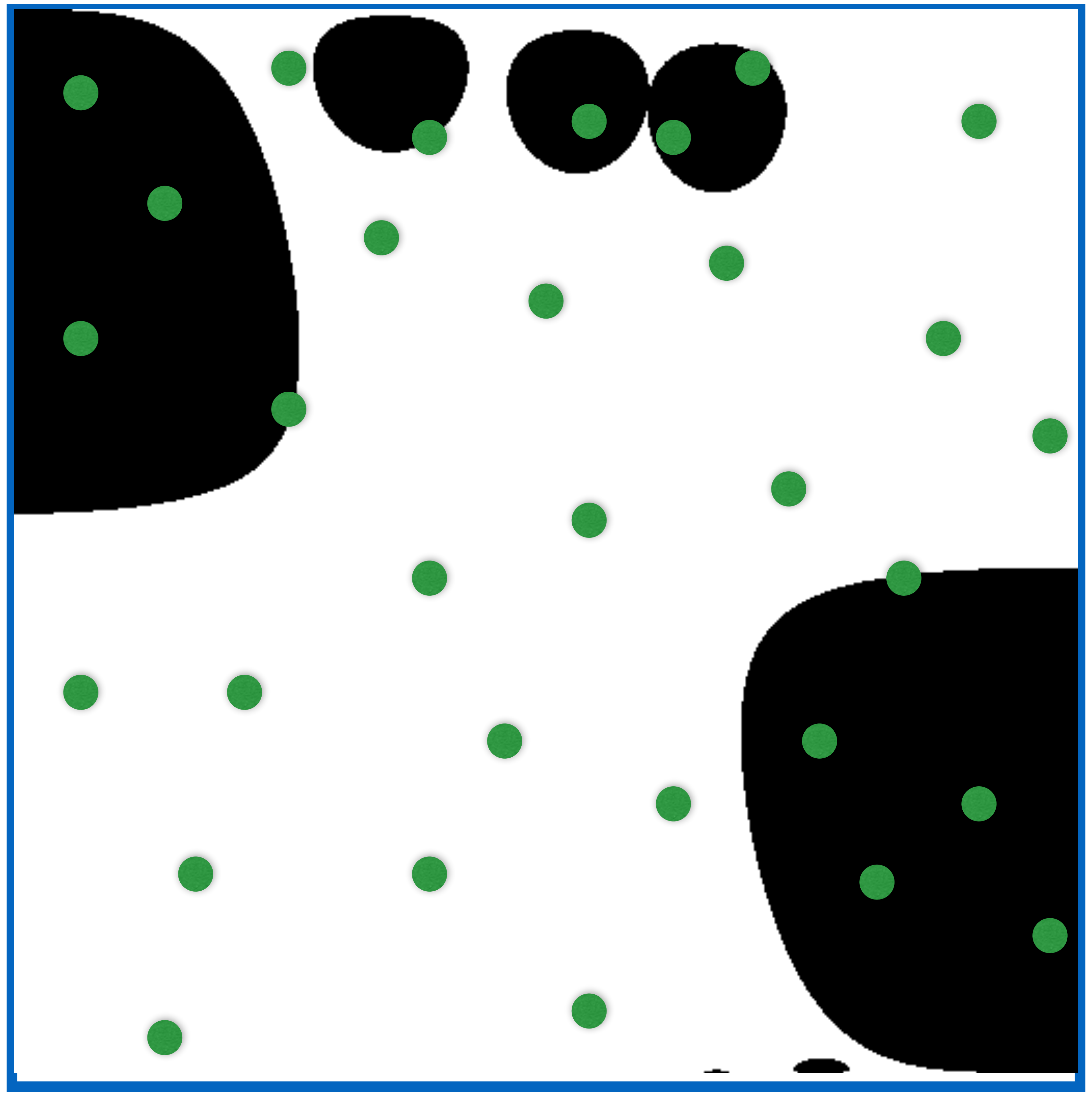
in 2D



$p(x)$: is the probability density function from which samples are drawn

Monte Carlo Estimator

in 2D



$$\mathbf{I} = \int_0^1 f(x)dx \approx \frac{1}{N}\sum_{i=1}^{N}\frac{f(x_i)}{p(x_i)}$$

$p(x)$: is the probability density function from which samples are drawn

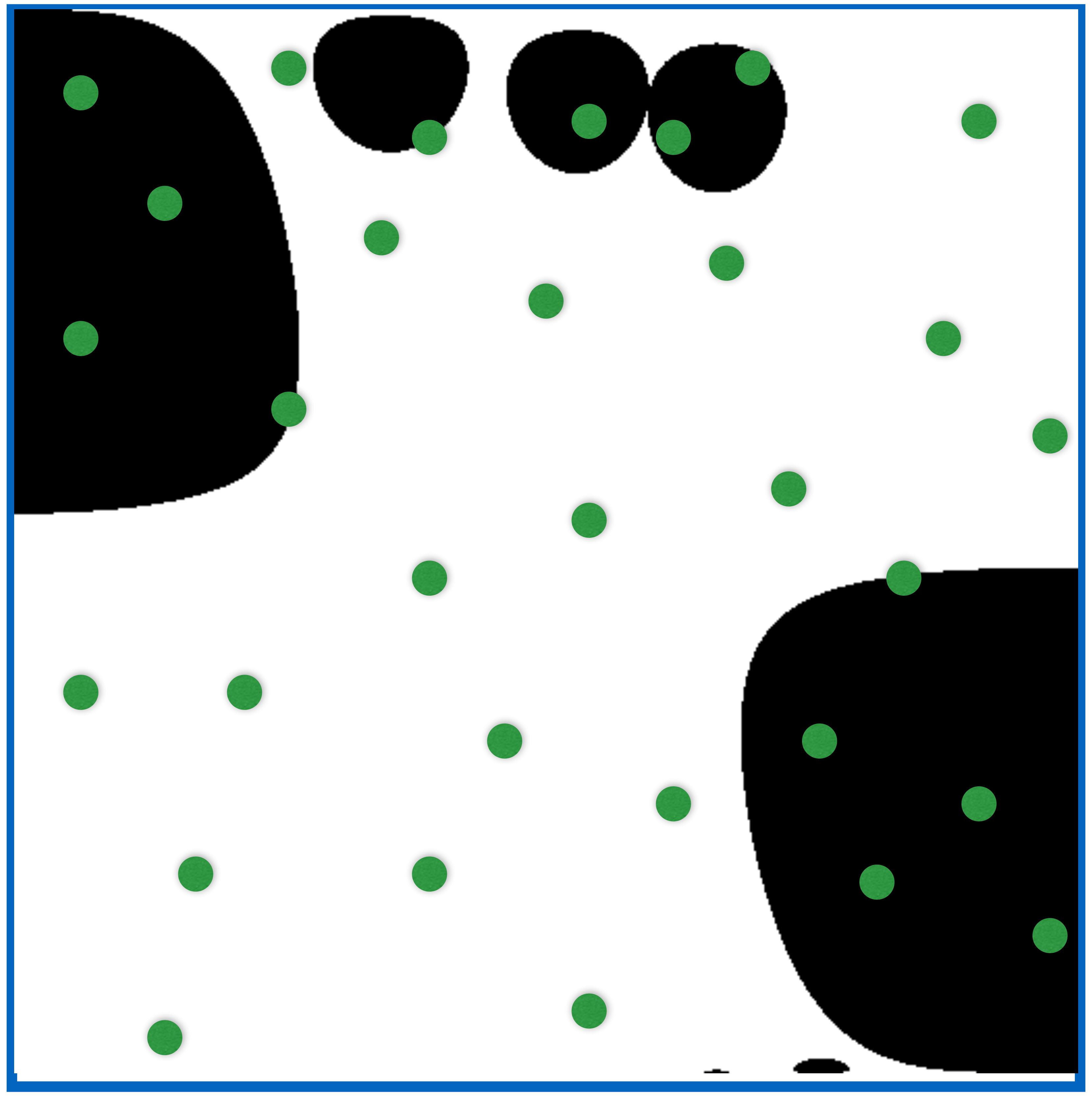
Monte Carlo Estimator

Secondary Estimator: $$\mathbf{I}_N = \frac{1}{N}\sum_{i=1}^{N}\frac{f(x_i)}{p(x_i)}$$

$$\mathbf{I}_N = \frac{1}{N}\sum_{i=1}^{N}\mathbf{I}_1^i$$

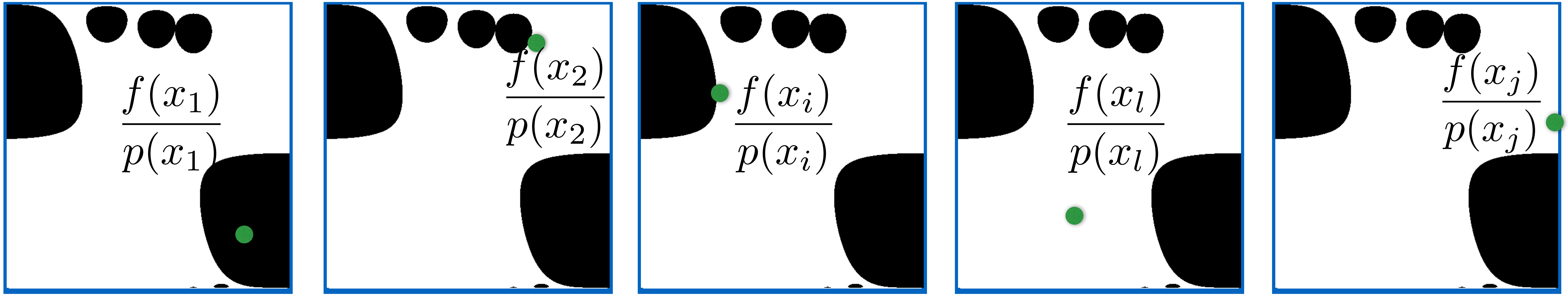
Primary Estimator: $$\mathbf{I}_1^i = \frac{f(x_i)}{p(x_i)}$$

$p(x)$: is the probability density function from which samples are drawn



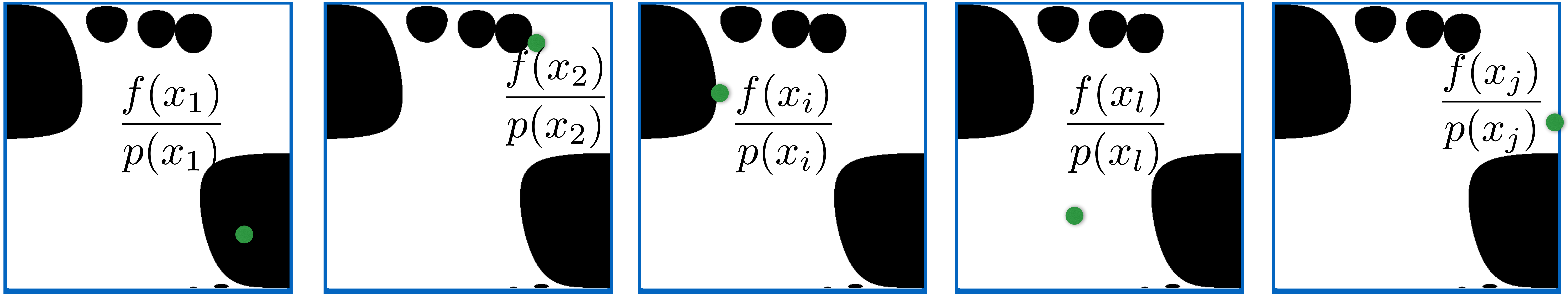
Monte Carlo Estimator

Primary Estimator: $$\mathbf{I}_1^i = \frac{f(x_i)}{p(x_i)}$$



Monte Carlo Estimator

Primary Estimator: $\mathbf{I}_1^i = \frac{f(x_i)}{p(x_i)}$



Monte Carlo Estimator

Primary Estimator: $\mathbf{I}_1^i = \frac{f(x_i)}{p(x_i)}$

$$\boldsymbol{\frac{f(x_1)}{p(x_1)}} \quad \boldsymbol{\frac{f(x_2)}{p(x_2)}}$$

Monte Carlo Estimator

Primary Estimator: $$\mathbf{I}_1^i = \frac{f(x_i)}{p(x_i)}$$

$$\mathbf{I}_N = \frac{1}{N} \sum_{i=1}^{N} \frac{f(x_i)}{p(x_i)}$$

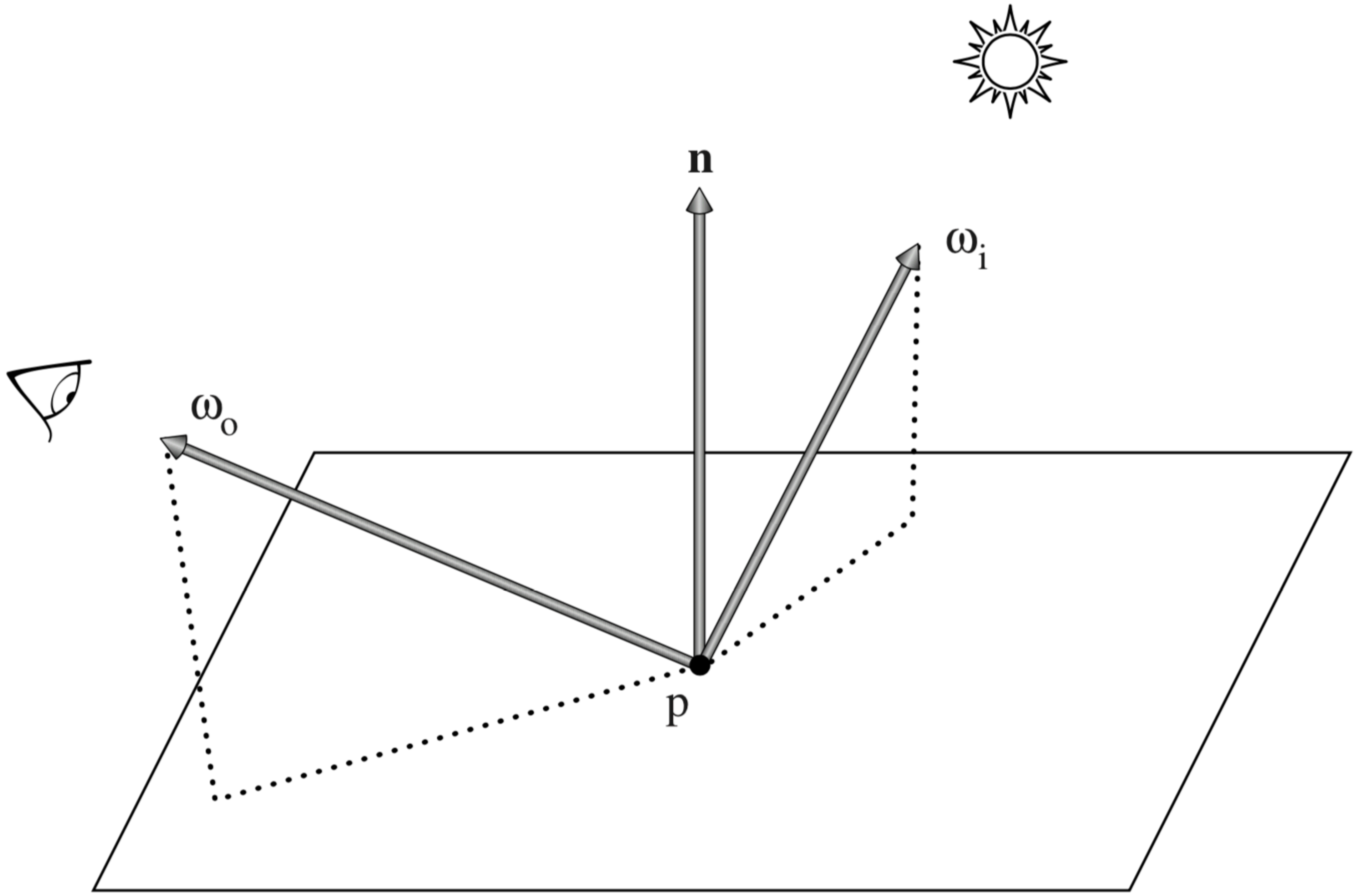
Monte Carlo Estimator

Due to the Strong law of large numbers, the arithmetic mean will converge to the expected value with probability 1 given enough samples:

$$\text{prob}\left\{ \lim_{N\to\infty} \mathbf{I}_N = \frac{1}{N}\sum_{i=1}^{N}\frac{f(x_i)}{p(x_i)} = \mathbf{E}\left[\frac{f(x)}{p(x)}\right] = \int_Q f(x)dx \right\} = 1$$

Rendering Equation

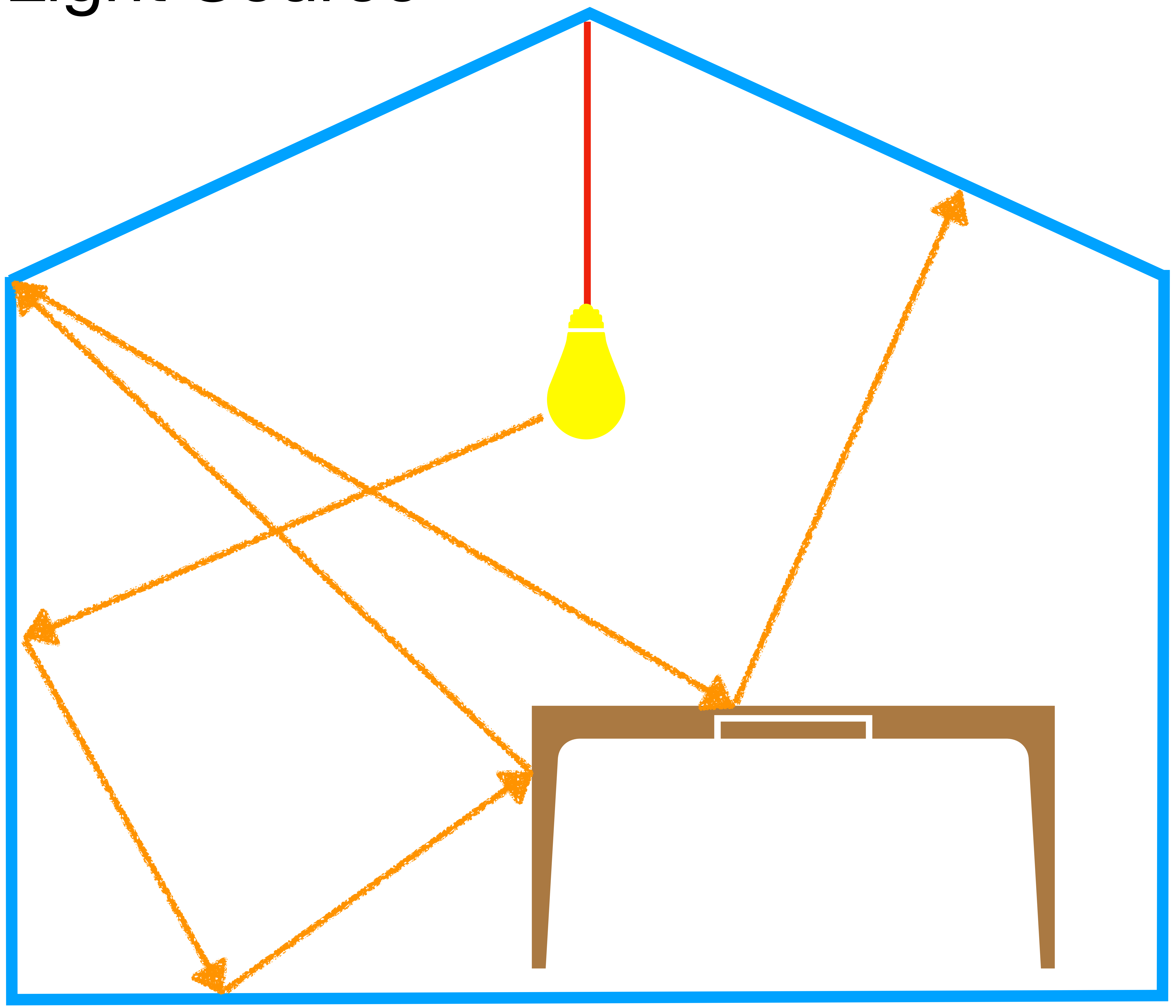
Scattering equation:



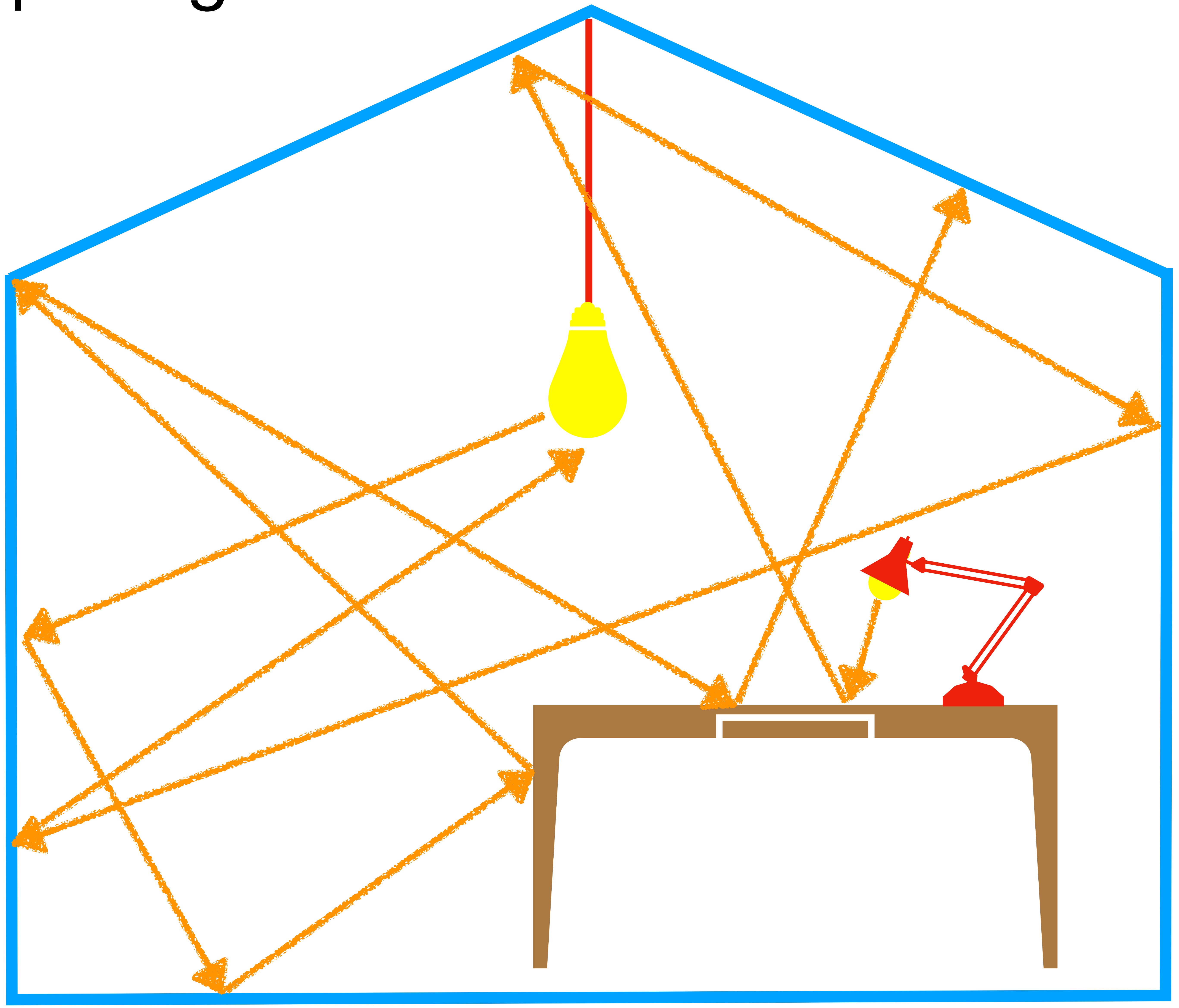
$$L_o(p, \omega_o) = \int_{\mathcal{S}^2} f(p, \omega_o, \omega_i)\, L_i(p, \omega_i)\, |\cos\theta_i|\, d\omega_i$$

Image from PBRT 2016

Global Illumination: One Light Source



Global Illumination: Multiple Light Source



Error in Monte Carlo Estimation

$$\text{Error} = \text{Bias}^2 + \text{Variance}$$

Error in Monte Carlo Estimation

$$\text{Error} = \text{Bias}^2 + \text{Variance}$$

- Monte Carlo estimation is unbiased due to it's "purely" stochastic nature
- We are left with variance, which is visible as stochastic unstructured noise in the rendered images

Error in Monte Carlo Estimation

$$\text{Error} = \text{Bias}^2 + \text{Variance}$$

- For biased techniques, it is important to have a consistent solution
    - This implies, the bias goes to zero with increase in sample count
    - Examples: Progressive photon mapping

Unbiased: Monte Carlo Estimator

$$\text{Error} = \mathbf{I}_N - \mathbf{I}$$

$$\text{Error} = \mathbf{I}_N - \int_Q f(x)dx$$

Unbiased: Monte Carlo Estimator

$$\text{Error} = \mathbf{I}_N - \int_Q f(x)dx$$

Bias by definition is the expected error:

$$\text{Bias} = \mathbf{E}[\text{Error}] = \mathbf{E}\left[\mathbf{I}_N - \int_Q f(x)dx\right]$$

$$\text{Bias} = \mathbf{E}\left[\mathbf{I}_N\right] - \left[\int_Q f(x)dx\right]$$

$$\text{Bias} = \mathbf{E}\left[\mathbf{I}_N\right] - \int_Q f(x)dx$$

Unbiased: Monte Carlo Estimator

$$\text{Bias} = \mathbf{E}\left[\mathbf{I}_N\right] - \int_Q f(x)dx$$

$$\mathbf{E}\left[\mathbf{I}_N\right] = \mathbf{E}\left[\frac{1}{N}\sum_{i=1}^{N}\frac{f(x_i)}{p(x_i)}\right] = \frac{1}{N}\sum_{i=1}^{N}\mathbf{E}\left[\frac{f(x_i)}{p(x_i)}\right] = \frac{1}{N}\sum_{i=1}^{N}\int_Q \frac{f(x)}{p(x)}p(x)dx$$

$$= \frac{1}{N}\sum_{i=1}^{N}\int_Q f(x)dx$$

$$= \int_Q f(x)dx$$

Unbiased: Monte Carlo Estimator

$$\text{Bias} = \mathbf{E}\left[\mathbf{I}_N\right] - \int_Q f(x)dx$$

$$\mathbf{E}\left[\mathbf{I}_N\right] = \int_Q f(x)dx$$

$$\text{Bias} = \mathbf{0}$$

Variance: Monte Carlo Estimator

For the variance of secondary Monte Carlo Estimator, the following holds:

$$\mathrm{Var}(\mathbf{I_N}) = \frac{1}{N^2} \sum_{i=1}^{N} \mathrm{Var}(\mathbf{I}_1^i)$$

Variance: Monte Carlo Estimator

$$\text{Var}(\mathbf{I_N}) = \frac{1}{N^2}\sum_{i=1}^{N}\text{Var}(\mathbf{I}_1^i)$$

$$\text{Var}(\mathbf{I_N}) = \text{Var}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{f(x_i)}{p(x_i)}\right)$$

$$\text{Var}(aX) = a^2\text{Var}(X)$$

Variance: Monte Carlo Estimator

$$\mathrm{Var}(\mathbf{I_N}) = \frac{1}{N^2}\sum_{i=1}^{N}\mathrm{Var}(\mathbf{I}_1^i)$$

$$\mathrm{Var}(\mathbf{I_N}) = \mathrm{Var}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{f(x_i)}{p(x_i)}\right) = \frac{1}{N^2}\mathrm{Var}\left(\sum_{I=1}^{N}\frac{f(x_i)}{p(x_i)}\right)$$

$\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$

$$= \frac{1}{N^2}\sum_{I=1}^{N}\mathrm{Var}\left(\frac{f(x_i)}{p(x_i)}\right)$$

Independent samples

$$= \frac{1}{N^2}\sum_{i=1}^{N}\mathrm{Var}(\mathbf{I}_1^i)$$

Convergence rate: MC Estimator

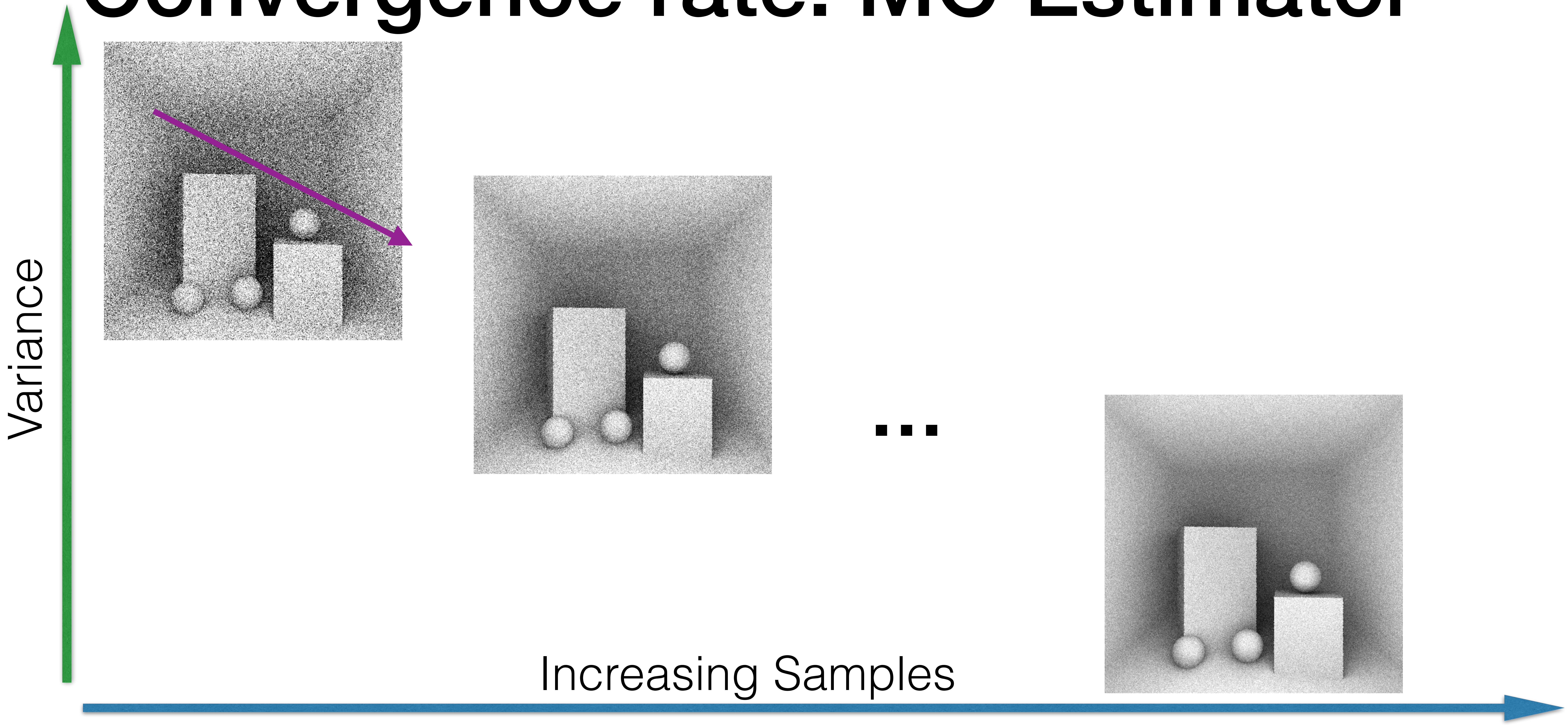
$$\mathrm{Var}(\mathbf{I_N}) = \frac{1}{N^2} \sum_{i=1}^{N} \mathrm{Var}(\mathbf{I}_1^i)$$

$$\mathrm{Error} = \sigma(\mathbf{I}_N) = \frac{1}{\sqrt{N^2}} \sqrt{\mathrm{Var}(\mathbf{I}_1^i)}$$

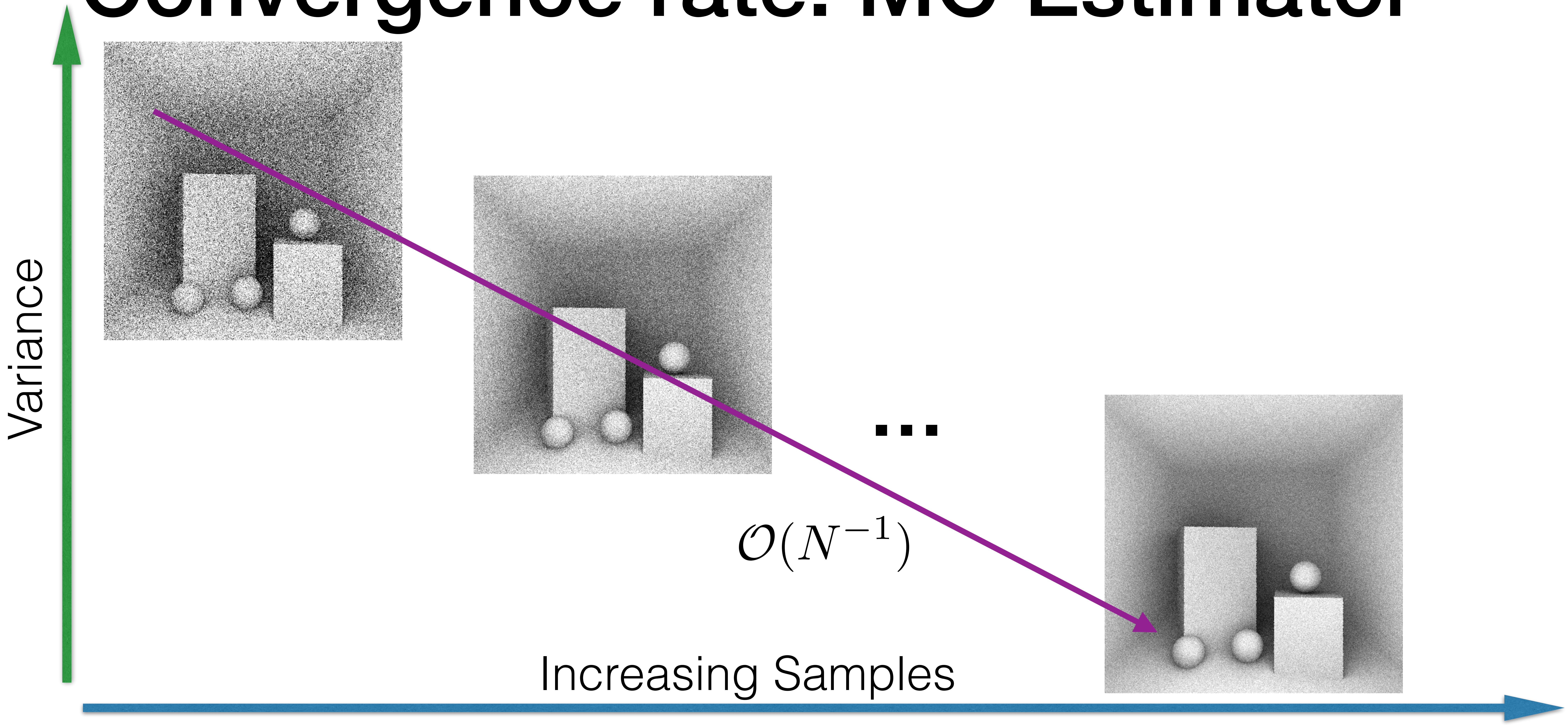
$$= \frac{1}{N} \sigma(\mathbf{I}_1^i)$$

$$\sigma(X) = \sqrt{\mathrm{Var}(X)}$$

Convergence rate: MC Estimator



Convergence rate: MC Estimator



Sampling Methods

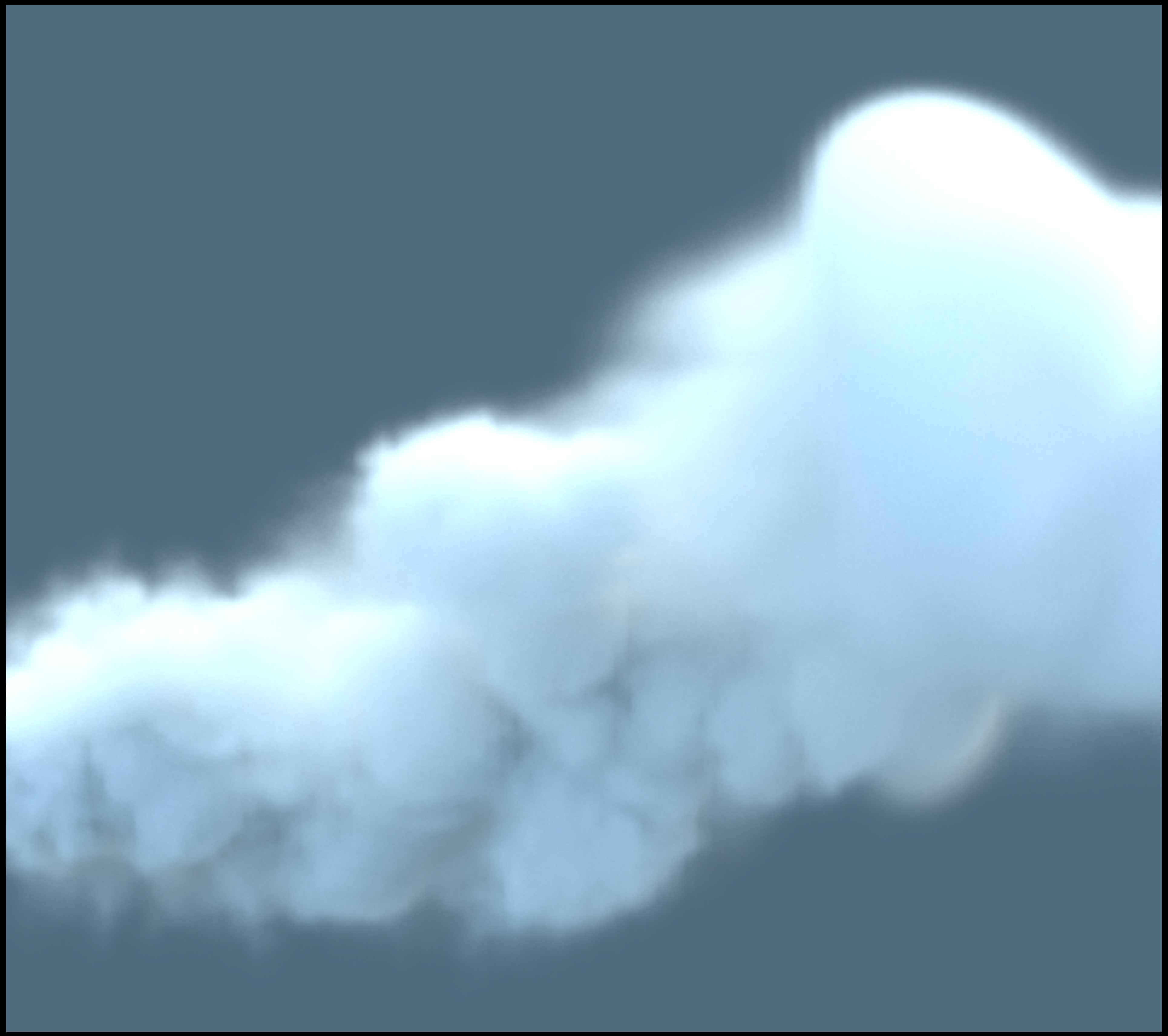
Sampling Methods

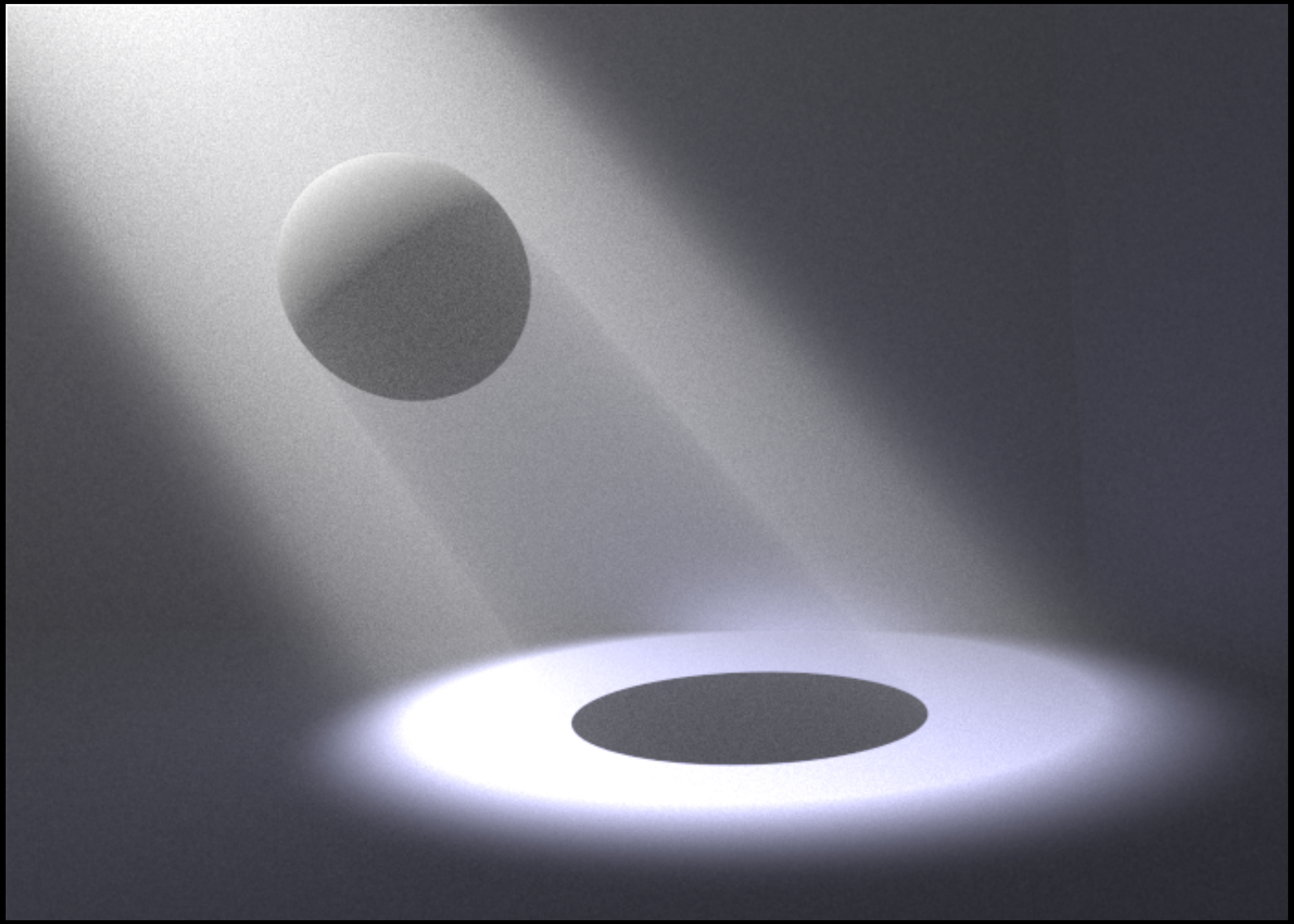
- Inversion methods
- Acceptance-rejection methods
- Metropolis sampling (later)
- Transforming distributions

Inversion Method

- Compute the CDF $P(x) = \int_0^x p(z)dz$
- Compute the inverse CDF $P^{-1}(x)$
- Obtain a uniformly distributed random number $\xi \in [0, 1)$
- Compute $X_i = P^{-1}(\xi)$

Rendering participating media





Inversion Method

$$p(x) \propto e^{-ax}$$

$$p(x) = ce^{-ax}$$

$$\int_0^\infty ce^{-ax} dx = \frac{c}{a} = 1$$

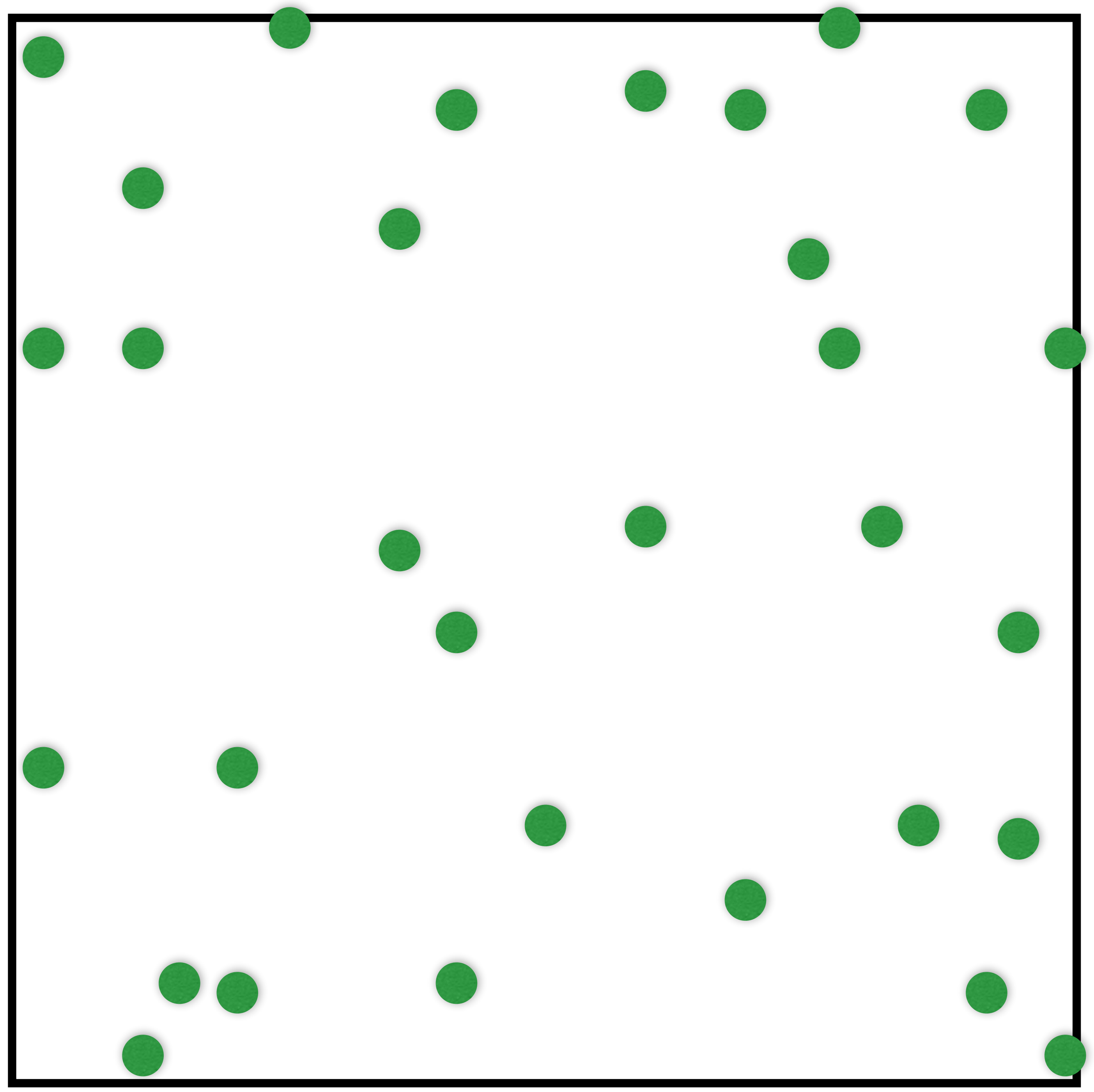
$$P(x) = \int_0^x ce^{-ax} dx = 1 - e^{-ax} = \xi$$

$$P^{-1}(x) = \frac{ln(1-\xi)}{a}$$

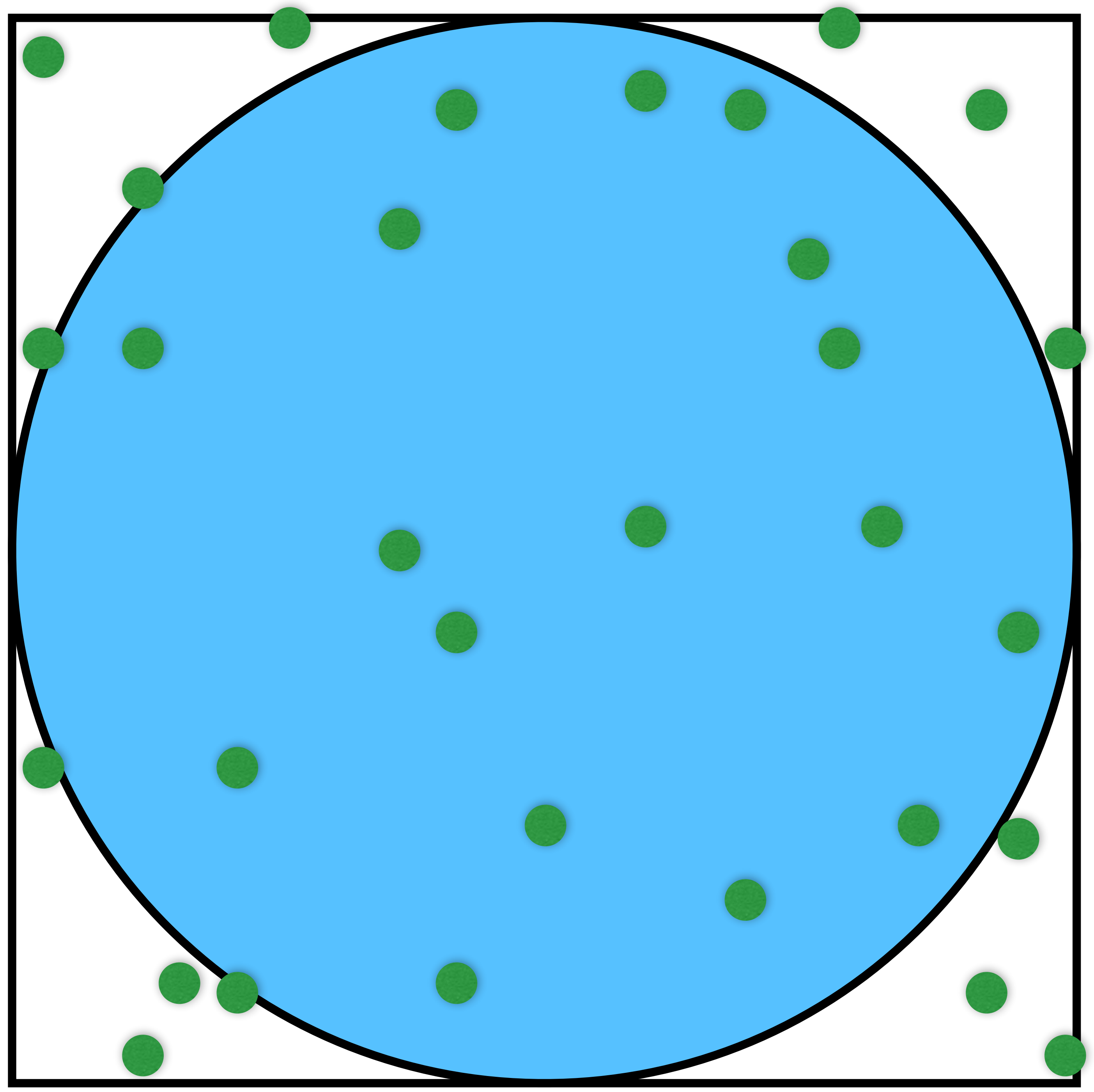
$$P^{-1}(x) = \frac{ln(1-x)}{a}$$

$$P^{-1}(x) = \frac{ln(\xi)}{a}$$

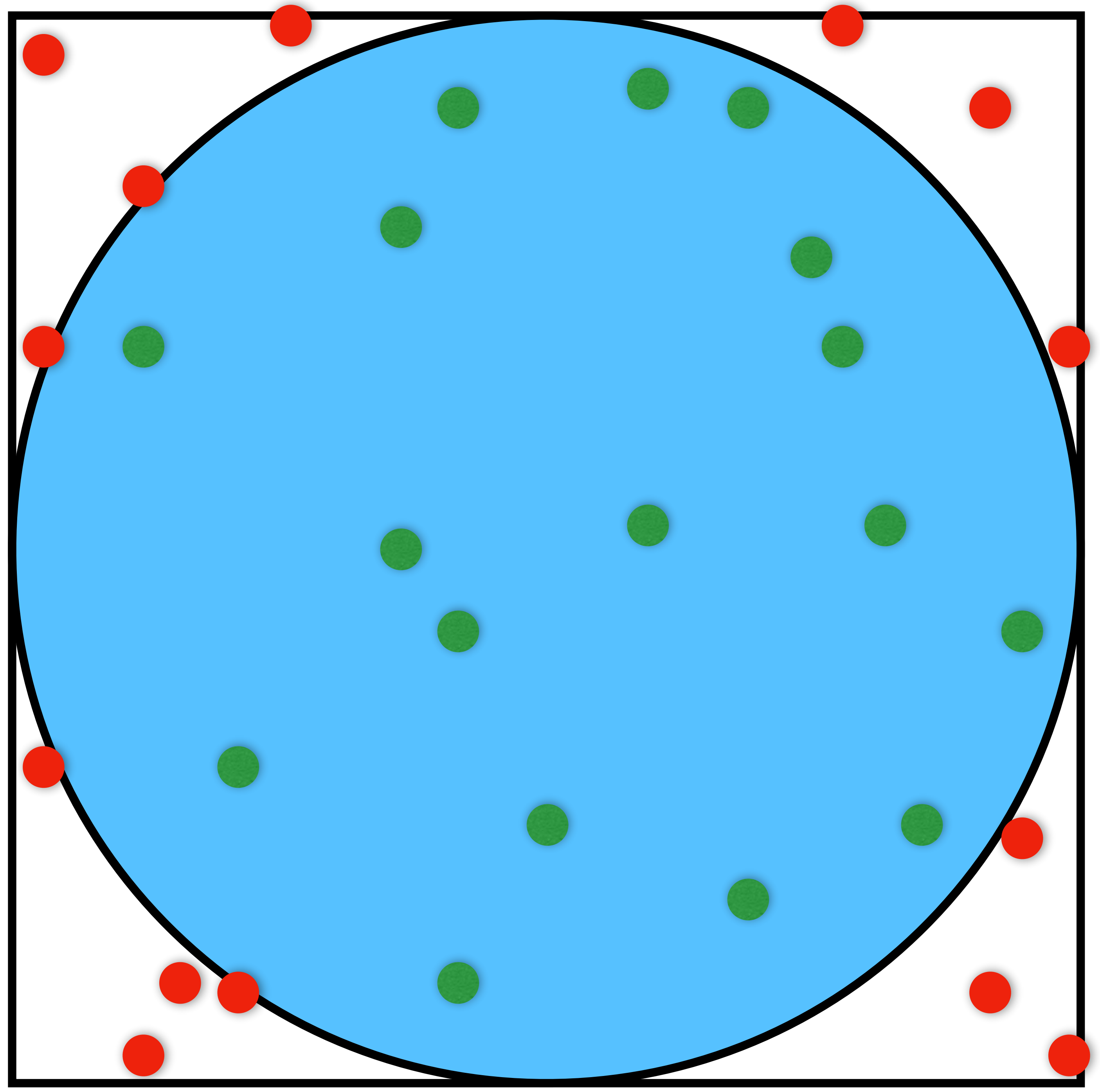
Rejection Sampling Method



Rejection Sampling Method



Rejection Sampling Method



- Many samples are wasted
- Very costly
- Not possible for arbitrary domains

Transformation Method

- General question: which distributions results when we transform samples from an arbitrary distributions to some other distribution with a function f.

$$X_i \sim p_x(x)$$

$$Y_i = y(X_i)$$

What is the distribution of Y_i ?

Transformation Method

- The function $y(x)$ must be a one-to-one transformation
  - It's derivative must either be strictly > 0 or strictly < 0

$$\text{prob}\{Y \leq y(x)\} = \text{prob}\{X \leq x\}$$

Transformation Method

$$\text{prob}\{Y \leq y(x)\} = \text{prob}\{X \leq x\}$$

$$P_y(y) = P_y(y(x)) = P_x(x)$$

This relationship between CDFs directly leads to the relationship between their PDFs:

$$p_y(y)\frac{dy}{dx} = p_x(x)$$

$$p_y(y) = \left(\frac{dy}{dx}\right)^{-1} p_x(x)$$

Transformation Method

$$p_y(y) = \left(\frac{dy}{dx}\right)^{-1} p_x(x)$$

In general, the derivative is strictly positive or negative, and the relationship between the densities is:

$$p_y(y) = \left|\frac{dy}{dx}\right|^{-1} p_x(x)$$

Transformation Method

$$p_y(y) = \left|\frac{dy}{dx}\right|^{-1} p_x(x)$$

How can we use this formula ?

$$p_x(x) = 2x \qquad x \in [0, 1]$$

$$Y = \sin X$$

$$\frac{dy}{dx} = \cos x$$

$$p_y(y) = \frac{p_x(x)}{|\cos x|} = \frac{2x}{\cos x} = \frac{2\arcsin y}{\sqrt{1-y^2}}$$

Transformation Method

- Usually we have some PDF that we want to sample from, not a given transformation

- For example, we might have given: $X \sim p_x(x)$ and we would like to compute $Y \sim p_y(y)$

$$P_y(y) = P_x(x) \qquad\qquad y(x) = P_y^{-1}(P_x(x))$$

- This is a generalization of the inversion method.

Transformation in Multiple dimensions

- Suppose we have an s-dimensional X with density function p_X
- Now let $Y = T(X)$ where T is a bijection.

$$p_y(y) = p_y(T(x)) = \frac{p_x(x)}{|J_T(x)|}$$

$$J_T(x) = \begin{pmatrix} \partial T_1/\partial x_1 & \cdots & \partial T_1/\partial x_n \\ \vdots & \ddots & \vdots \\ \partial T_n/\partial x_1 & \cdots & \partial T_n/\partial x_n \end{pmatrix}$$

Polar Coordinates

$$x = r\cos\theta$$
$$y = r\sin\theta$$

Suppose we draw samples from some density $p(r,\theta)$

What is the corresponding density $p(x,y)$?

$$J_T = \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$$

$$p(x,y) = p(r,\theta)/J_T$$

$$p(x,y) = p(r,\theta)/r$$

Spherical Coordinates

$$x = r \sin\theta \cos\phi$$
$$y = r \sin\theta \sin\phi$$
$$z = r \cos\theta,$$

$$|J_T| = r^2 \sin\theta$$

$$p(r, \theta, \phi) = r^2 \sin\theta \; p(x, y, z)$$

Spherical Coordinates

Spherical coordinates

$$x = r \sin\theta \cos\phi$$
$$y = r \sin\theta \sin\phi$$
$$z = r \cos\theta,$$

$$|J_T| = r^2 \sin\theta$$

$$d\omega = \sin\theta \, \mathrm{d}\theta \, \mathrm{d}\phi$$

$$p(r, \theta, \phi) = r^2 \sin\theta \; p(x, y, z)$$

$$Pr\,\{\omega \in \Omega\} = \int_\Omega p(\omega)\, \mathrm{d}\omega$$

$$p(\theta, \phi)\, \mathrm{d}\theta\, \mathrm{d}\phi = p(\omega)\, \mathrm{d}\omega$$
$$p(\theta, \phi) = \sin\theta \; p(\omega)$$

Uniformly sampling a hemisphere

Here, the task is to choose a direction on the hemisphere uniformly w.r.t. solid angle.
Using the fact that, PDF must integrate to one over its domain:

$$\int_{\mathcal{H}^2} p(\omega)\, \mathrm{d}\omega = 1 \Rightarrow c \int_{\mathcal{H}^2} \mathrm{d}\omega = 1 \Rightarrow c = \frac{1}{2\pi}$$

$$p(\omega) = 1/(2\pi)$$

$$p(\theta, \phi) = \sin\theta/(2\pi)$$

Marginal density function: $p(\theta) = \int_0^{2\pi} p(\theta, \phi)\, \mathrm{d}\phi = \int_0^{2\pi} \frac{\sin\theta}{2\pi}\, \mathrm{d}\phi = \sin\theta$

Conditional density function: $p(\phi|\theta) = \frac{p(\theta, \phi)}{p(\theta)} = \frac{1}{2\pi}$

Uniformly sampling a hemisphere

Corresponding CDFs:

$$P(\theta) = \int_0^\theta \sin\theta' \, \mathrm{d}\theta' = 1 - \cos\theta$$

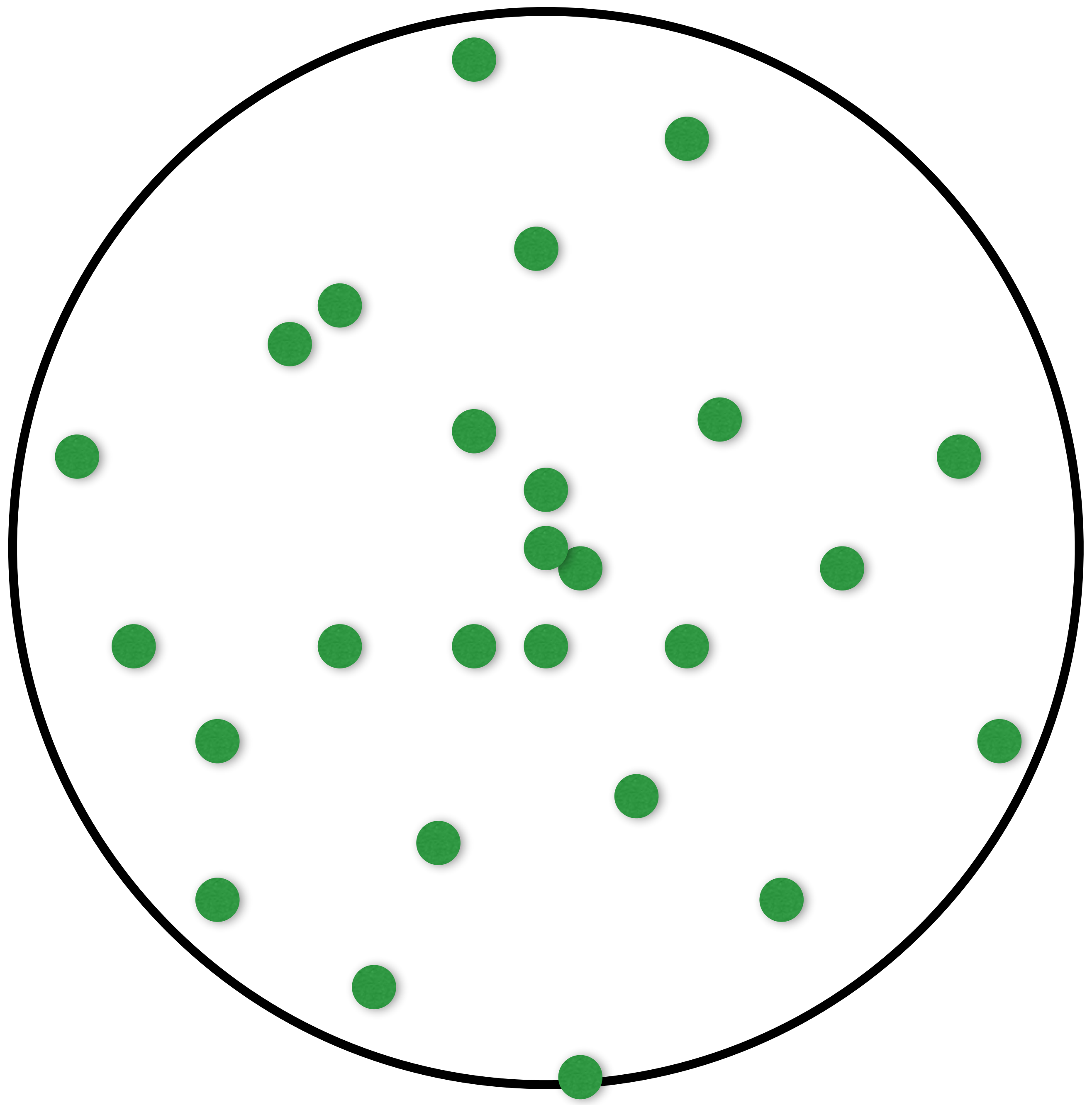
$$P(\phi|\theta) = \int_0^\phi \frac{1}{2\pi} \, \mathrm{d}\phi' = \frac{\phi}{2\pi}.$$

Inverting these functions is straightforward, and here we can safely write:

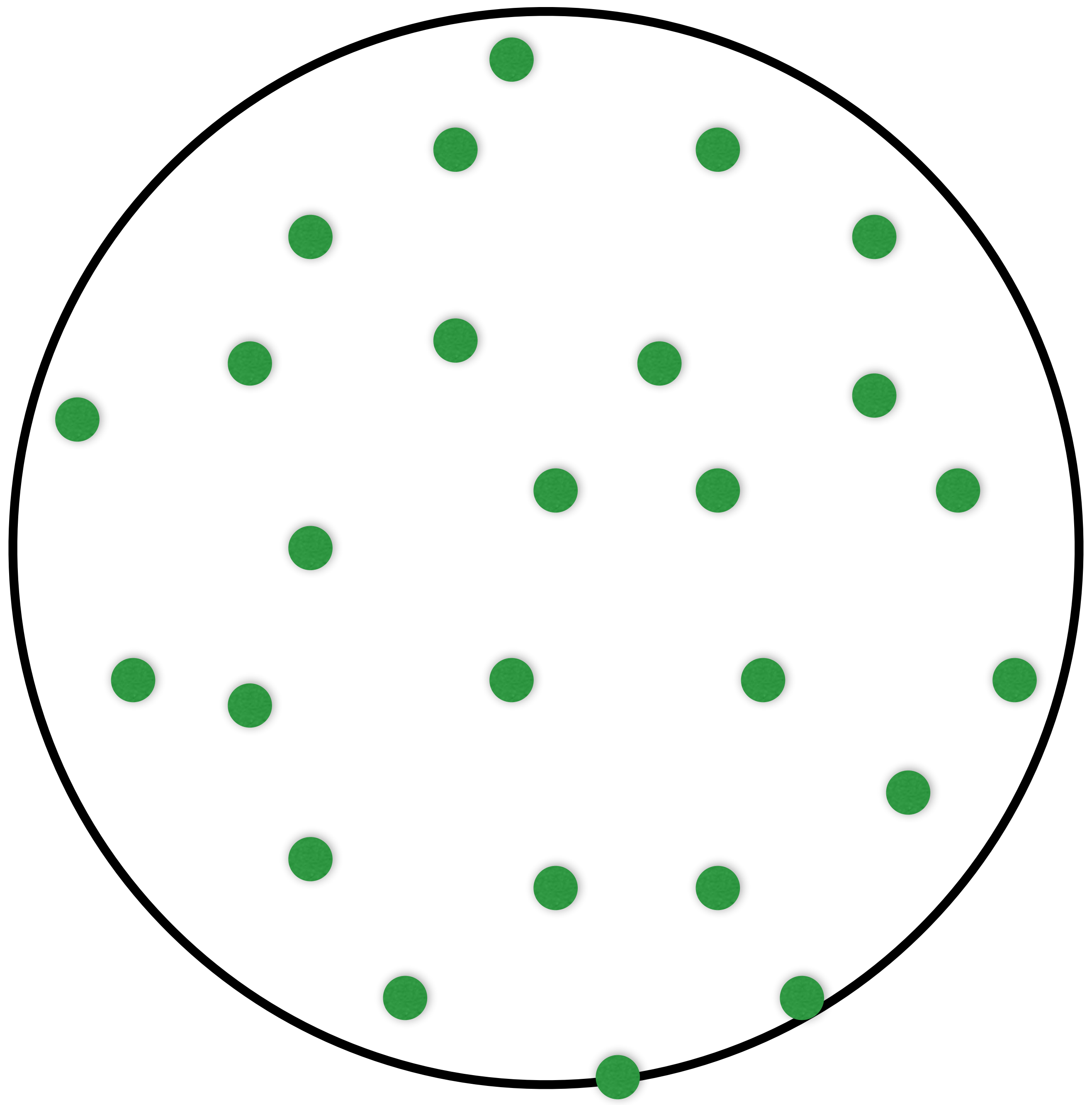
$$\theta = \cos^{-1} \xi_1$$
$$\phi = 2\pi\xi_2.$$

$$x = \sin\theta\cos\phi = \cos(2\pi\xi_2)\sqrt{1-\xi_1^2}$$
$$y = \sin\theta\sin\phi = \sin(2\pi\xi_2)\sqrt{1-\xi_1^2}$$
$$z = \cos\theta = \xi_1.$$

Uniformly sampling a disk



$$r = \xi_1, \theta = 2\pi\xi_2$$



Correct PDF ???

Uniformly sampling a disk

$$p(x, y) = 1/\pi$$

$$p(r, \theta) = r/\pi$$

Marginal density function: $$p(r) = \int_0^{2\pi} p(r, \theta)\, \mathrm{d}\theta = 2r$$

Conditional density function: $$p(\theta|r) = \frac{p(r, \theta)}{p(r)} = \frac{1}{2\pi}.$$

$$r = \sqrt{\xi_1}$$
$$\theta = 2\pi\xi_2.$$

Variance Reduction Techniques

Variance Reduction Techniques

- Importance Sampling
- Multiple Importance Sampling
- Control Variates
- Stratified Sampling

Variance reduction: Importance sampling

$$\mathbf{I}_N = \frac{1}{N} \frac{f(\vec{x}_i)}{p(\vec{x}_i)}$$

- Importance Sampling doesn't always reduce variance.
- The pdf $p(\vec{x})$ must be carefully chosen to gain improvements

Variance reduction: Importance sampling

$$\mathbf{I}_N = \frac{1}{N}\frac{f(\vec{x_i})}{p(\vec{x_i})}$$

$$p(\vec{x}) \propto f(\vec{x})$$

$$p(\vec{x}) = cf(\vec{x})$$

$$c = \frac{1}{\int_{-\infty}^{\infty} f(\vec{x})} d\vec{x}$$

$$\int_{-\infty}^{\infty} p(\vec{x})d\vec{x} = 1$$

$$\int_{-\infty}^{\infty} cf(\vec{x})d\vec{x} = 1$$

this seems like a no-op since the PDF computation requires the integral of the function that we are interested in estimating.

Variance reduction: Importance sampling

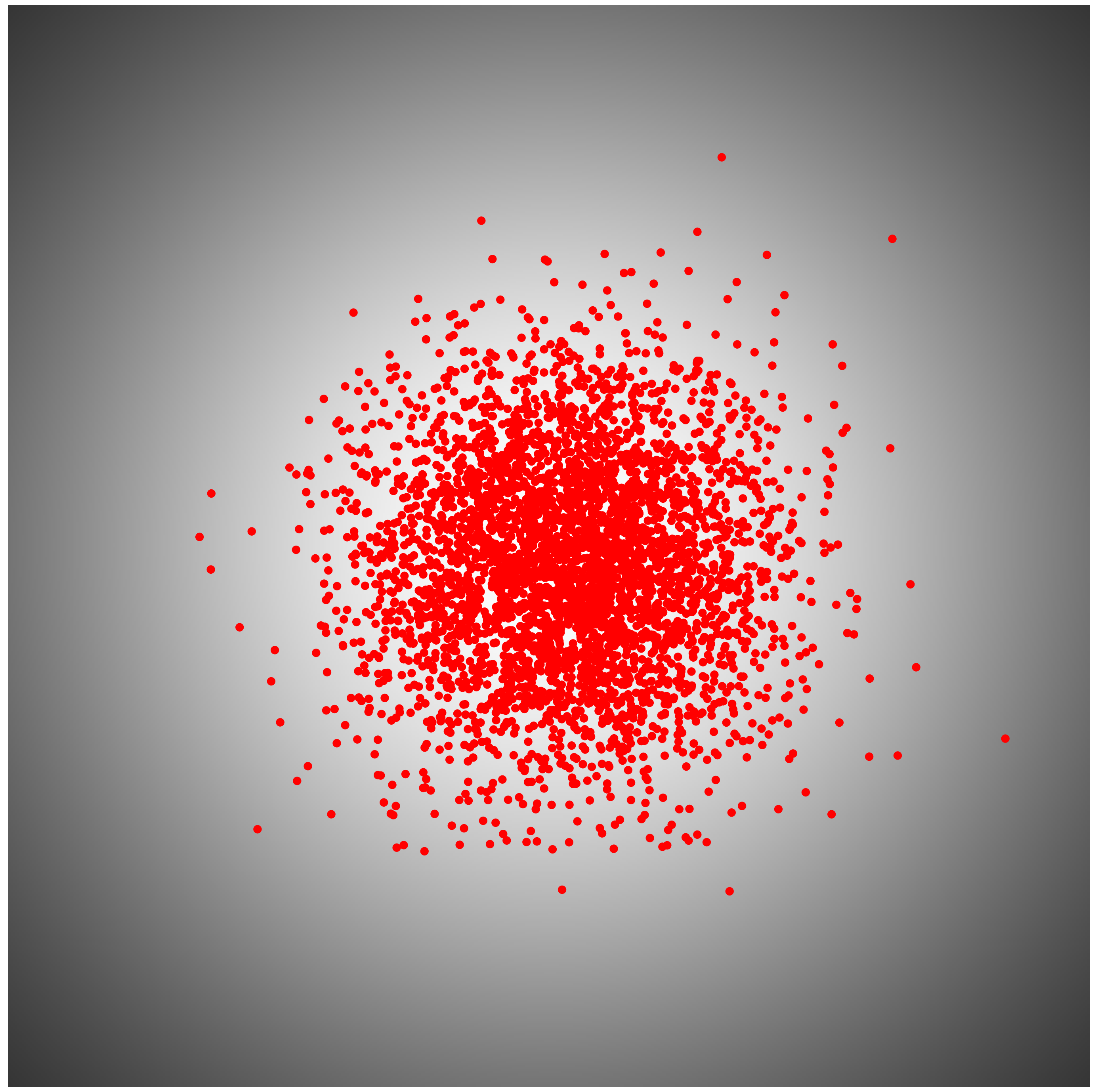
$$\mathbf{I}_N = \frac{1}{N}\frac{f(\vec{x_i})}{p(\vec{x_i})}$$

$$p(\vec{x}) = \frac{f(\vec{x})}{\int_{-\infty}^{\infty} f(\vec{x})} d\vec{x}$$

$$\mathbf{I}_N = \int_{-\infty}^{\infty} f(\vec{x}) d\vec{x}$$

- However, this is a very special case that we are encountering here.
- This is referred to as Perfect Importance Sampling, for which the variance is zero.

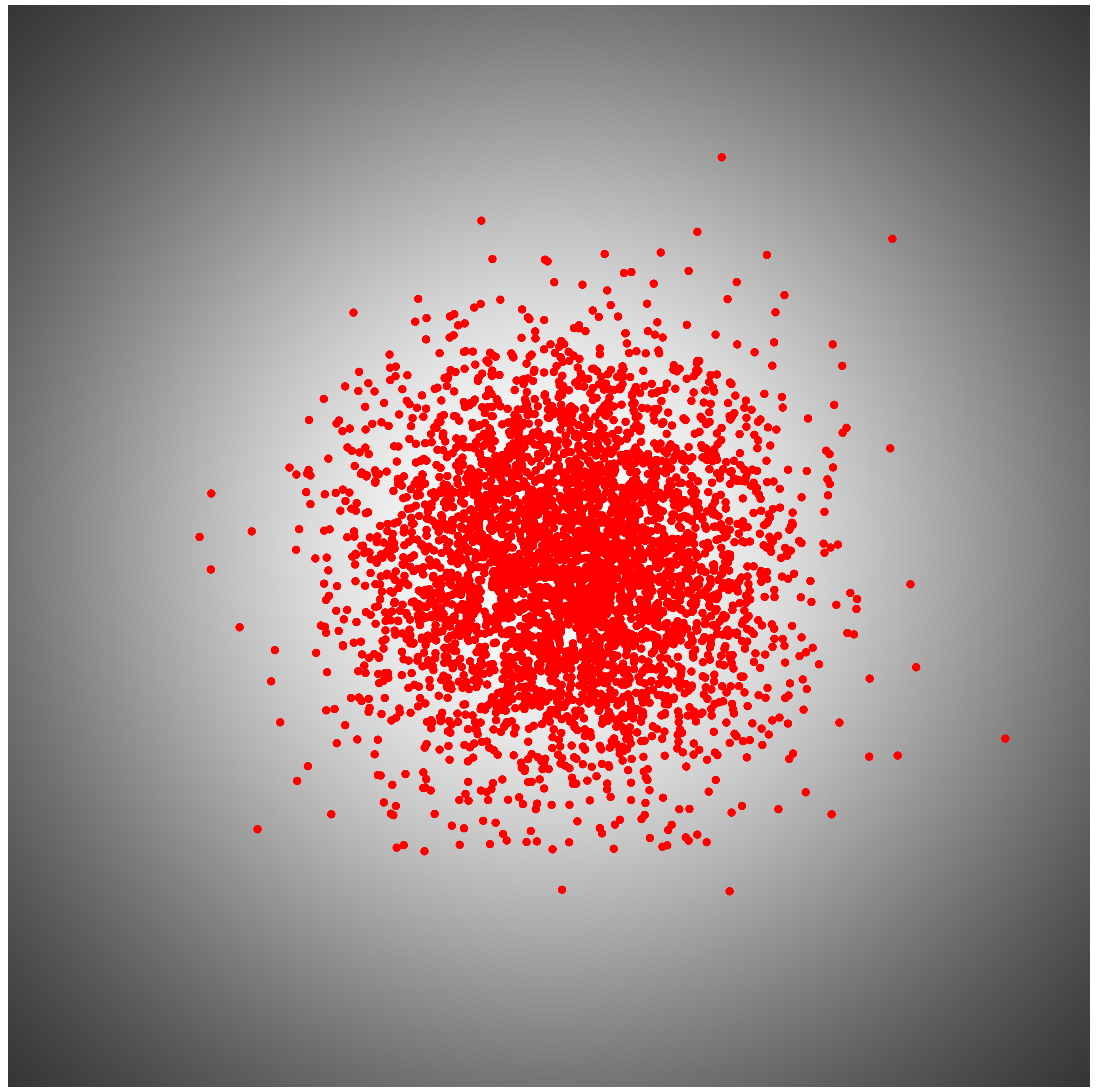
Variance reduction: Importance sampling



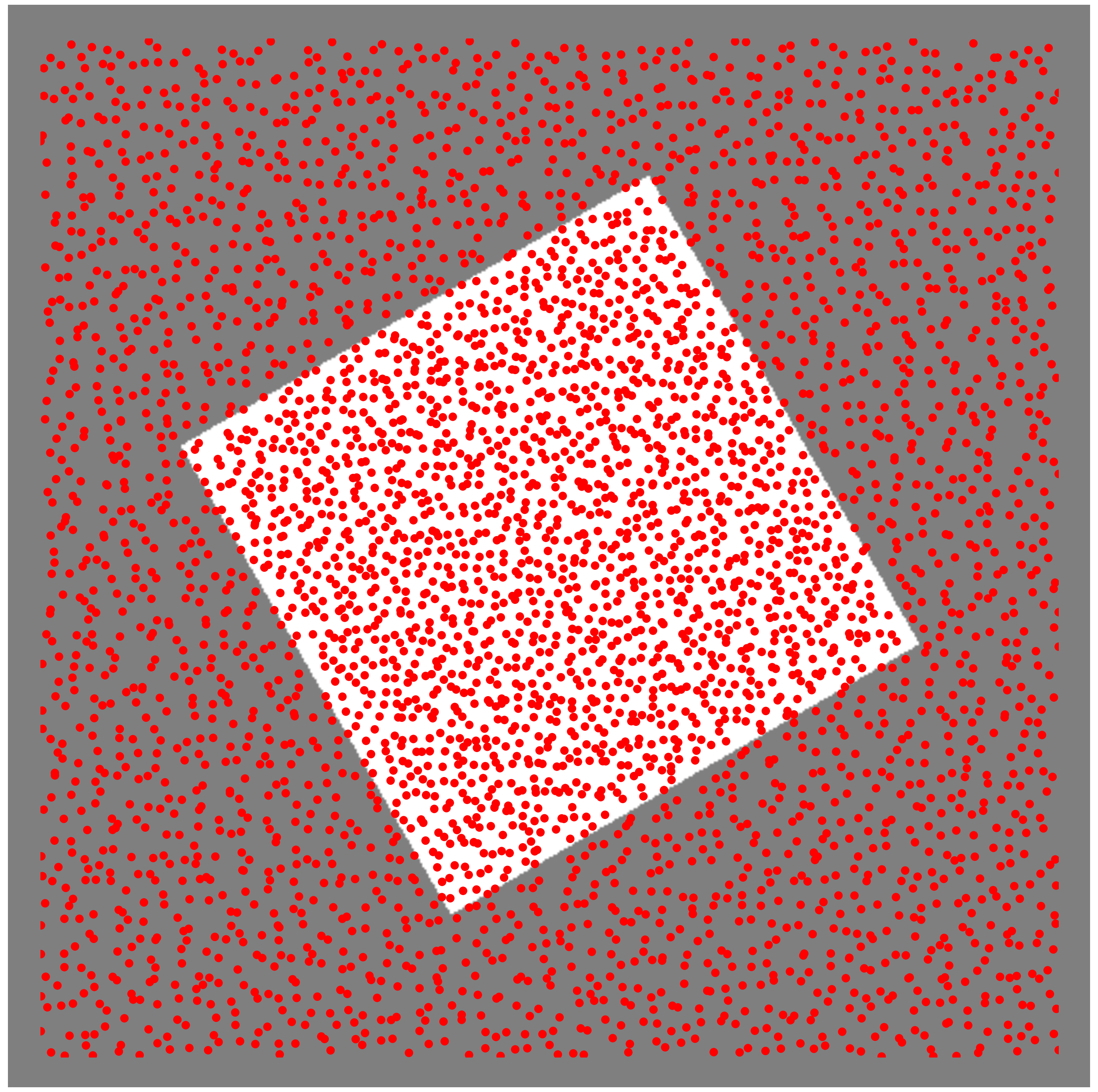
$f(\vec{x})$

Examples of **perfect importance sampling** for which the variance is zero

Variance reduction: Importance sampling



$f(\vec{x})$

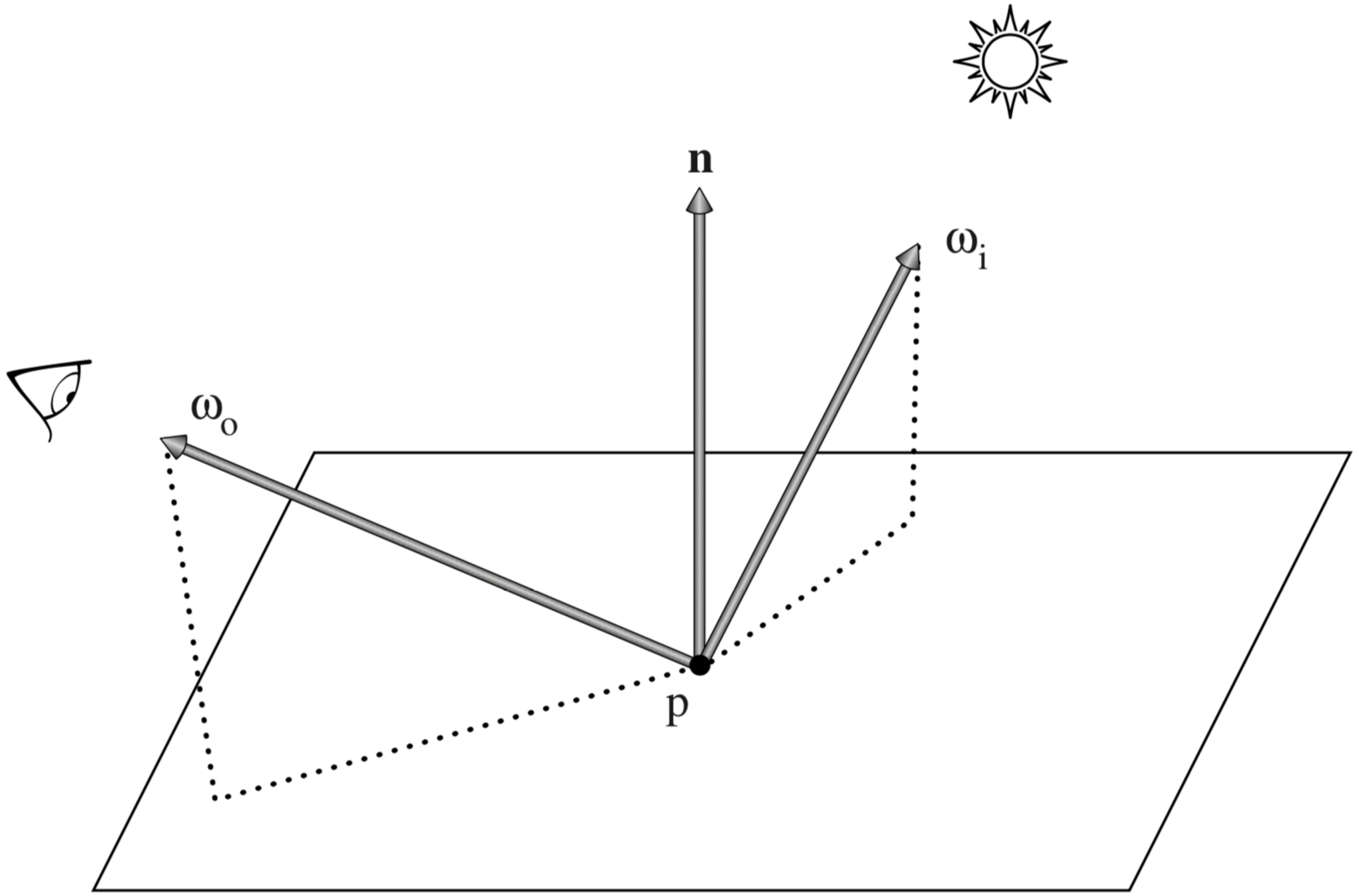


$g(\vec{x})$

Examples of **perfect importance sampling** for which the variance is zero

Variance reduction: Importance sampling

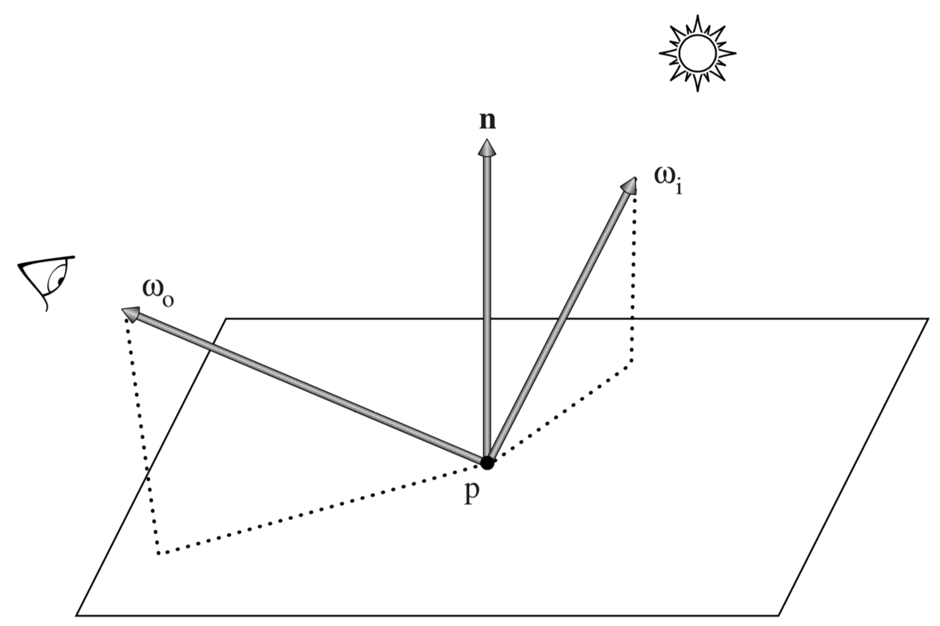
Scattering equation:



$$L_o(p, \omega_o) = \int_{S^2} f(p, \omega_o, \omega_i)\, L_i(p, \omega_i)\, |\cos \theta_i|\, d\omega_i$$

Image from PBRT 2016

Variance reduction: Importance sampling



Scattering equation:

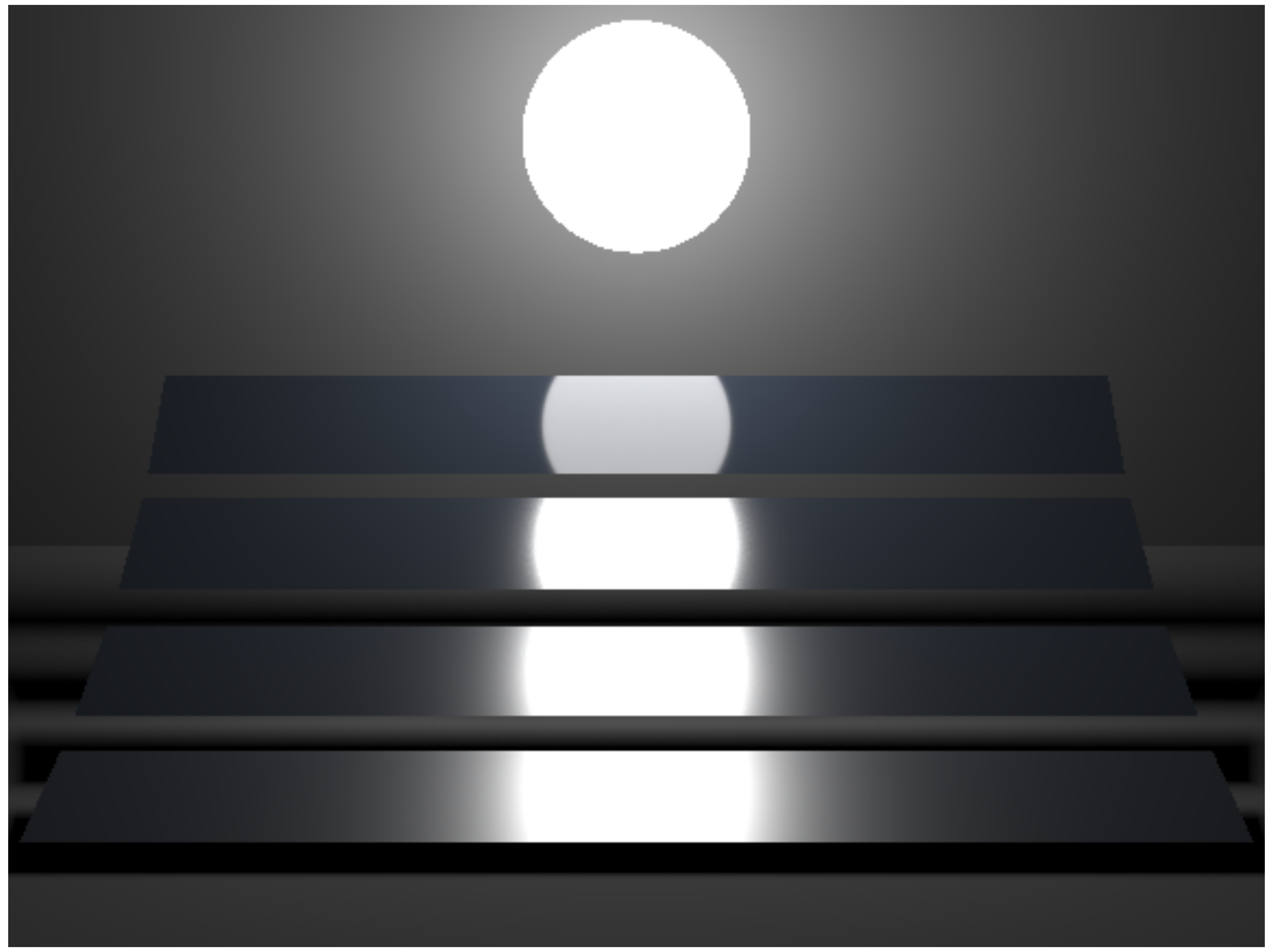
$$L_o(p, \omega_o) = \int_{\mathcal{S}^2} f(p, \omega_o, \omega_i)\, L_i(p, \omega_i)\, |\cos\theta_i|\, d\omega_i$$

$$\approx \frac{1}{N} \sum_{j=1}^{N} \frac{f(p, \omega_o, \omega_j)\, L_i(p, \omega_j)\, |\cos\theta_j|}{p(\omega_j)}$$

$$p(\omega) \propto \cos\theta$$

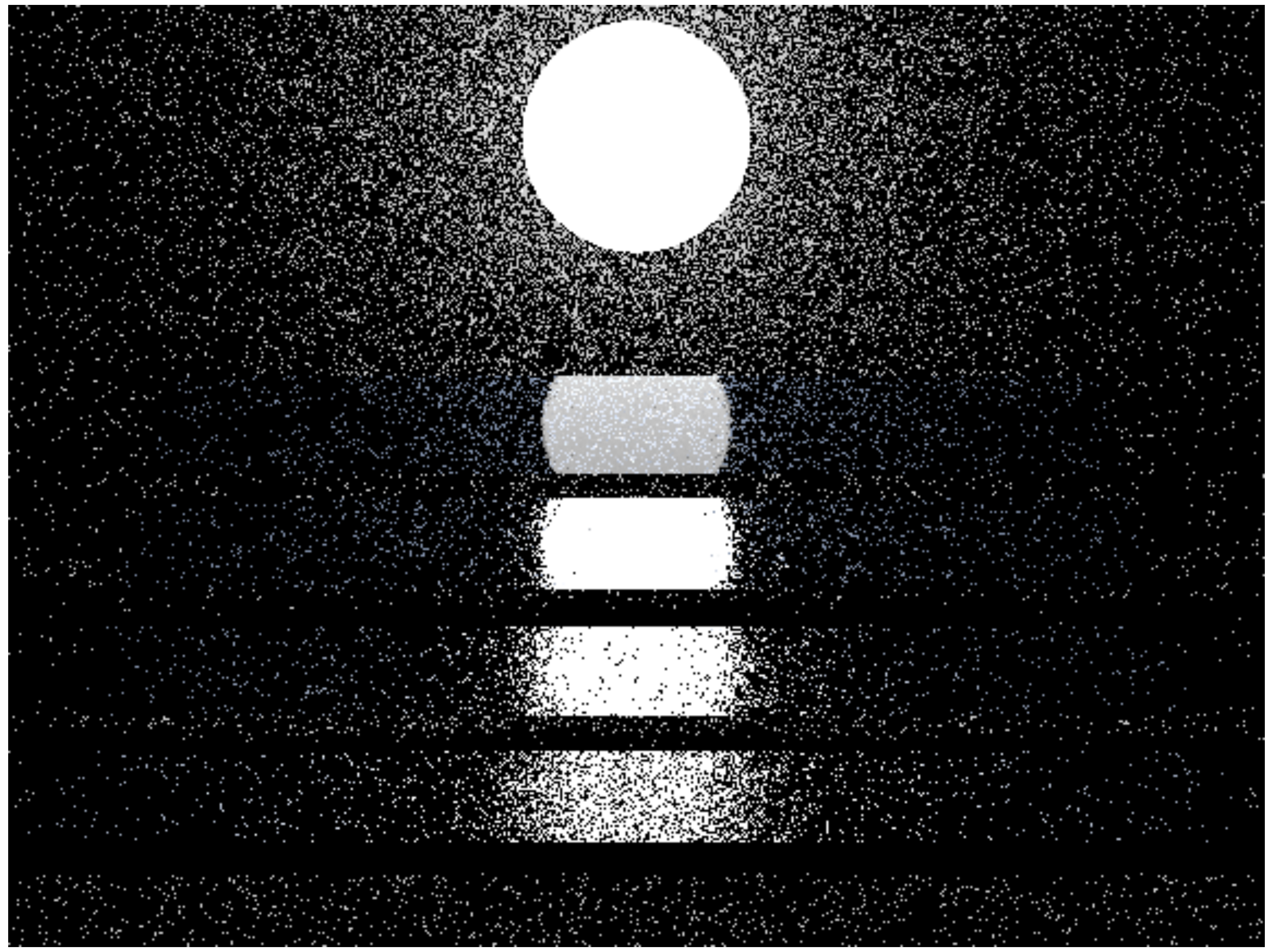
Cosine weighted spherical/hemispherical sampling

Variance reduction: Importance sampling



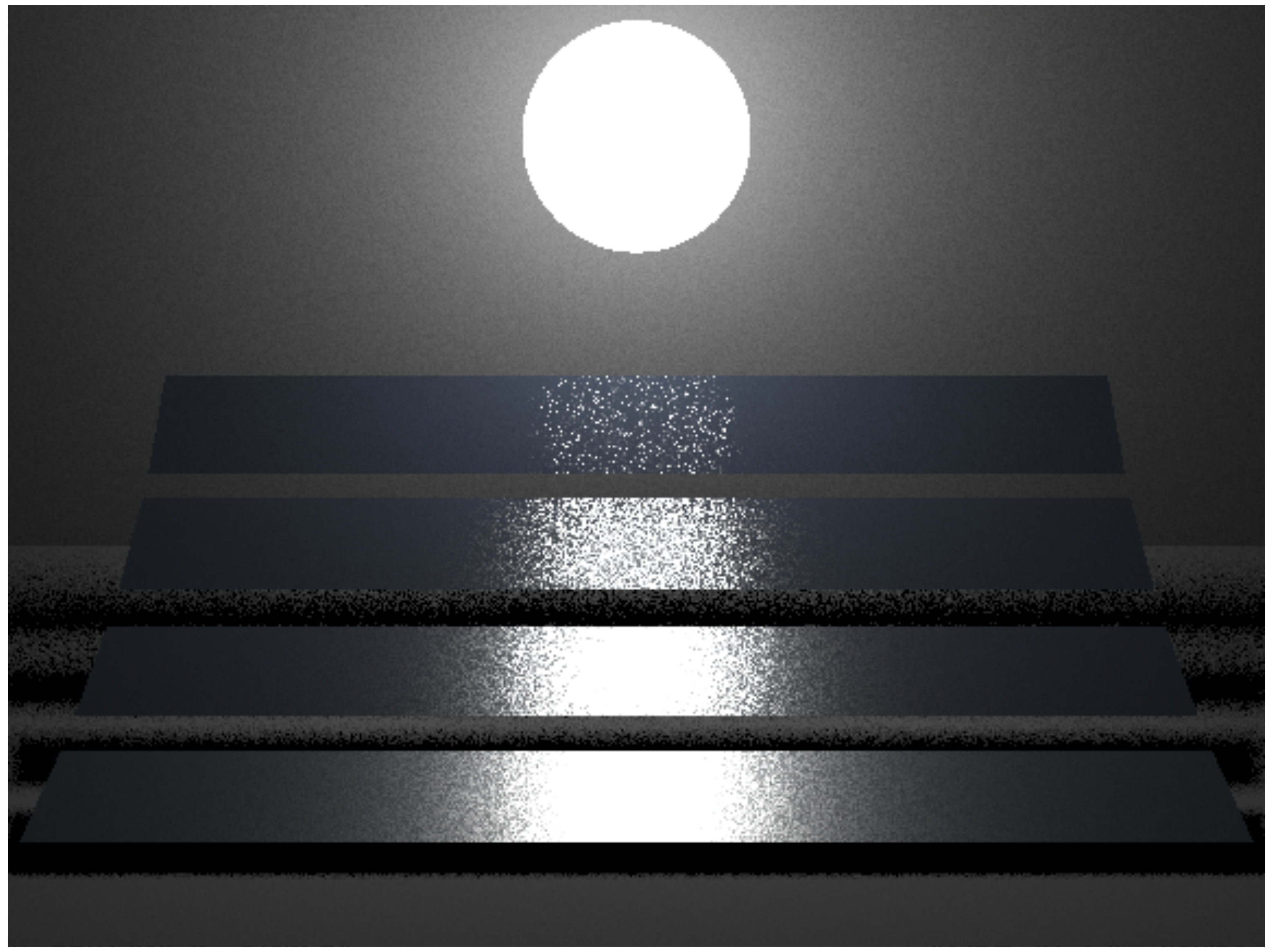
Reference image

N = 1024 spp



BSDF importance sampling

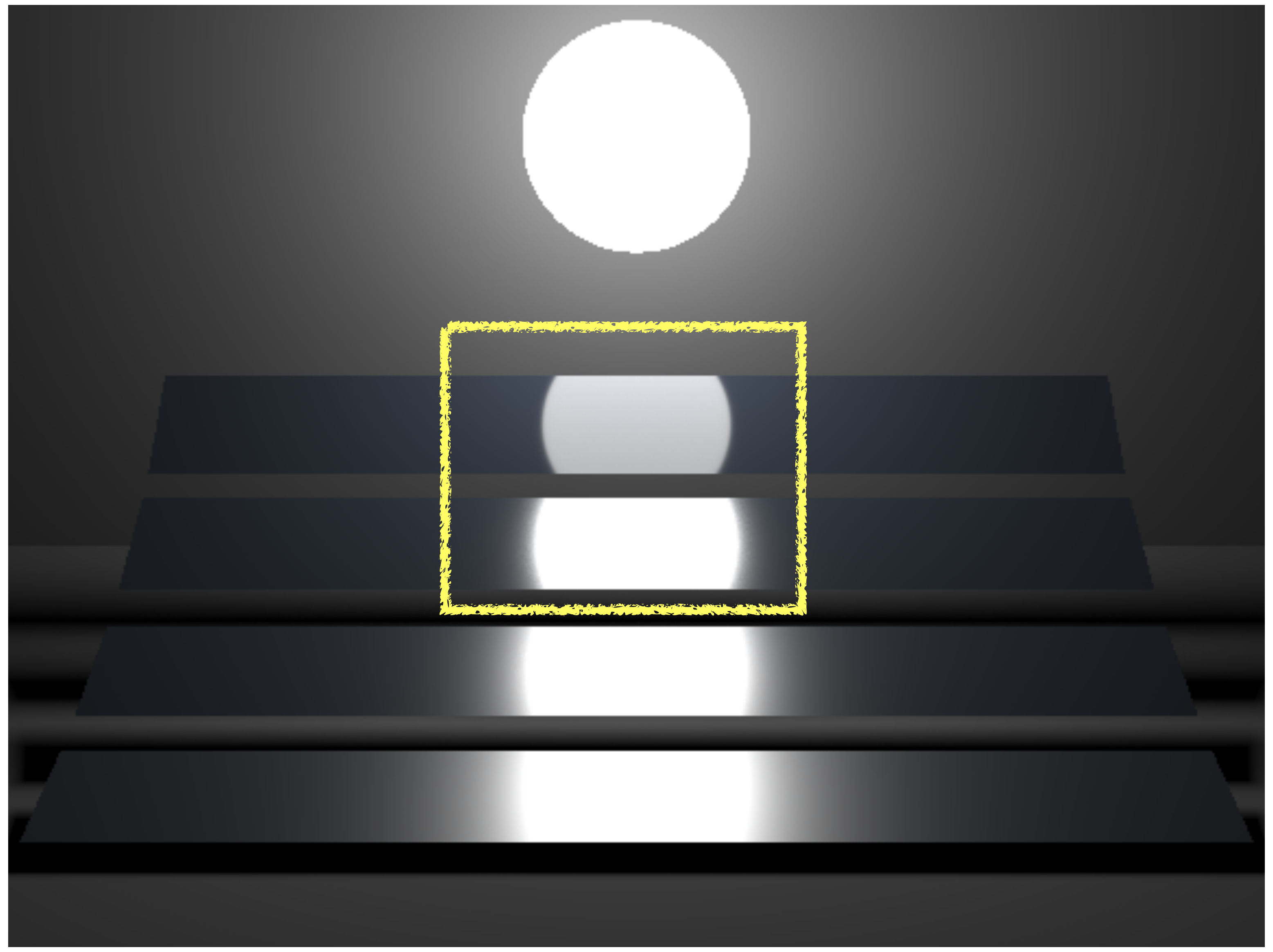
N = 4 spp



Light importance sampling

N = 4 spp

Variance reduction: Importance sampling

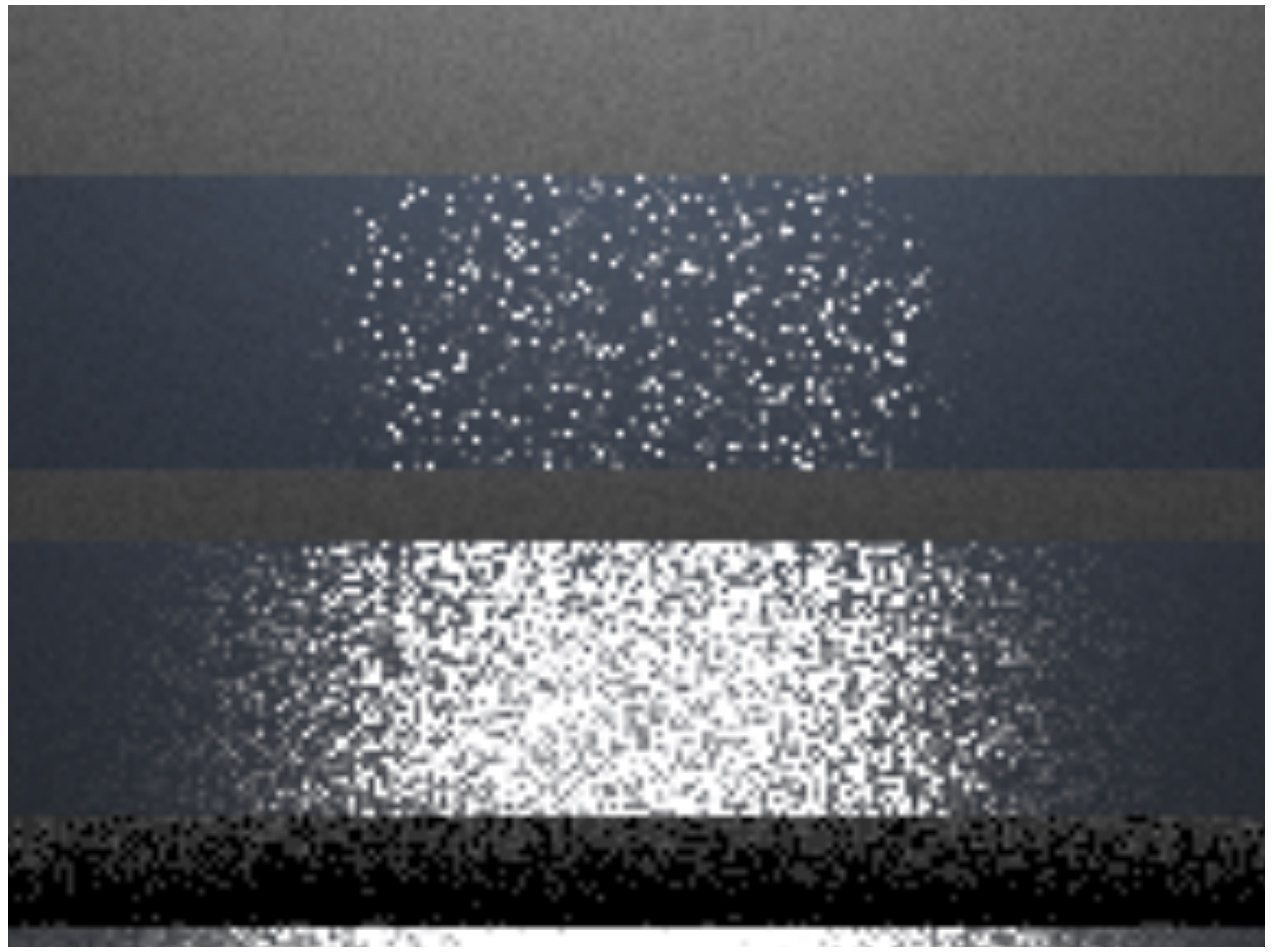


Reference image

N = 1024 spp

BSDF importance sampling

N = 4 spp

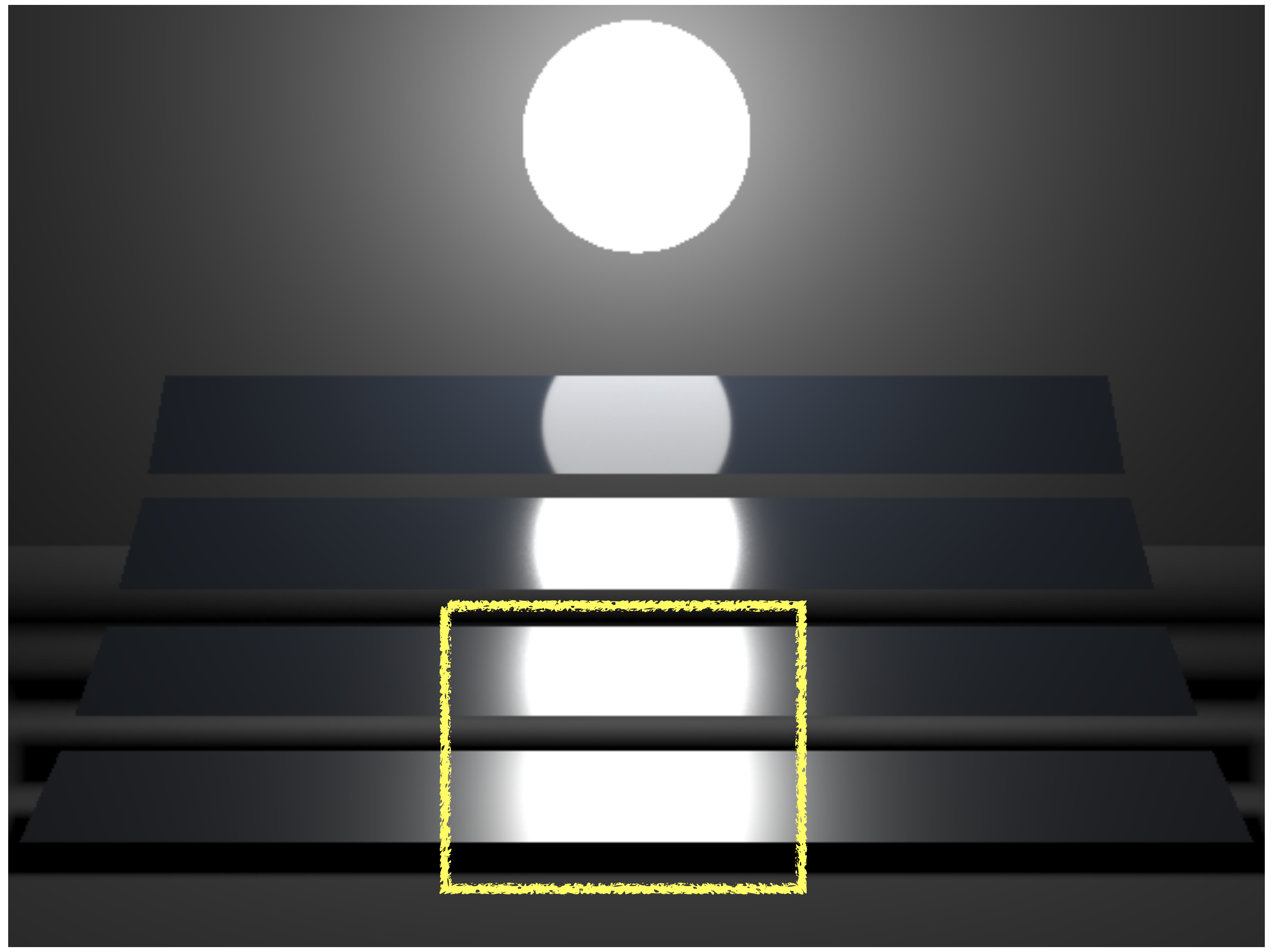


Light importance sampling

N = 4 spp

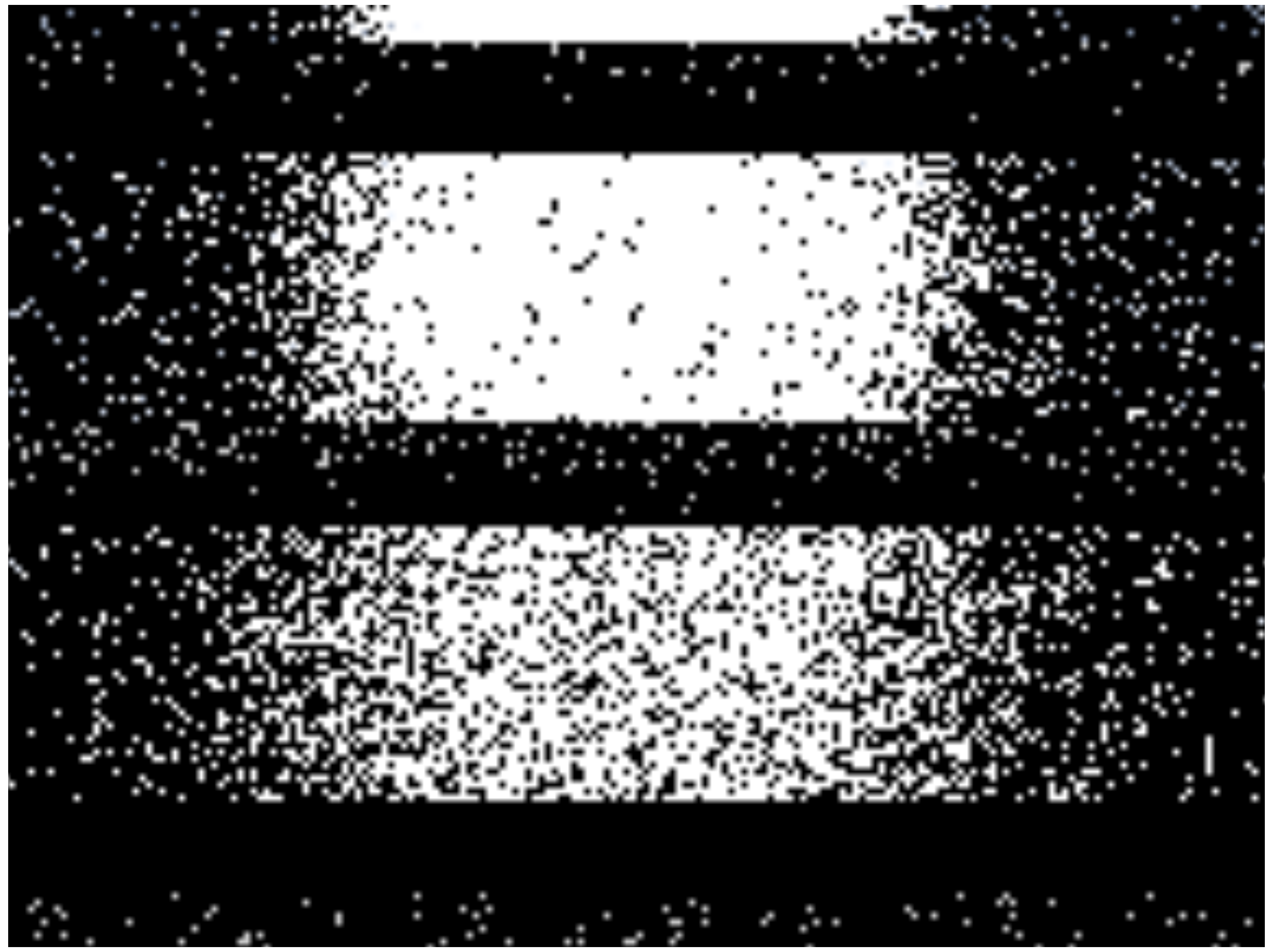
BSDF sampling is better in some regions

Variance reduction: Importance sampling



Reference image

N = 1024 spp



BSDF importance sampling

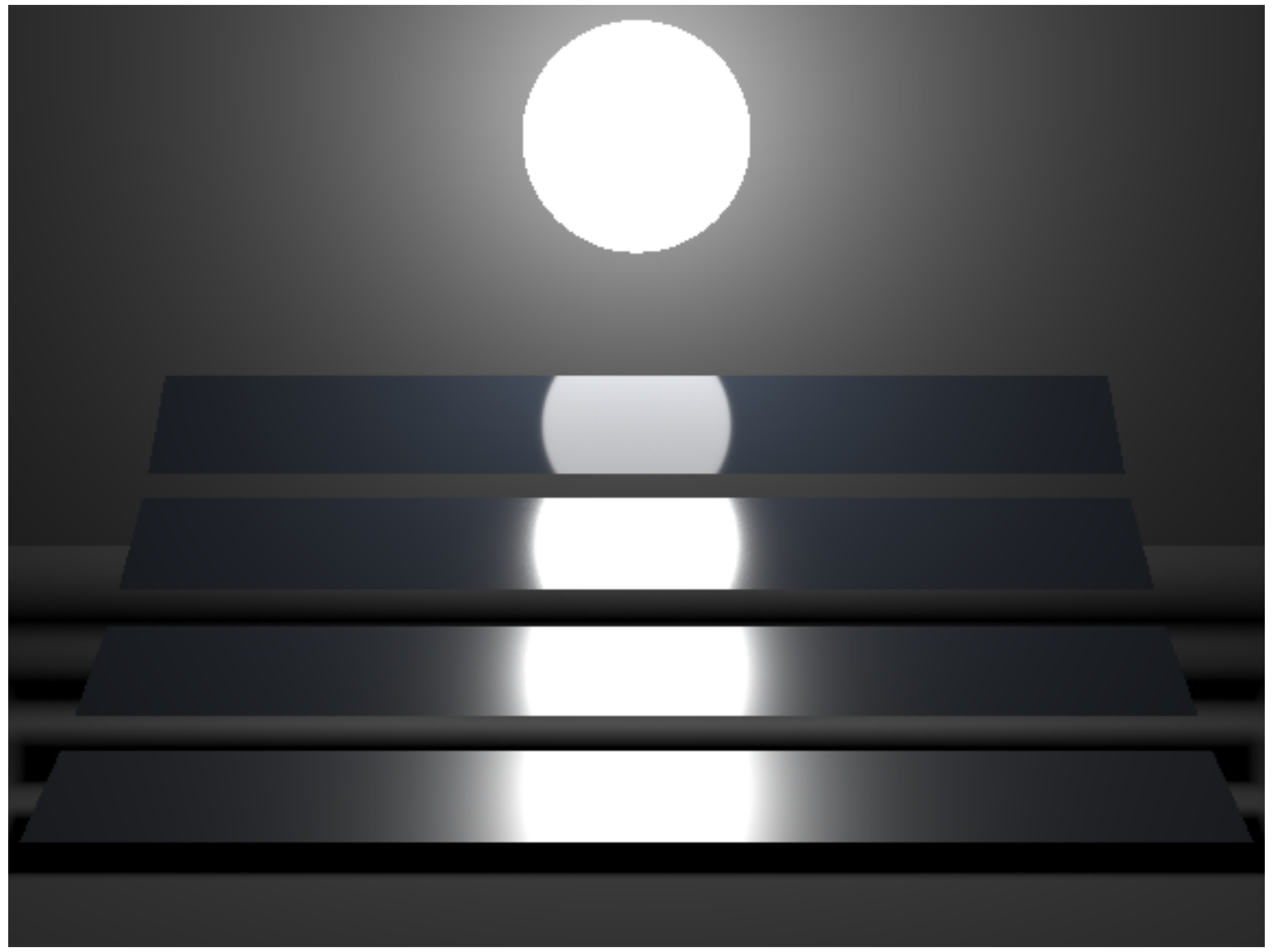
N = 4 spp

Light importance sampling

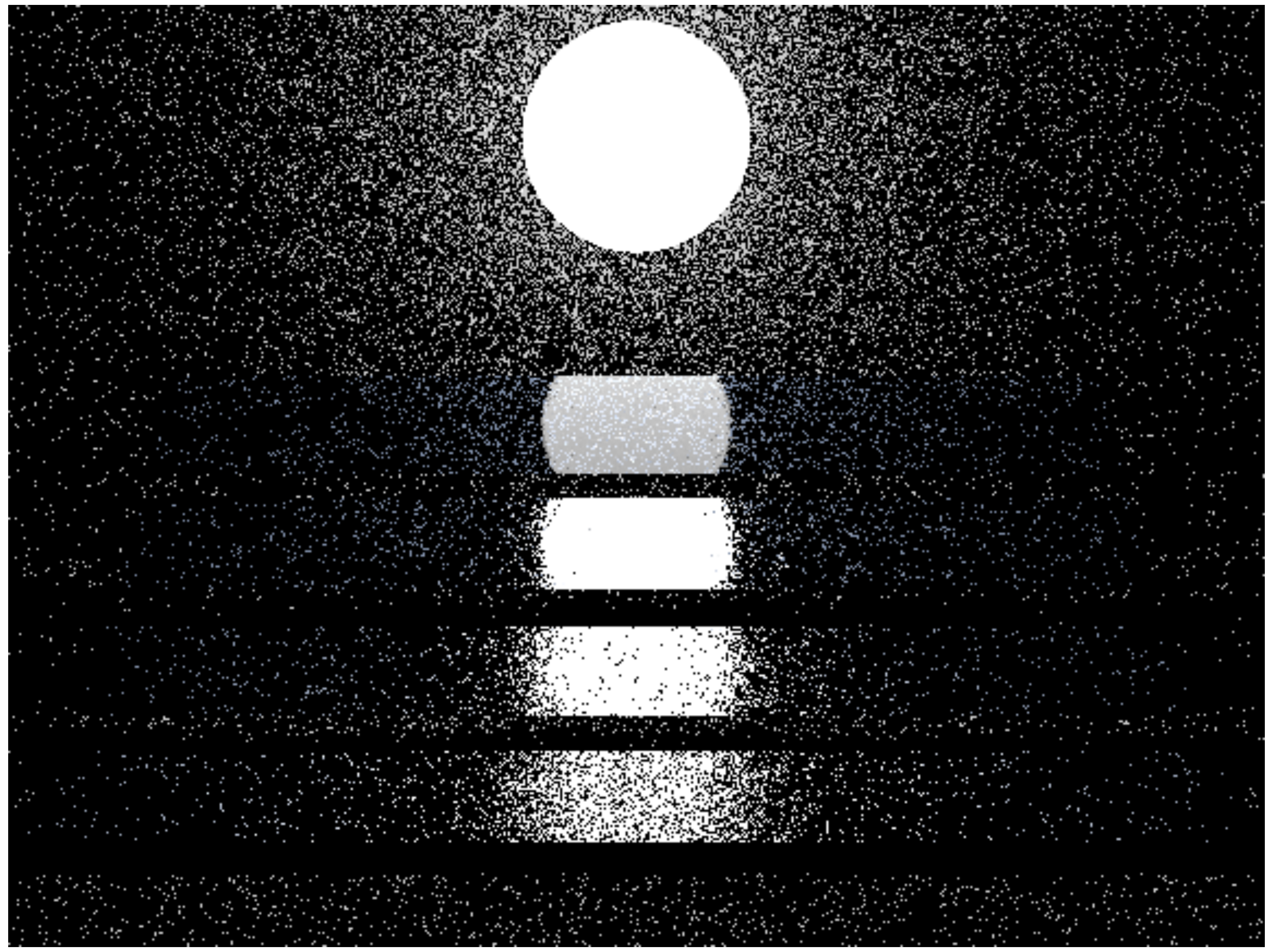
N = 4 spp

Light sampling is better in other regions

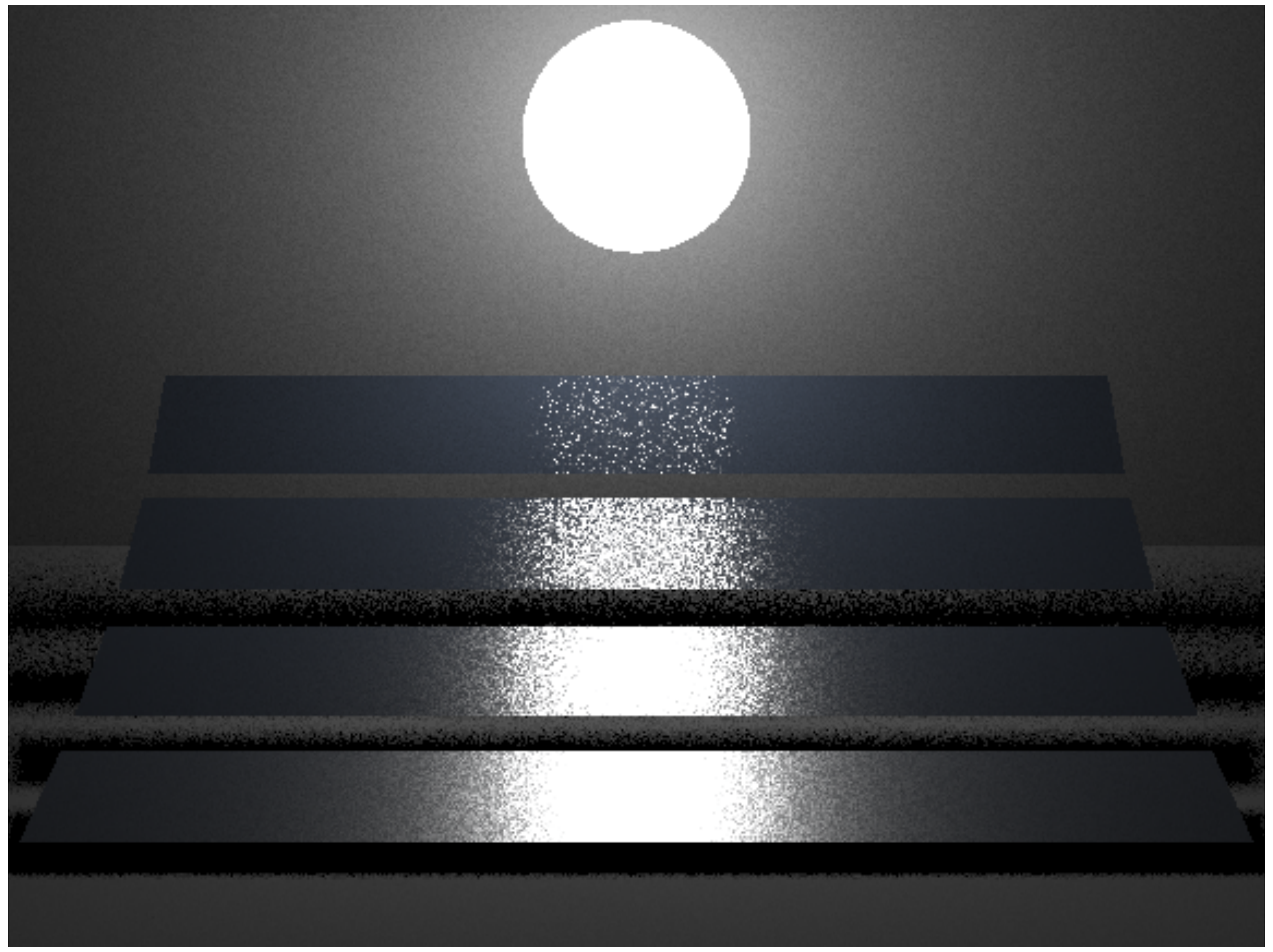
Variance reduction: Importance sampling



Reference image



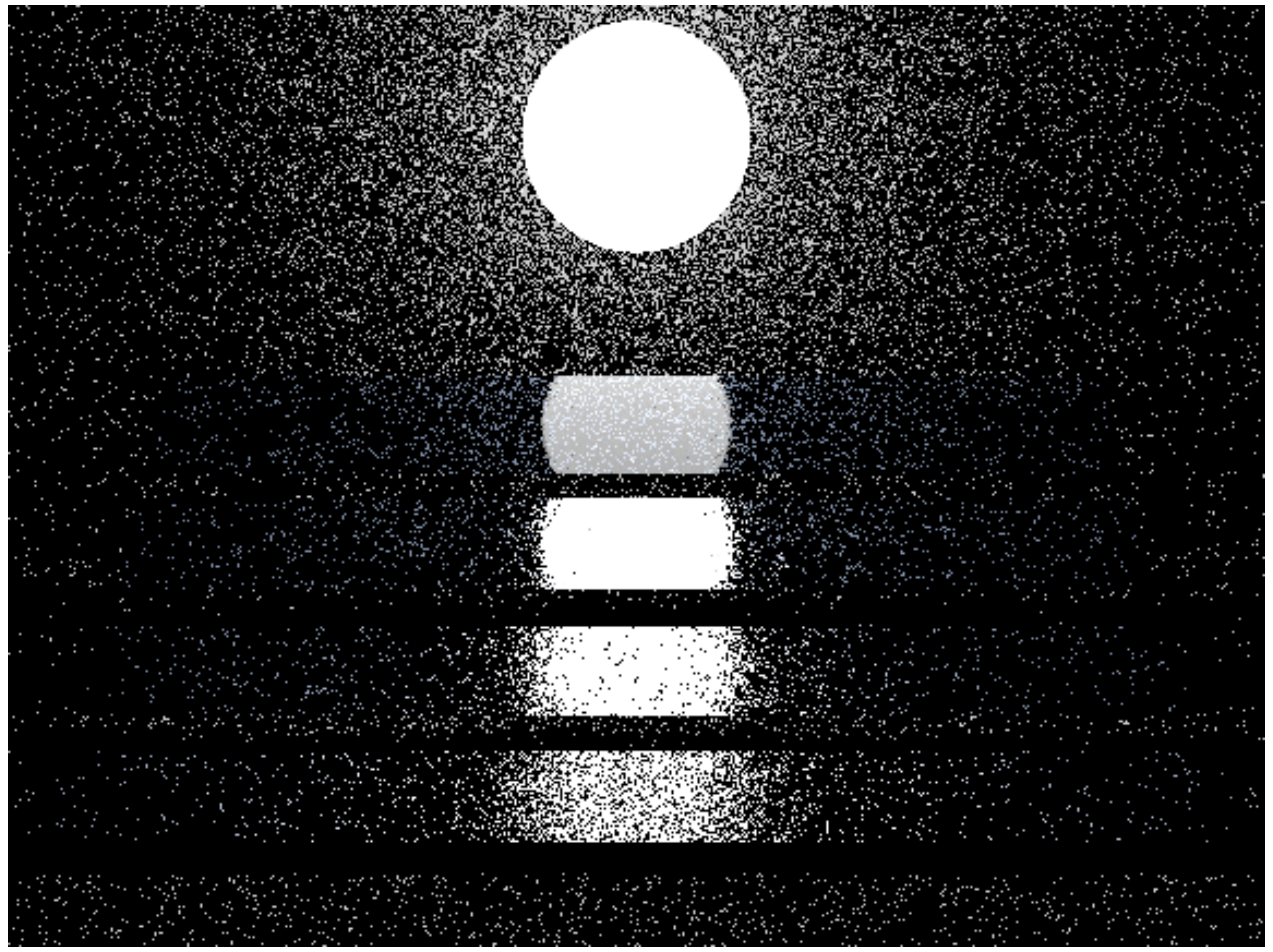
BSDF importance sampling



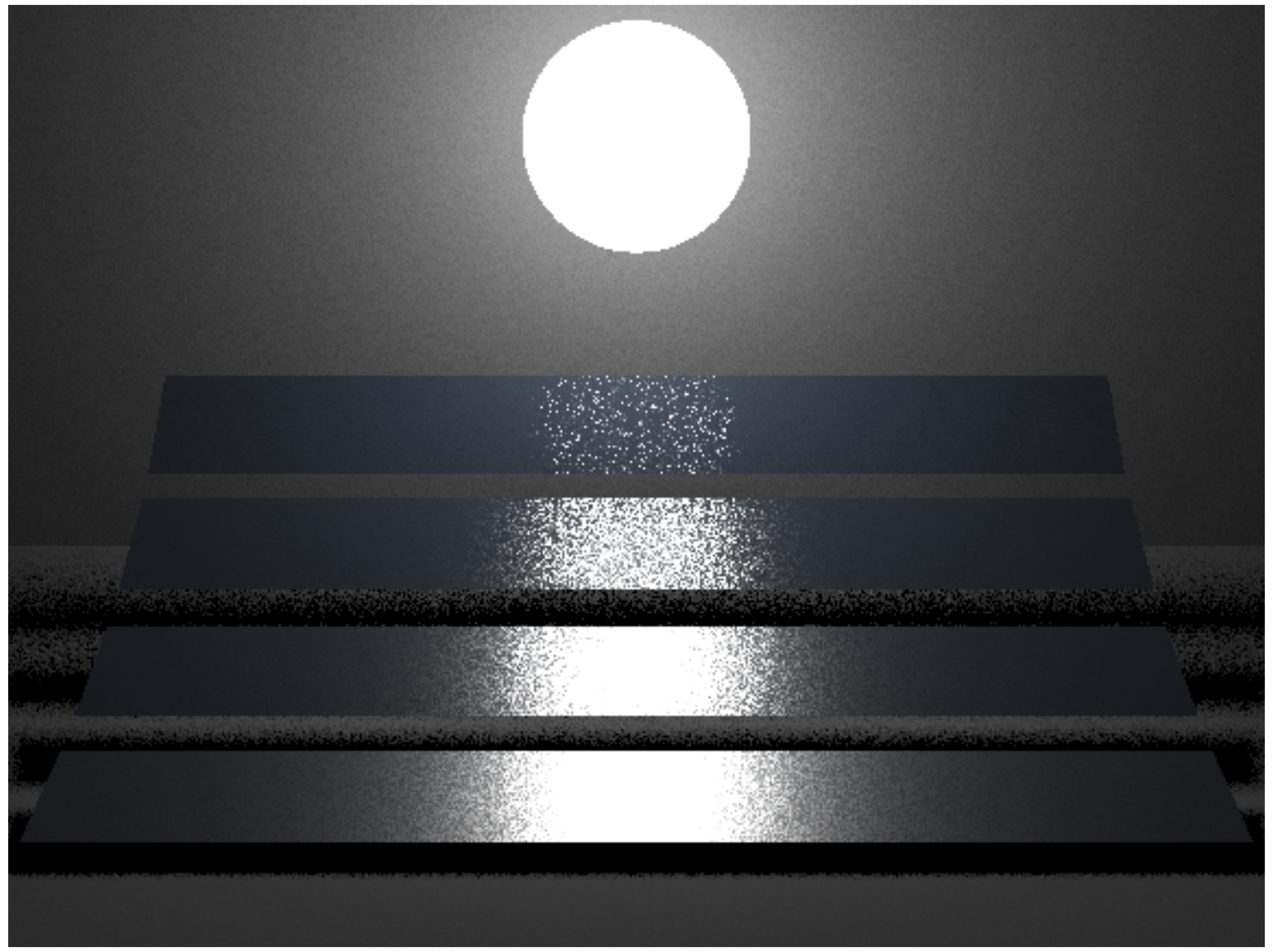
Light importance sampling

Can we combine the benefits of different PDFs ? **Yes!**

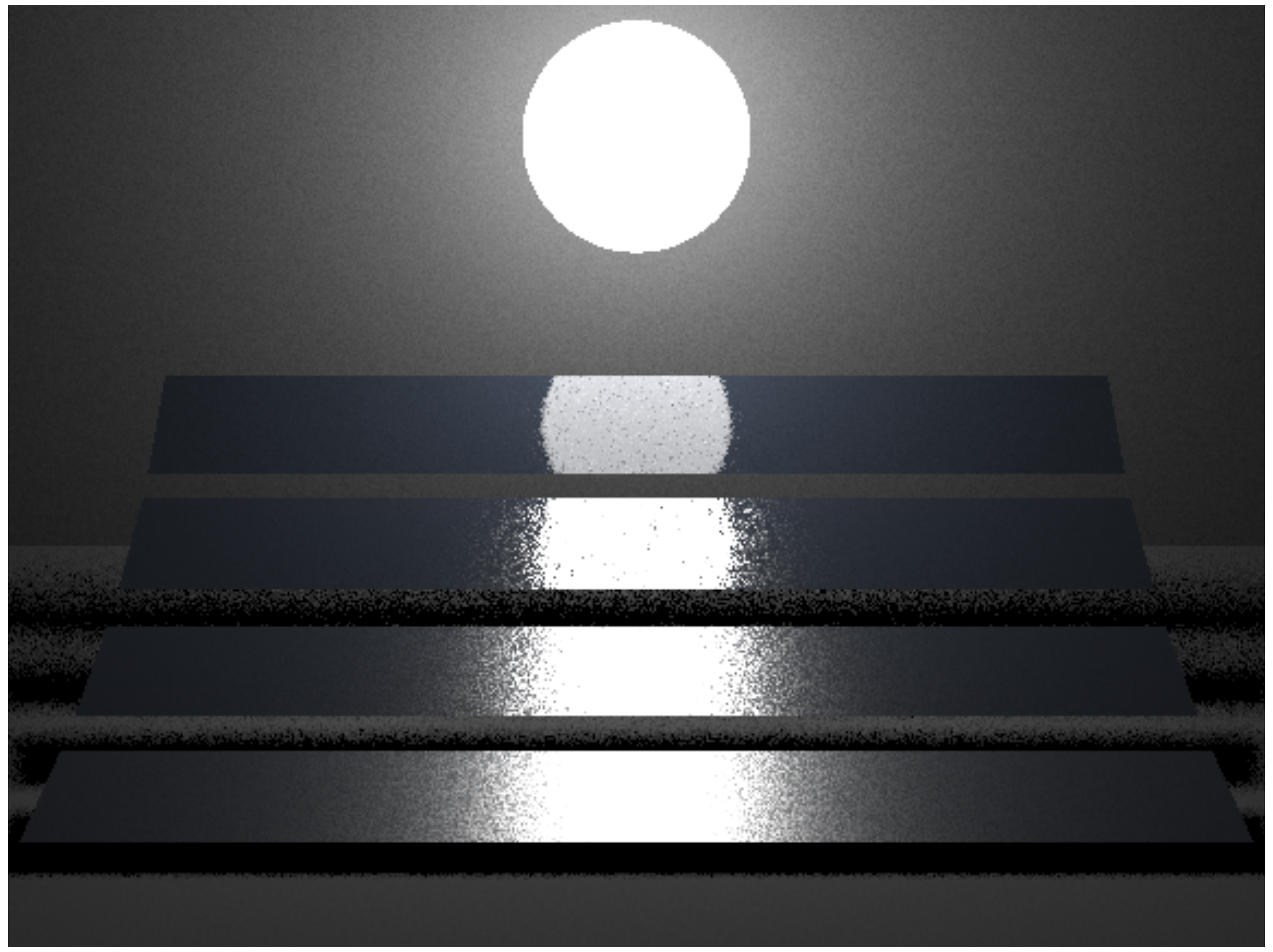
Variance reduction: Importance sampling



BSDF importance sampling



Light importance sampling



Multiple Importance Sampling

Can we combine the benefits of different PDFs ? **Yes!**

Variance reduction: Multiple Importance sampling

Multiple Importance Sampling

$$I_N = \frac{1}{N} \sum_{i=1}^{N} \frac{f(x)g(x)}{p(x)}$$

$$p(x) \propto ???$$

$$\mathbf{I}_N = \frac{1}{n_f} \sum_{i=1}^{n_f} \frac{f(X_i)g(X_i)w_f(X_i)}{p_f(X_i)} + \frac{1}{n_g} \sum_{j=1}^{n_g} \frac{f(Y_j)g(Y_j)w_g(Y_j)}{p_g(Y_j)}$$

Variance reduction: Multiple Importance sampling

Multiple Importance Sampling

$$\mathbf{I}_N = \frac{1}{n_f} \sum_{i=1}^{n_f} \frac{f(X_i)g(X_i)w_f(X_i)}{p_f(X_i)} + \frac{1}{n_g} \sum_{j=1}^{n_g} \frac{f(Y_j)g(Y_j)w_g(Y_j)}{p_g(Y_j)}$$

Balance heuristic: $$w_s(x) = \frac{n_s p_s(x)}{\sum_i n_i p_i(x)}$$

Power heuristic: $$w_s(x) = \frac{(n_s p_s(x))^\beta}{\sum_i (n_i p_i(x))^\beta}$$ $\beta = 2$

Variance reduction: Control Variate

Variance reduction: Control Variate

- To reduce variance, an easily evaluated approximation to the integrand is sought
- Instead sampling all points independently, control variates make use of correlated points in the sampling
- The mathematical basis of control variates is the linearity property of the Lebesgue integral, i.e., one try to find an analytically Lebesgue-integrable function g that is similar to the integral under study.

Variance reduction: Control Variate

$$\int_Q f(x)dx = \int_Q g(x)dx + \int_Q (f(x) - g(x))dx$$

$$= \int_Q g(x)dx + \int_Q \frac{(f(x) - g(x))}{p(x)} p(x)dx$$

$$= \int_Q g(x)dx + \mathbf{E}\left[\frac{(f(x) - g(x))}{p(x)}\right]$$

Variance reduction: Control Variate

$$\int_Q f(x)dx = \int_Q g(x)dx + \mathbf{E}\left[\frac{(f(x) - g(x))}{p(x)}\right]$$

Since we don't know the analytic integral solution of $f(x)$ the corresponding **estimator** can be written as:

$$\mathbf{I}_N^{CV} = \int_Q g(x)dx + \frac{1}{N}\sum_{i=1}^{N}\left[\frac{(f(x_i) - g(x_i))}{p(x_i)}\right]$$

Variance reduction: Control Variate

$$\mathbf{I}_N^{CV} = \int_Q g(x)dx + \frac{1}{N}\sum_{i=1}^{N}\left[\frac{(f(x_i) - g(x_i))}{p(x_i)}\right]$$

The integral on the right hand side can be evaluated exactly,
where as the variance of the estimator is given by:

$$\mathrm{Var}(\mathbf{I}_N^{CV}) = \frac{1}{N^2}\sum_{i=1}^{N}\mathrm{Var}\left(\frac{(f(x_i) - g(x_i))}{p(x_i)}\right)$$

Variance can be reduced if:

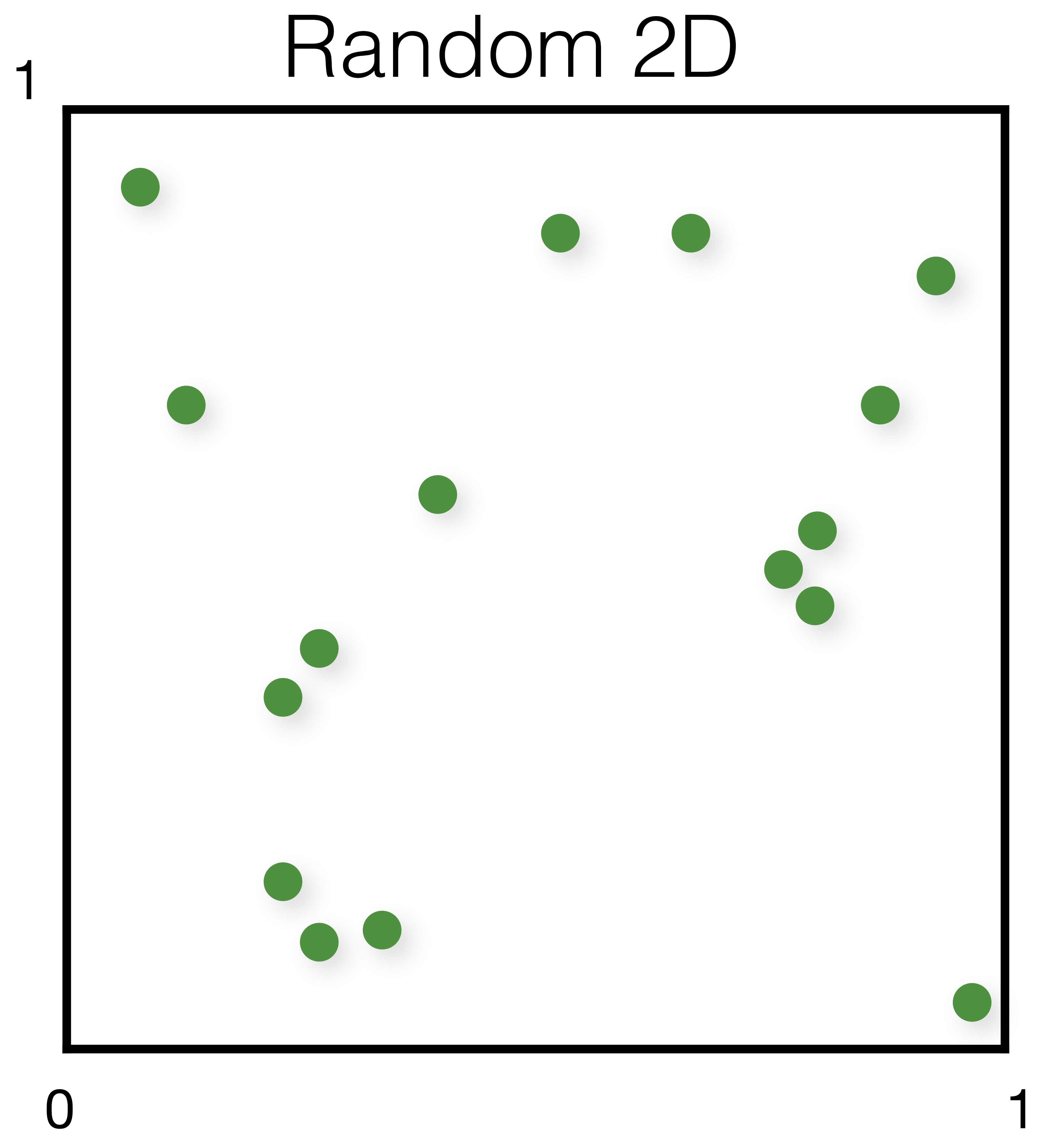
$$\mathrm{Var}\left(\frac{(f(x_i) - g(x_i))}{p(x_i)}\right) < \mathrm{Var}\left(\frac{f(x_i)}{p(x_i)}\right)$$

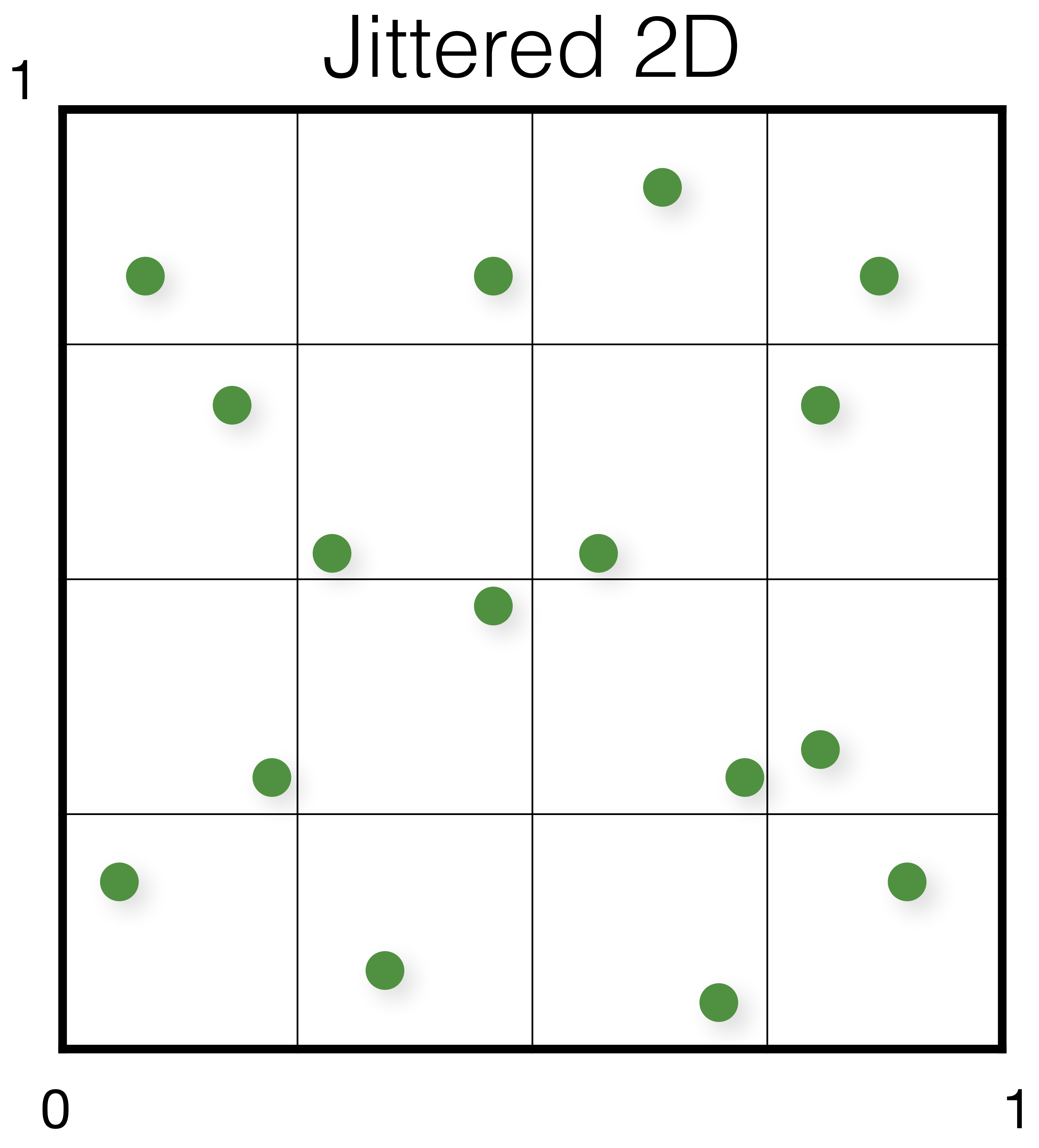
Variance reduction: Stratified Sampling

Jittered Sampling

Latin Hypercube Sampling

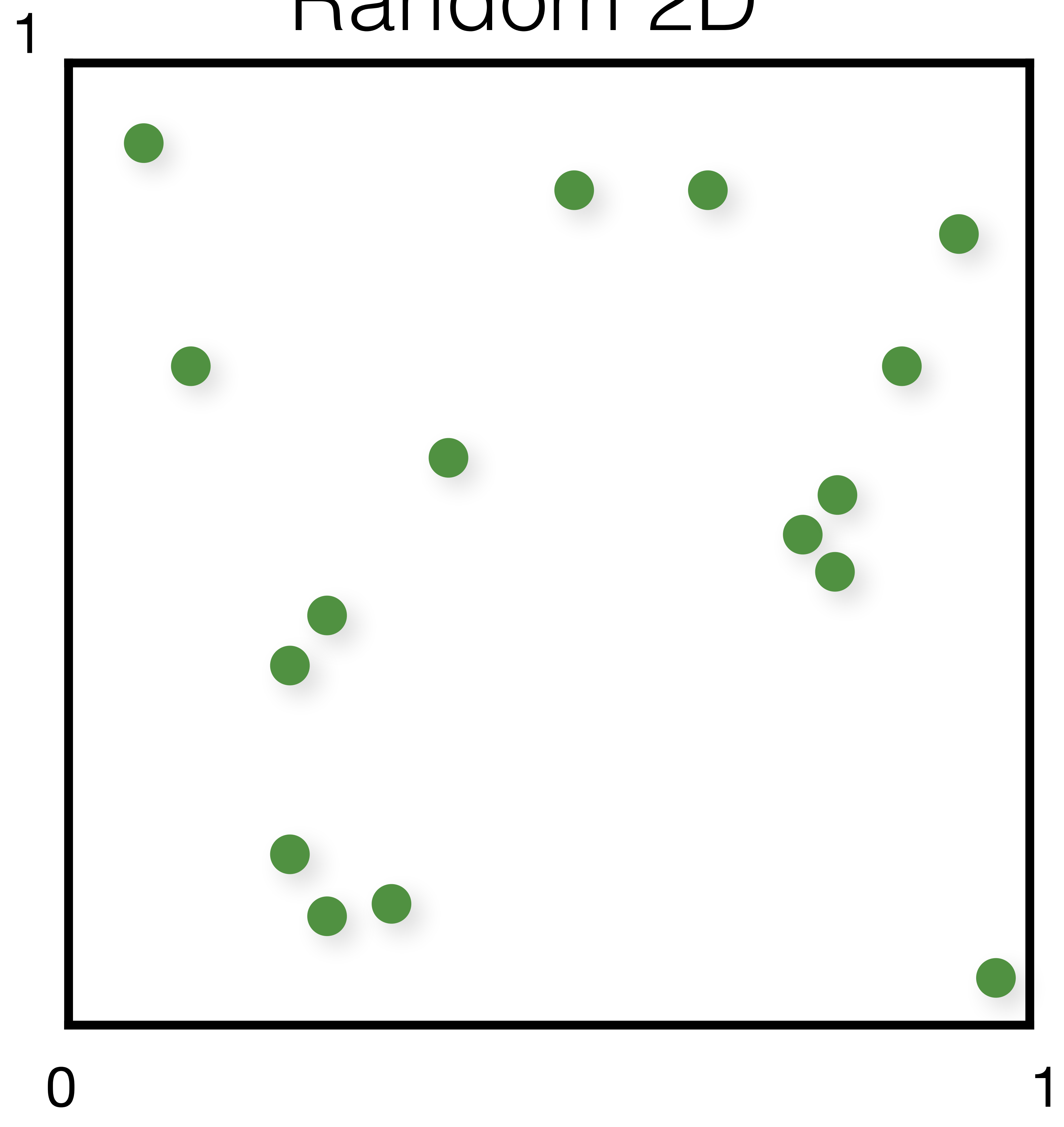
Variance reduction: Stratified Sampling



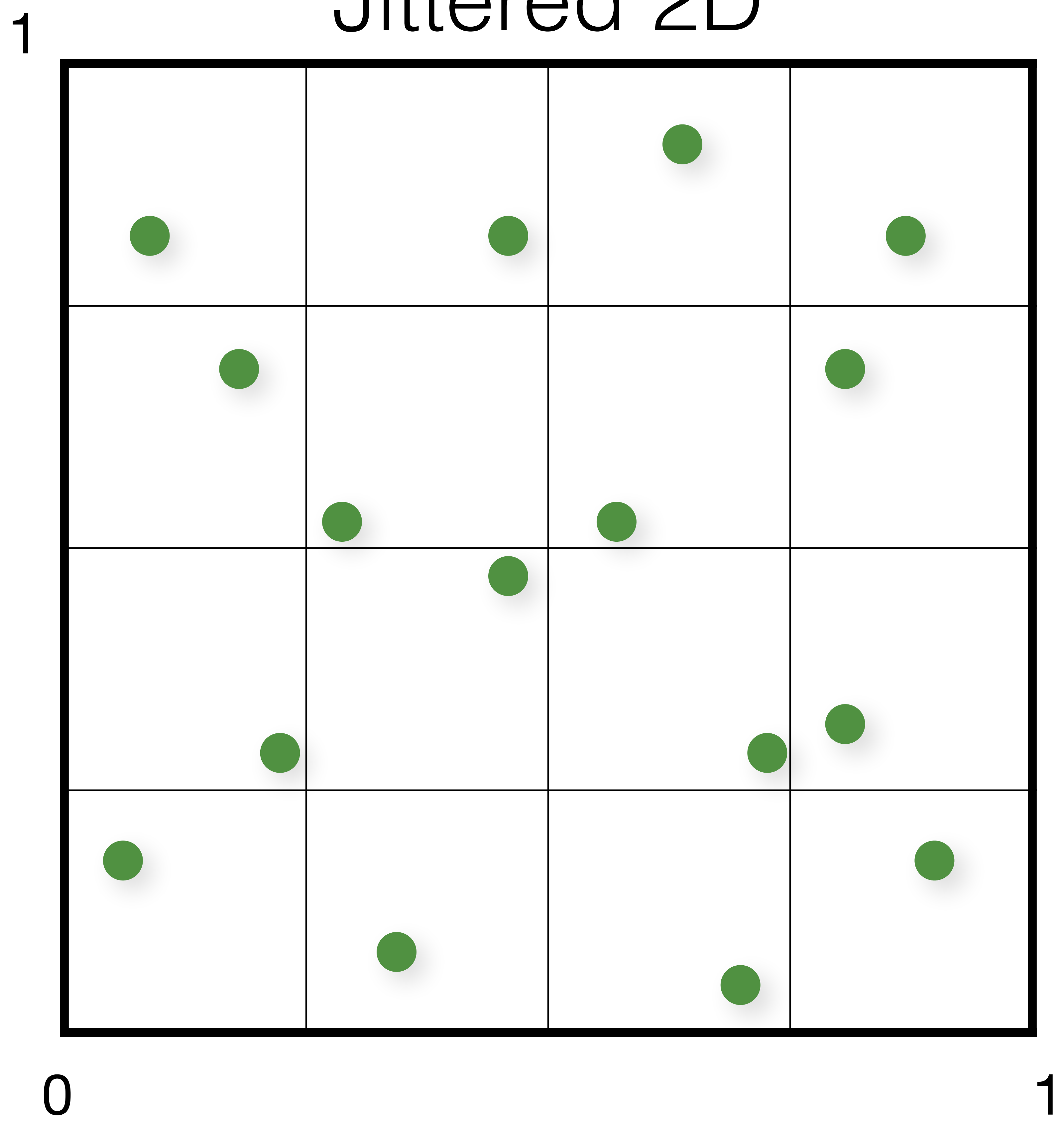


Variance reduction: Stratified Sampling

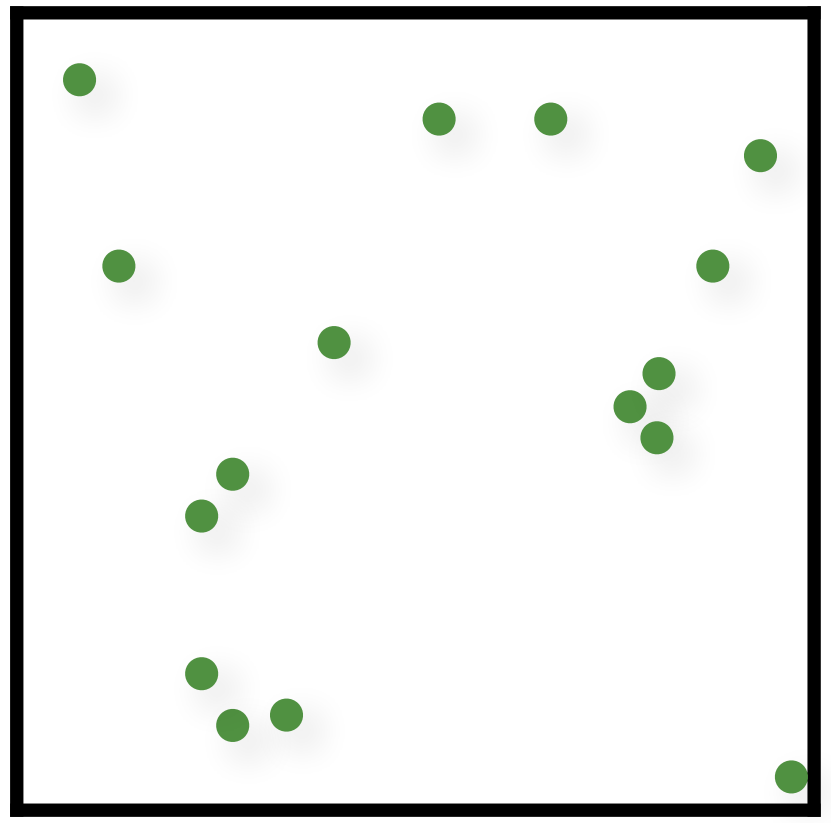
Random 2D



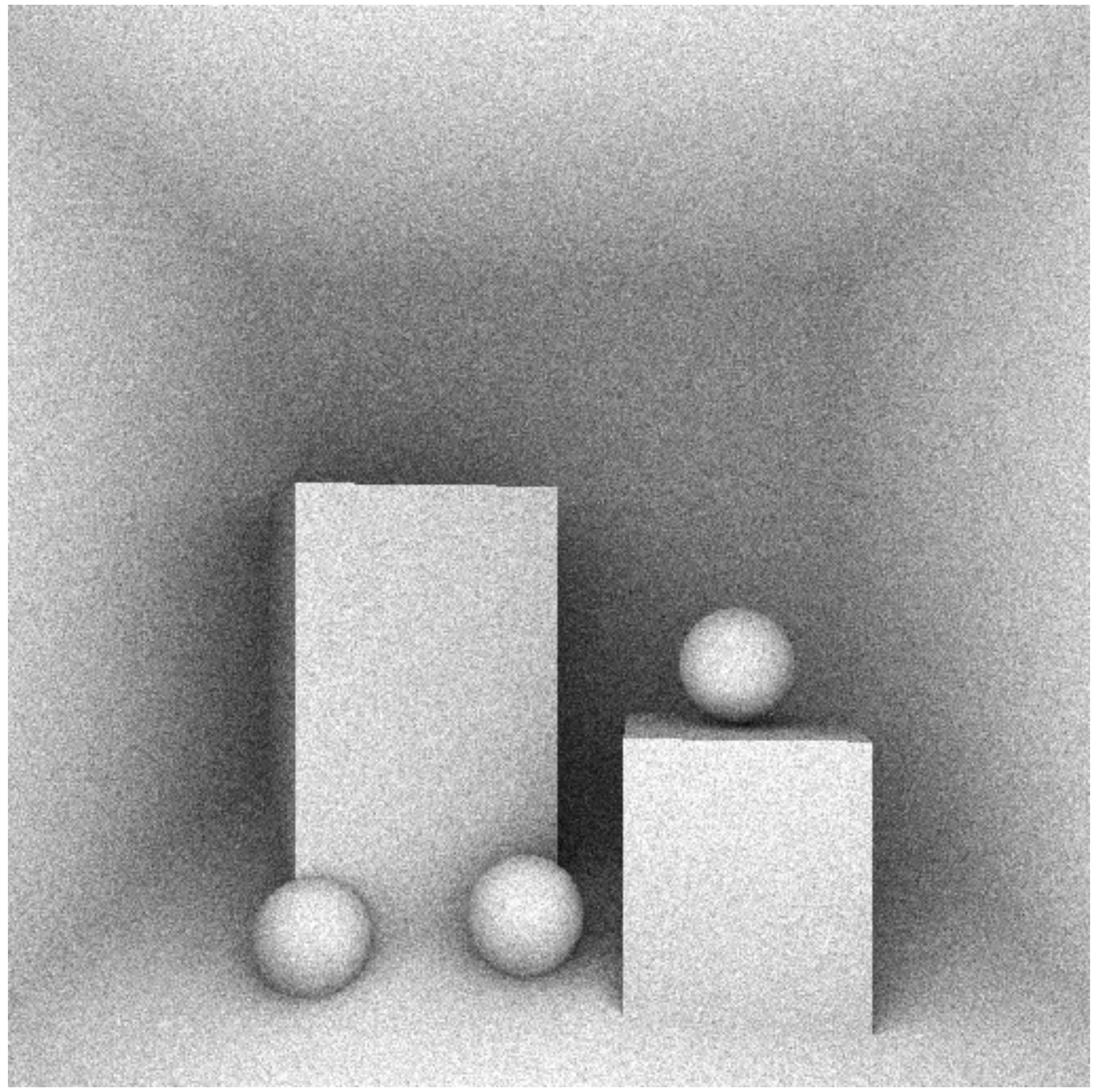
Jittered 2D



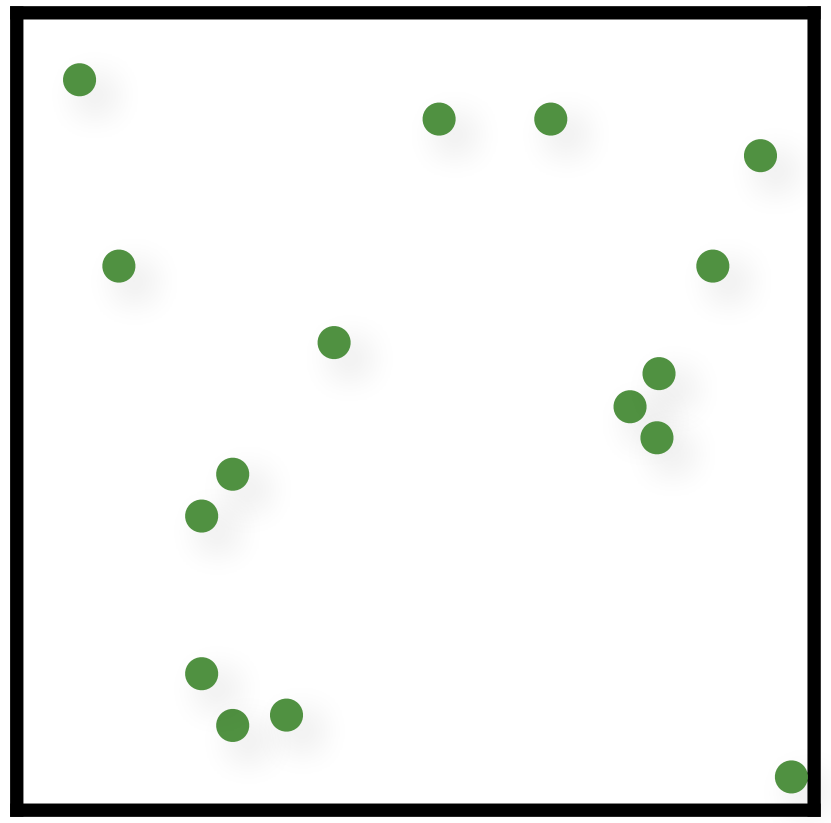
Variance reduction: Stratified sampling



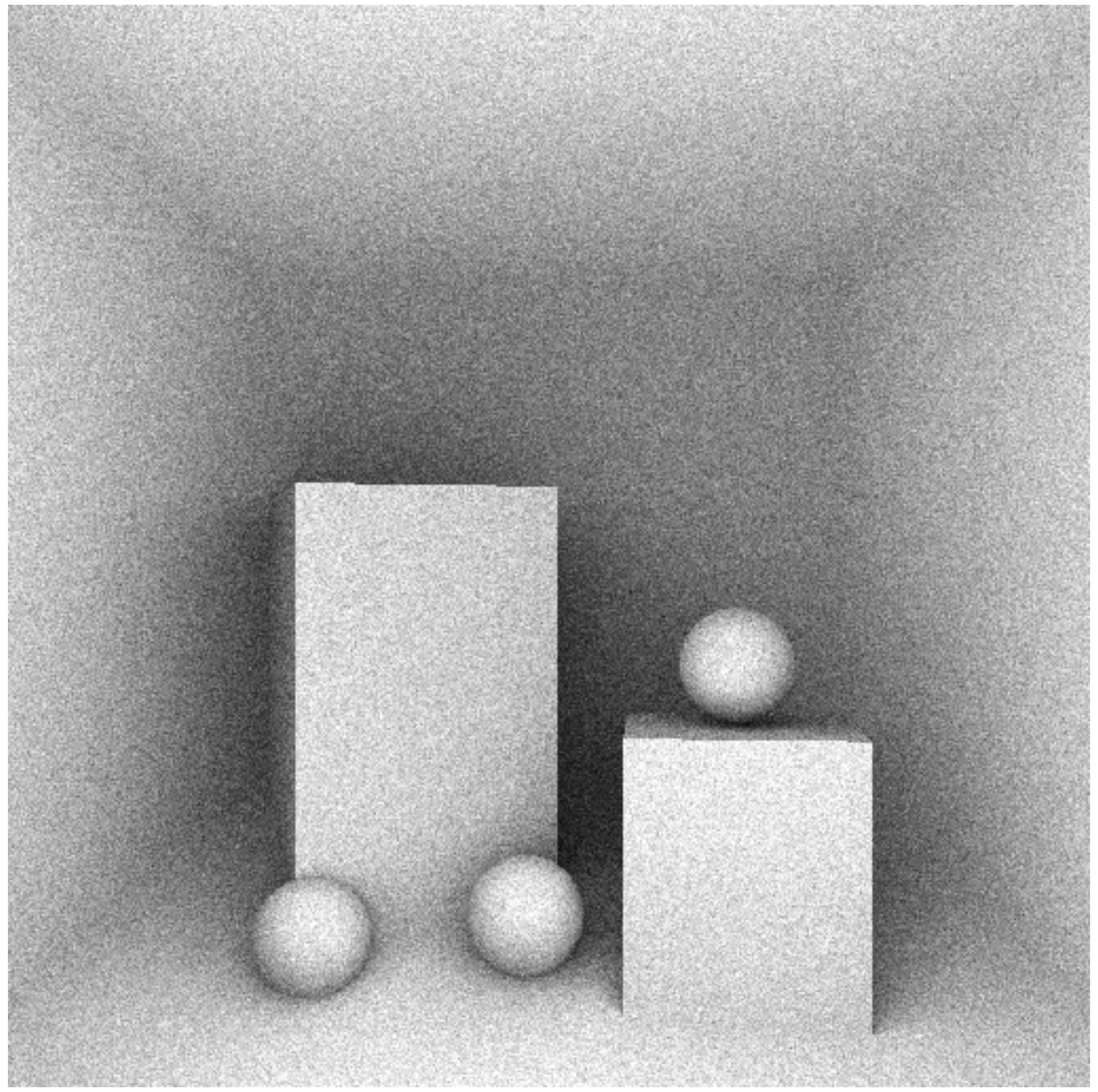
Random Samples



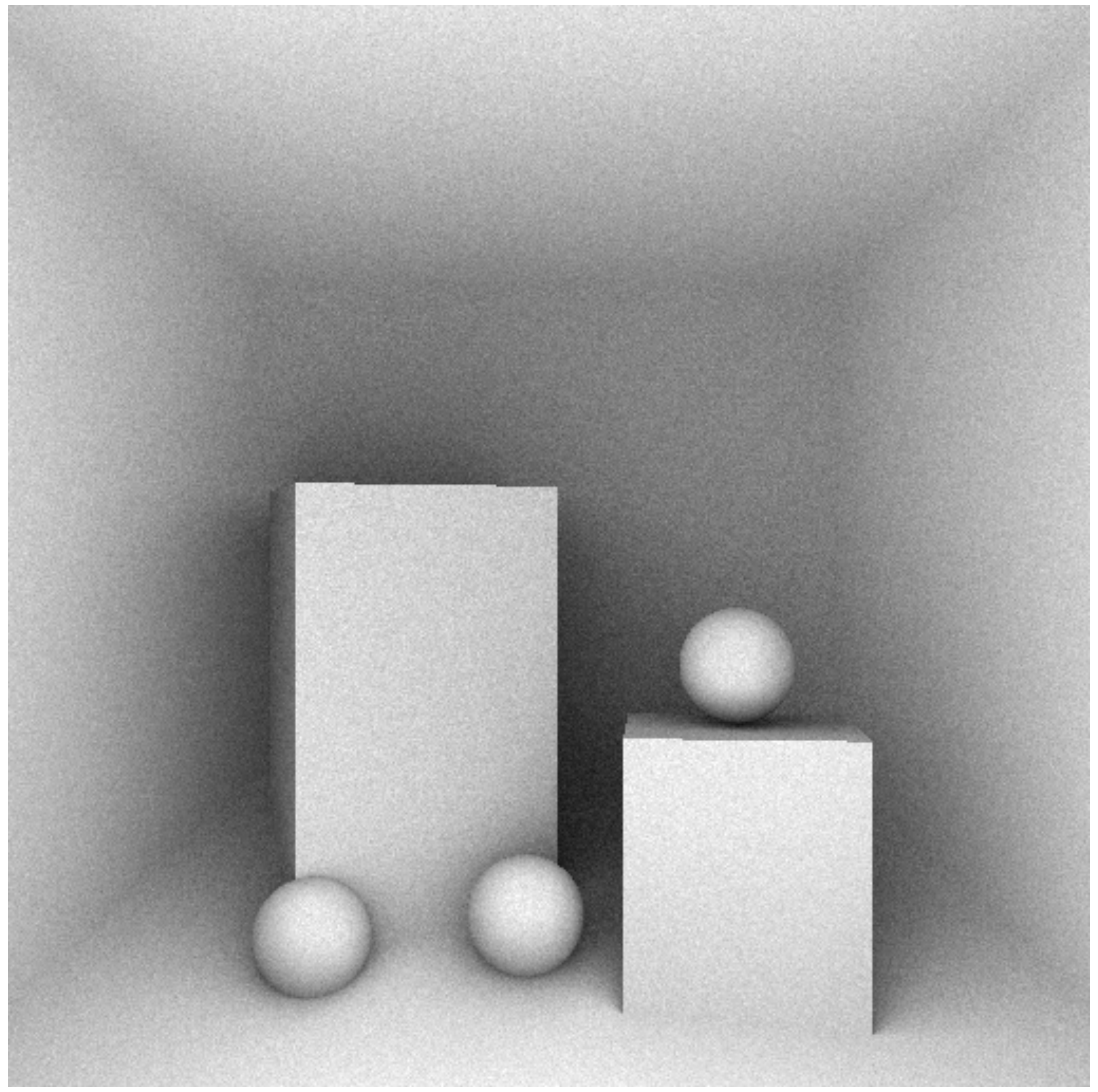
Variance reduction: Stratified sampling

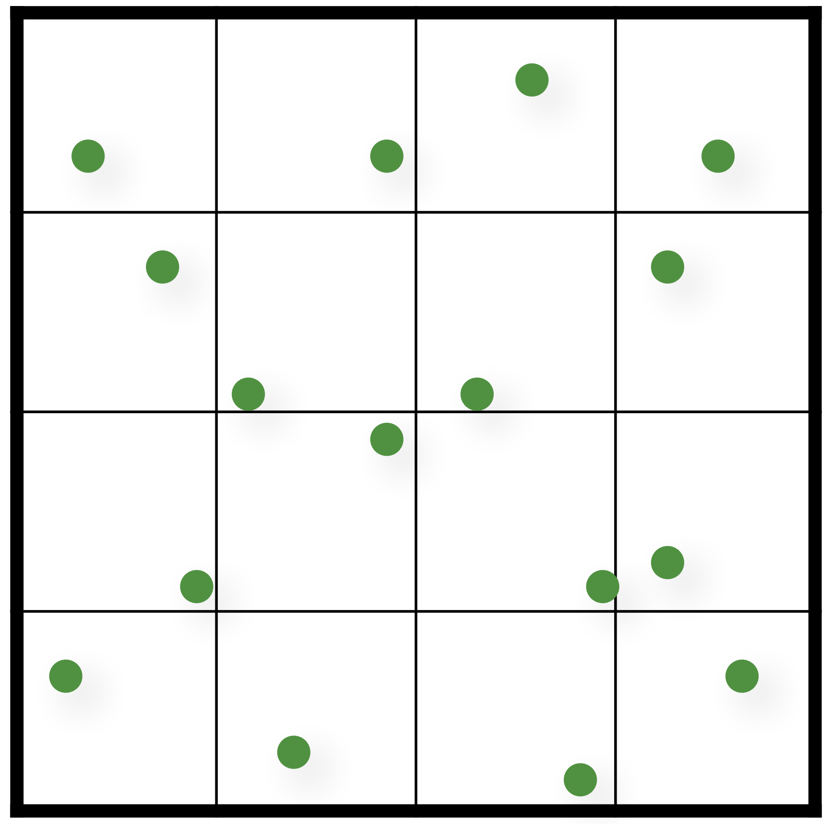


Random Samples



Jittered Samples





N = 64 spp

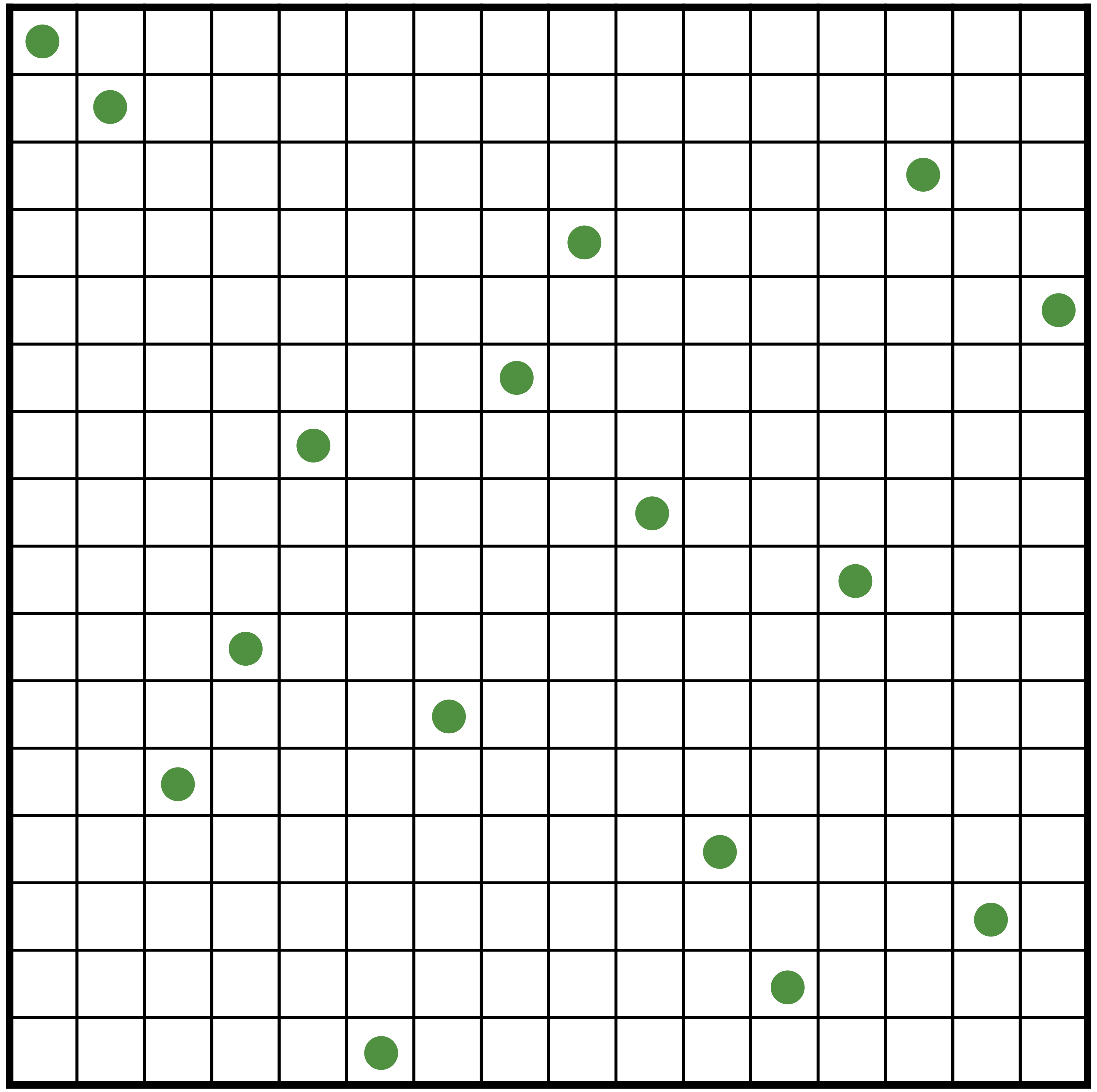
Stratified sampling suffers from the curse of dimensionality

Variance reduction: Stratified Sampling

Jittered Sampling

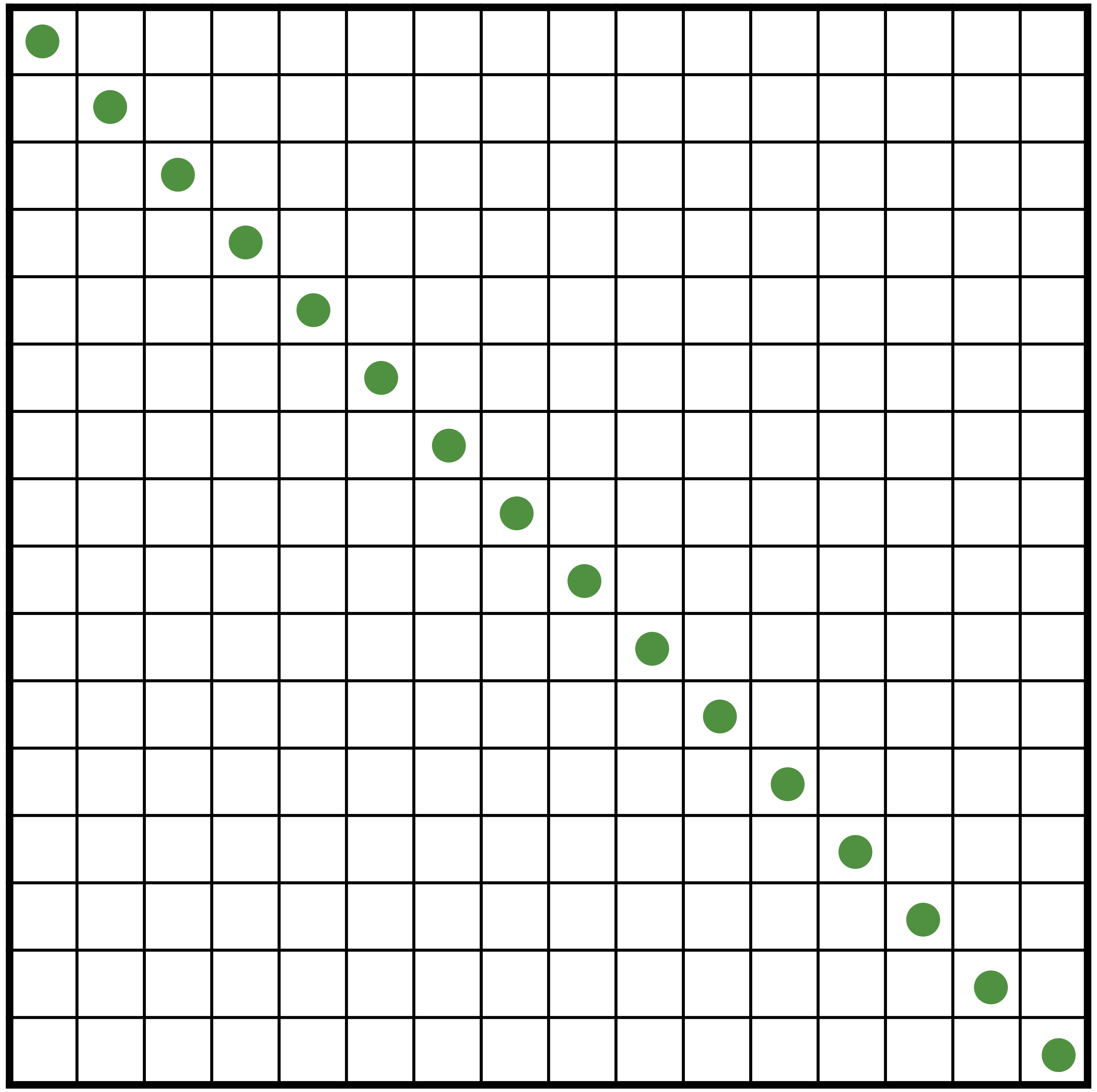
Latin Hypercube Sampling

Latin Hypercube Sampler (N-rooks)



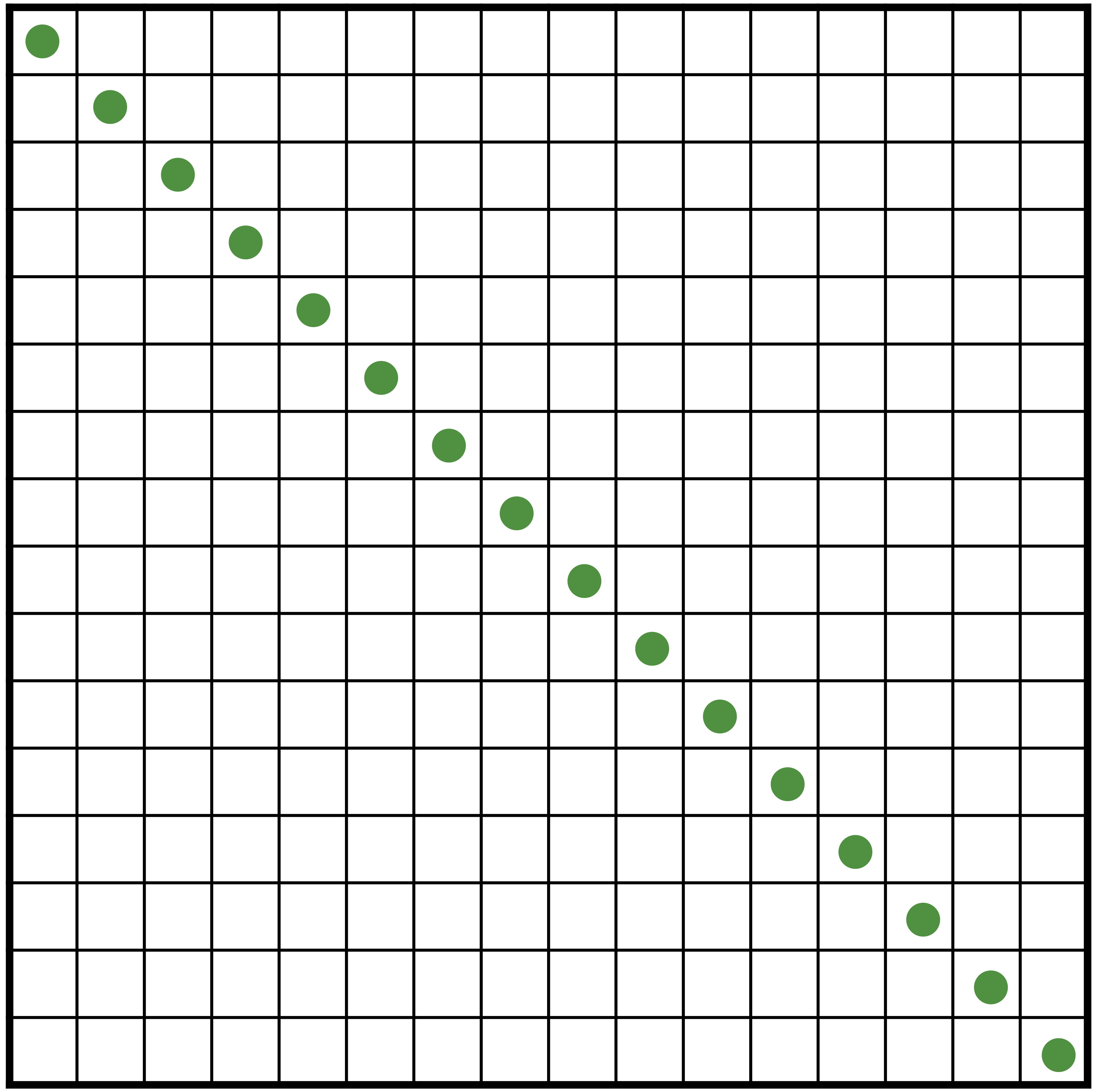
Latin Hypercube Sampler (N-rooks)

Initialize

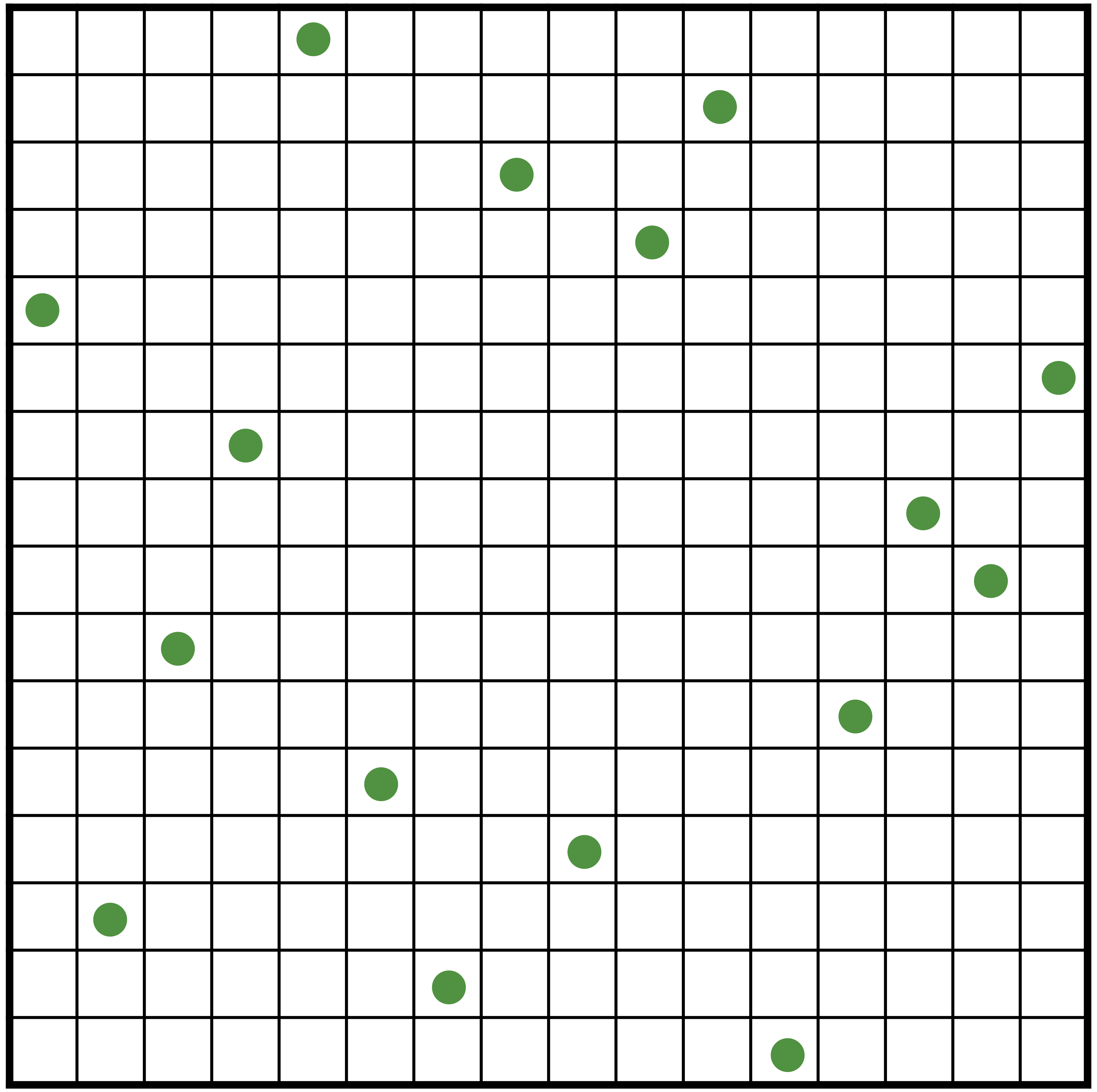


Latin Hypercube Sampler (N-rooks)

Shuffle rows

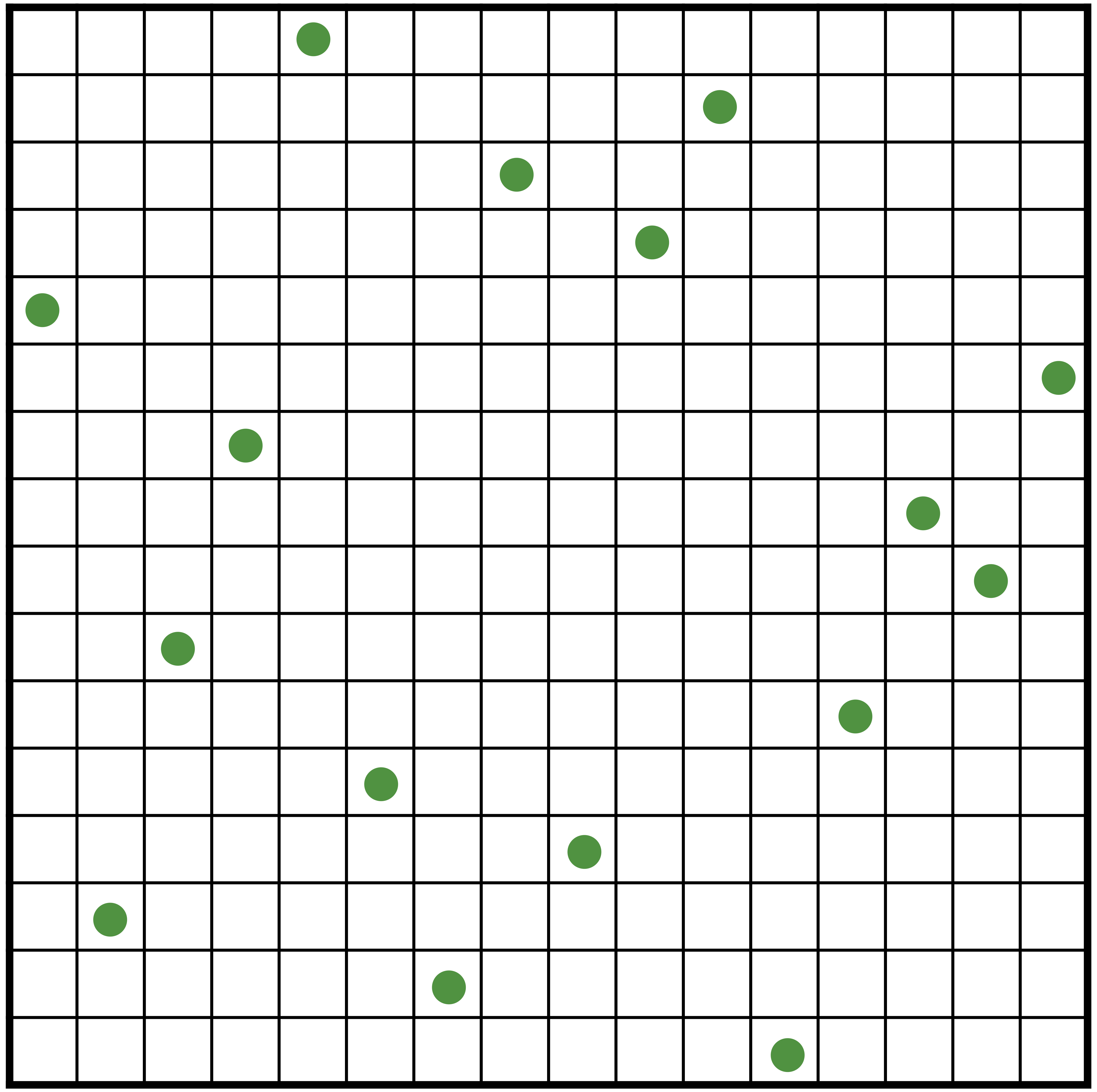


Latin Hypercube Sampler (N-rooks)

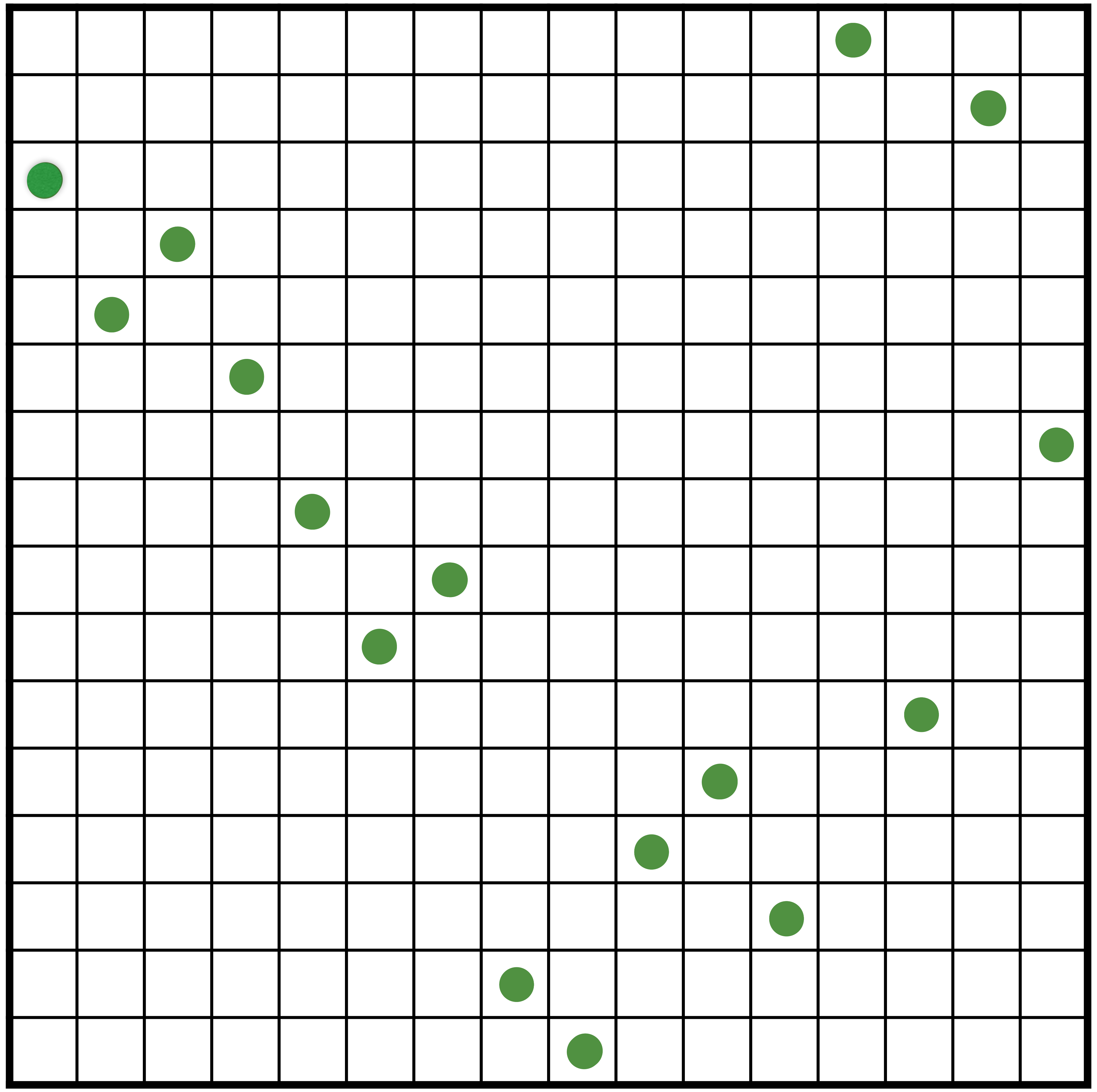


Latin Hypercube Sampler (N-rooks)

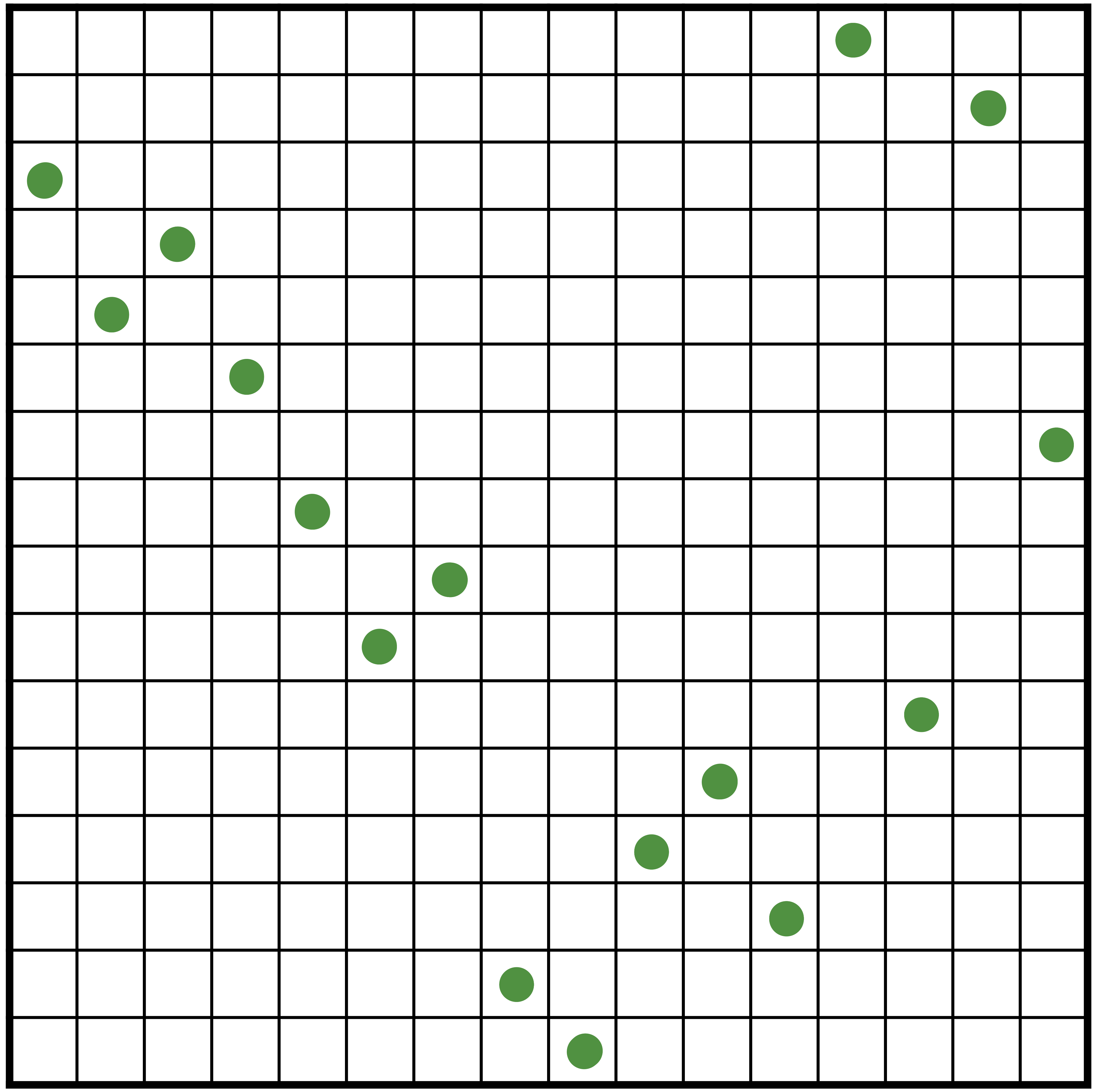
Shuffle columns



Latin Hypercube Sampler (N-rooks)



Latin Hypercube Sampler (N-rooks)



UNIVERSITÄT DES SAARLANDES

Variants of stratified sampling

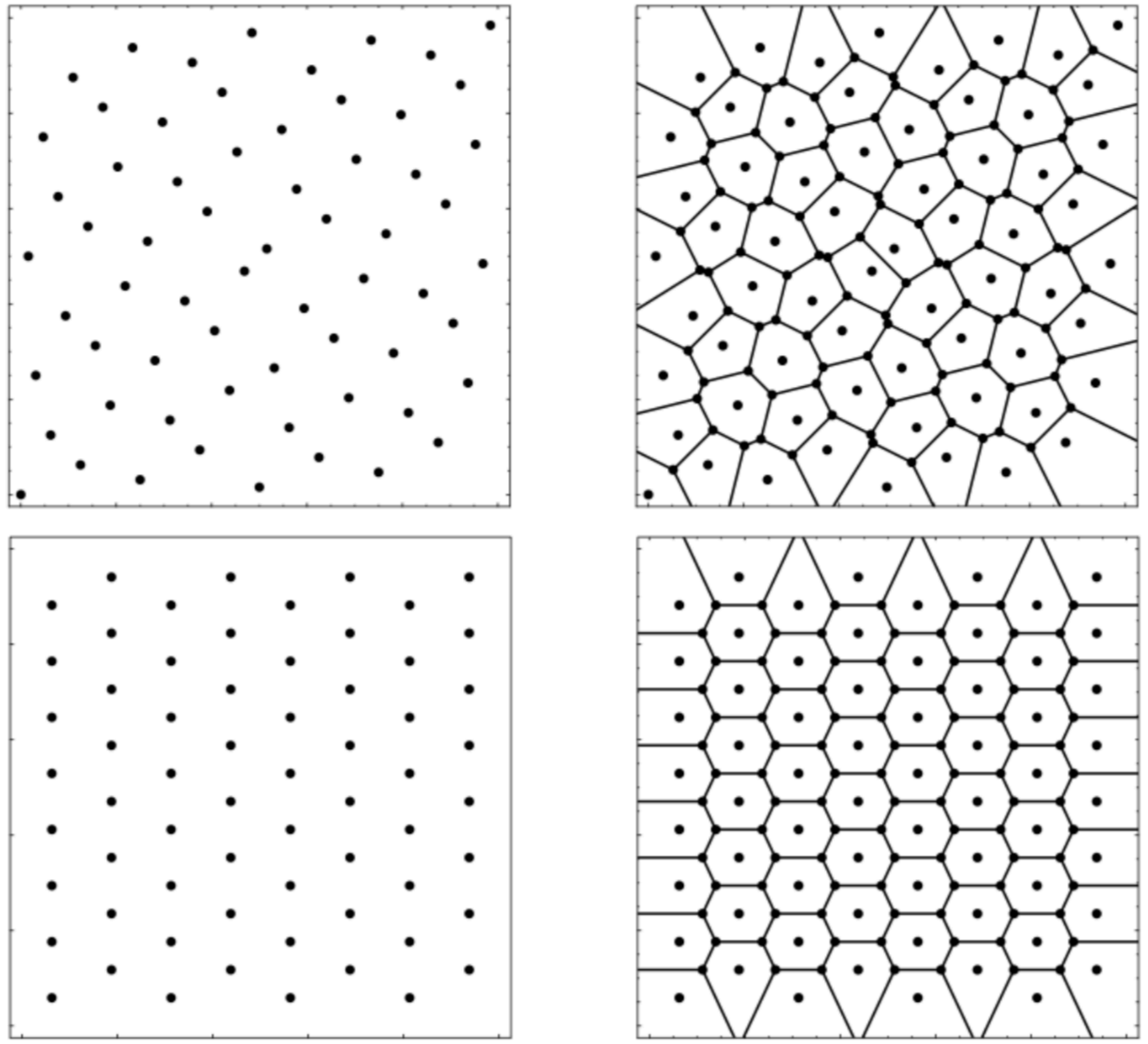


Figure 2.25: Stratification of I^2 with Voronoi diagrams. (a) 64-element Hammersley point set; (b) Voronoi diagram implied through (a); (c) 64-element hexagonal grid; (d) Voronoi diagram implied through (c).

Slide from Philipp Slusallek

Quasi-Monte Carlo Integration

Quasi-Monte Carlo Integration

- Monte Carlo integration suffers, apart from the slow convergence rate, from the disadvantages that only probabilistic statements on convergence and error boundaries are possible

- The success of any Monte Carlo procedure stands or falls with the quality of these random samples

- If the distribution of the sample points is not uniform then there are large regions where there are no samples at all, which can increases the error

- Closely related to this is the fact that a smooth function is evaluated at unnecessary many locations if samples are clumped

Quasi-Monte Carlo Integration

- Deterministic generation of samples, while making sure uniform distributions
- Based on number-theoretic approaches
- Samples with good uniform properties can be generated in very high dimensions.
- Sample generation is pretty fast: (almost) no pre-processing

Quasi-Monte Carlo Integration

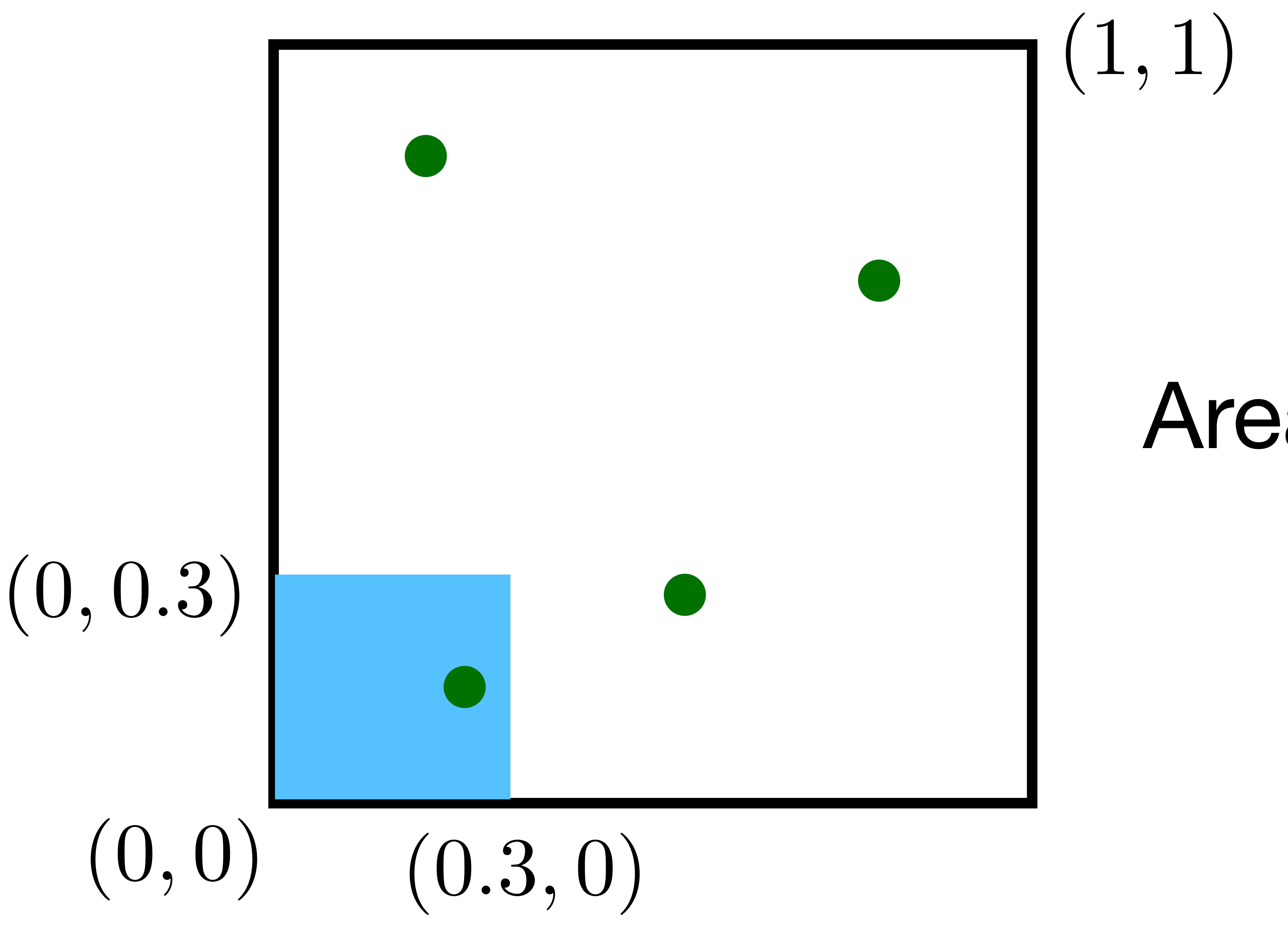
- Low discrepancy sequences
    - Halton and Hammerslay sequences
    - Scrambled sequences
- Discrepancy
- Koksma-Hlawka Inequality (later)

Discrepancy: Basic idea

- The concept of discrepancy can be viewed as a quantitative measure for the deviation of a given point set from a uniform distribution

Discrepancy: Basic idea

- The concept of discrepancy can be viewed as a quantitative measure for the deviation of a given point set from a uniform distribution



Area of the blue box: 0.09

Area associated to each sample: 0.25

Discrepancy: 0.25 - 0.09 = 0.16

Radical Inverse

Techniques based on a construction called as **radical inverse**

Any integer can be represented in the form:

$$n = \sum_{i=1}^{\infty} d_i b^{i-1}$$

n	Binary	$\Phi_b(n)$
1	1	
2	01	
3	11	
4	001	
5	101	

Radical Inverse

Techniques based on a construction called as **radical inverse**

Any integer can be represented in the form:

$$n = \sum_{i=1}^{\infty} d_i b^{i-1}$$

Radical inverse:

$$\Phi_b(n) = 0.d_1 d_2 \ldots d_m$$

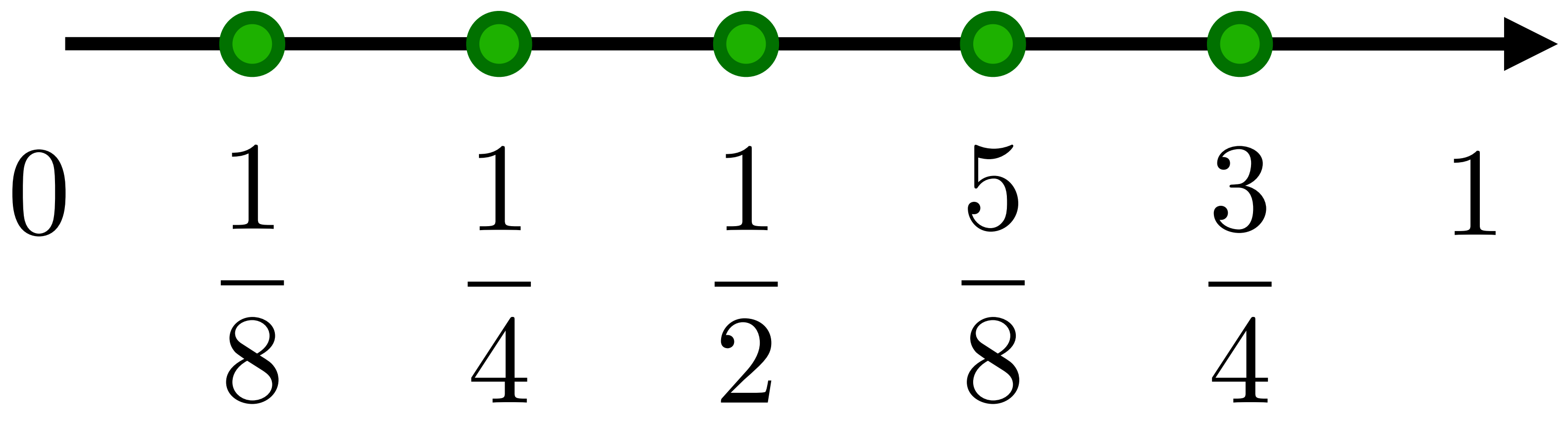
n	**Binary**	$\Phi_b(n)$
1	1	0.1
2	01	0.01
3	11	0.11
4	001	0.001
5	101	0.101

Radical Inverse

Techniques based on a construction called as **radical inverse**

Radical inverse:

$$\Phi_b(n) = 0.d_1 d_2 ... d_m$$



n	Binary	$\Phi_b(n)$
1	1	0.1 = 1/2
2	01	0.01 = 1/4
3	11	0.11 = 3/4
4	001	0.001 = 1/8
5	101	0.101 = 5/8

Halton and Hammerslay Sequence

Techniques based on a construction called as **radical inverse**

Radical inverse: $\Phi_b(n) = 0.d_1 d_2 ... d_m$

Halton Sequence: For n-dimensional sequence, we use different base b for each dimension

$$x_i = (\Phi_2(i), \Phi_3(i), \Phi_5(i), \ldots, \Phi_{p_n}(i))$$

Halton and Hammerslay Sequence

Techniques based on a construction called as **radical inverse**

Radical inverse: $\Phi_b(n) = 0.d_1 d_2 ... d_m$

Halton Sequence: For n-dimensional sequence, we use different base b for each dimension

$$x_i = (\Phi_2(i), \Phi_3(i), \Phi_5(i), \ldots, \Phi_{p_n}(i))$$

Hammerslay Sequence: All except the first dimension has co-prime bases

$$x_i = \left(\frac{i}{N}, \Phi_{b_1}(i), \Phi_{b_2}(i), \ldots, \Phi_{b_n}(i) \right)$$

Halton and Hammerslay Sequence

Techniques based on a construction called as **radical inverse**

Radical inverse: $\Phi_b(n) = 0.d_1 d_2 ... d_m$

Halton Sequence:

$$x_i = (\Phi_2(i), \Phi_3(i), \Phi_5(i), \ldots, \Phi_{p_n}(i))$$

Hammerslay Sequence:

$$x_i = \left(\frac{i}{N}, \Phi_{b_1}(i), \Phi_{b_2}(i), \ldots, \Phi_{b_n}(i)\right)$$

Hammerslay has slightly **lower** discrepancy than Halton

Halton 4spp

Halton 8spp


Sobol 4spp


Sobol 8spp


Random 8spp

Visualizing samples

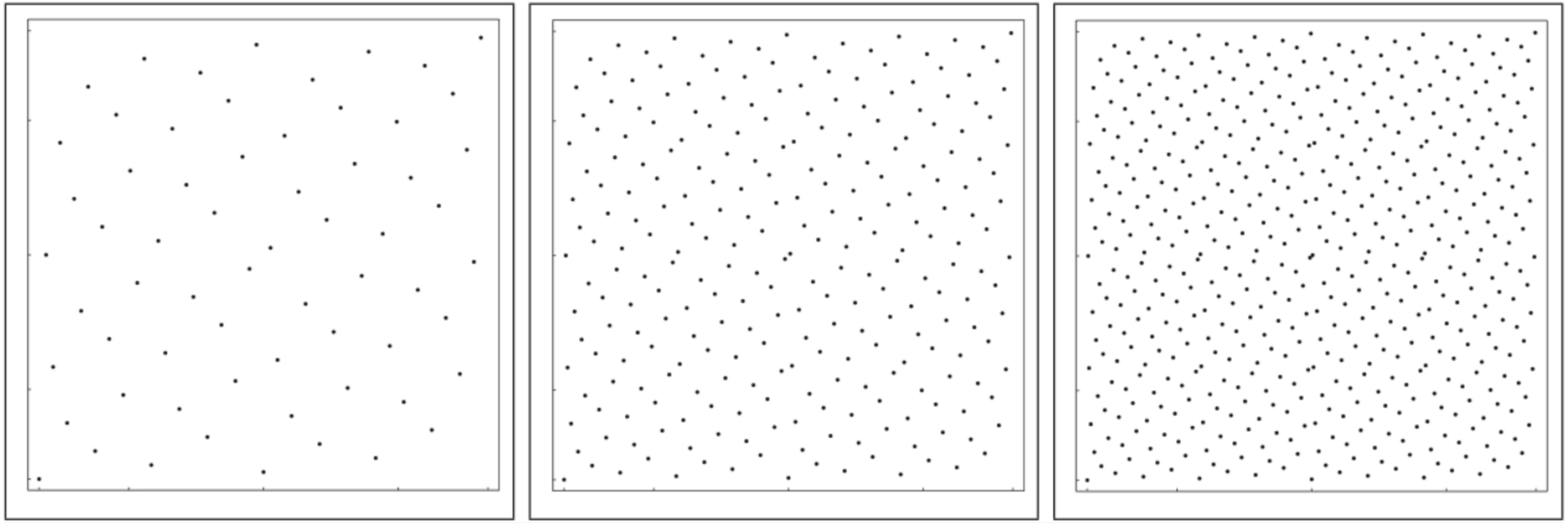


Figure 2.7: Hammersley Point Set on the 2D Plane. Three 2-dimensional Hammersley point sets $\mathbf{P}^2_{\mathrm{HAM}} = \left(\frac{i}{N}, \Phi_2(i)\right)_{i \in (0,\ldots,N-1)}$ of sizes $N = 64$-element, $N = 256$-element and $N = 512$-element.

Slide from Philipp Slusallek

Visualizing samples

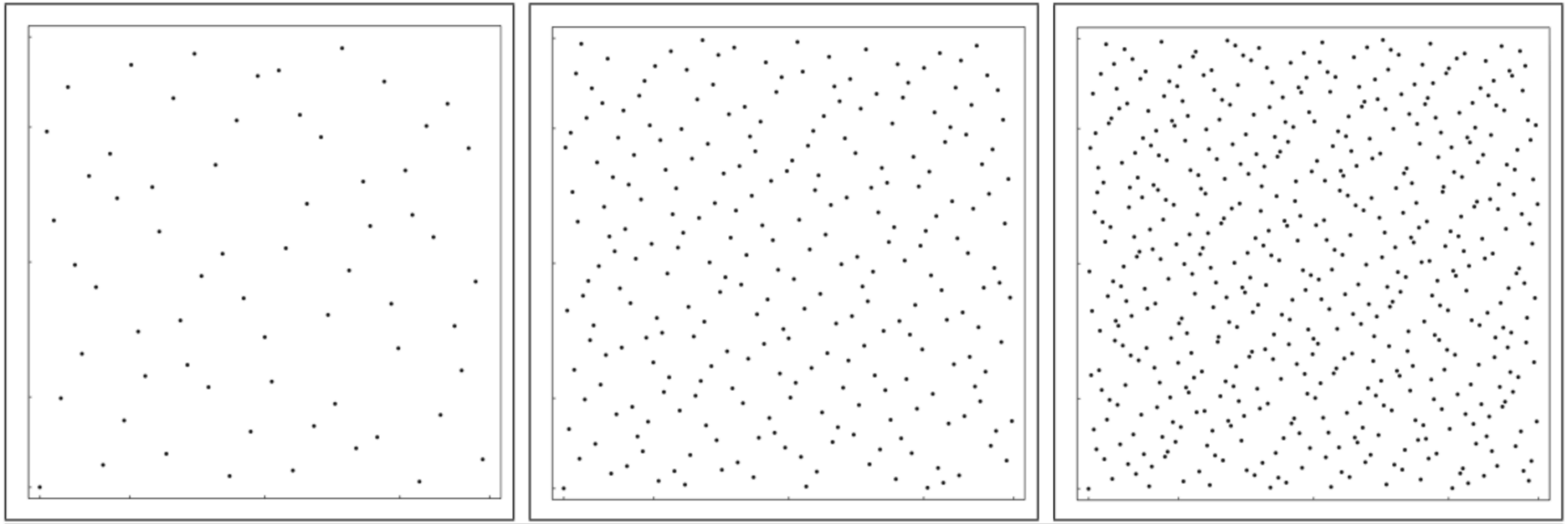
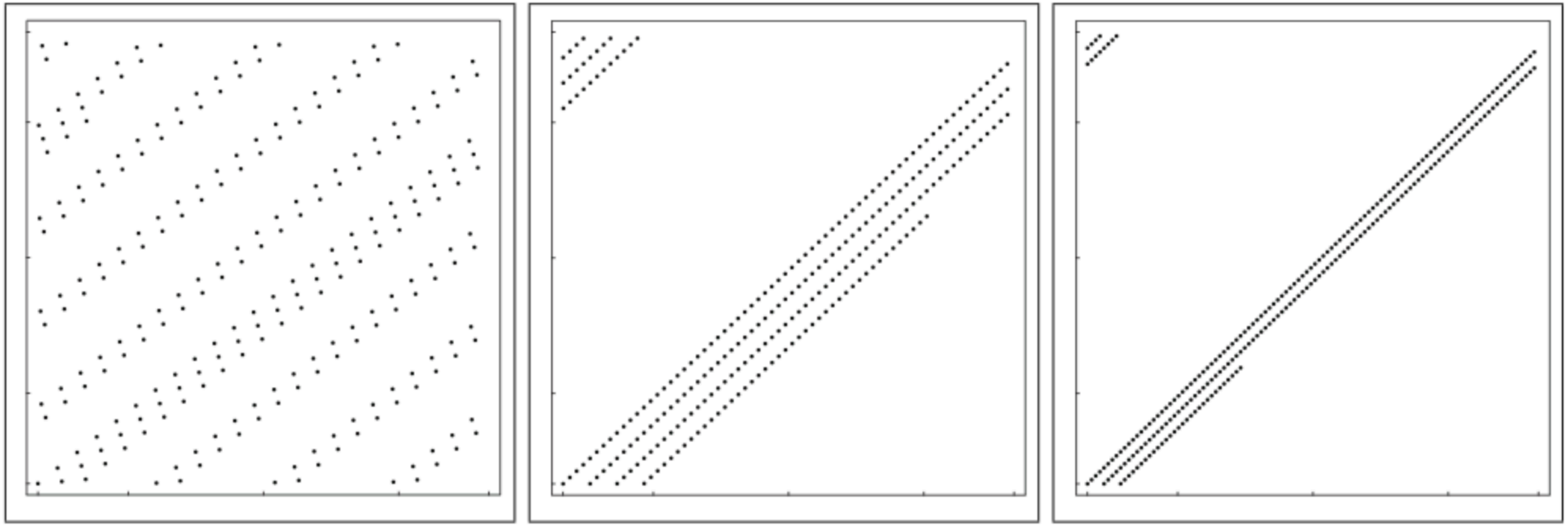


Figure 2.5: Halton sequence. The first 64, 256, and 512 points of the 2-dimensional Halton Sequence $\mathbf{P}^2_{\mathrm{HAL}} = (\Phi_2(i), \Phi_3(i))_{i \in \mathbb{N}_0}$.

Slide from Philipp Slusallek

Visualizing samples

Projection: (9,10) Projection: (19,20) Projection: (29,30)



Halton Sequence

Slide from Philipp Slusallek

Faure's permutation

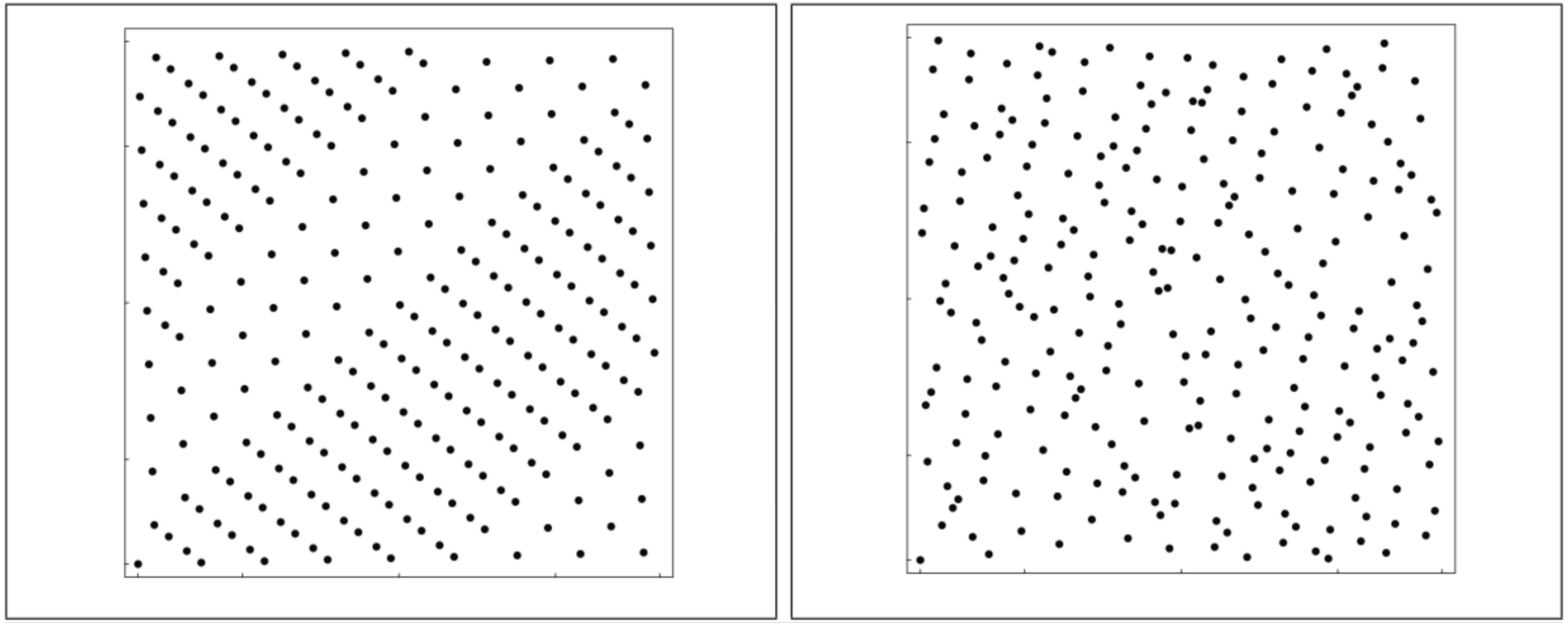


Figure 2.12: Halton Sequence and Scrambled Halton Sequence, Dimensions 7 and 8. **(a)** The first 256 elements of the 2-dimensional Halton sequence $\mathbf{P}^2_{\text{HAL}} = (\Phi_7(i), \Phi_8(i))$ and the scrambled versions of dimension 7 and 8 generated according to procedure of Faure.

Slide from Philipp Slusallek

Quasi-Monte Carlo Integration

- Low discrepancy sequences
  - Van der Corpus, Sobol sequences
  - (t,m,s)-nets & (t-s)-sequences

Discrepancy

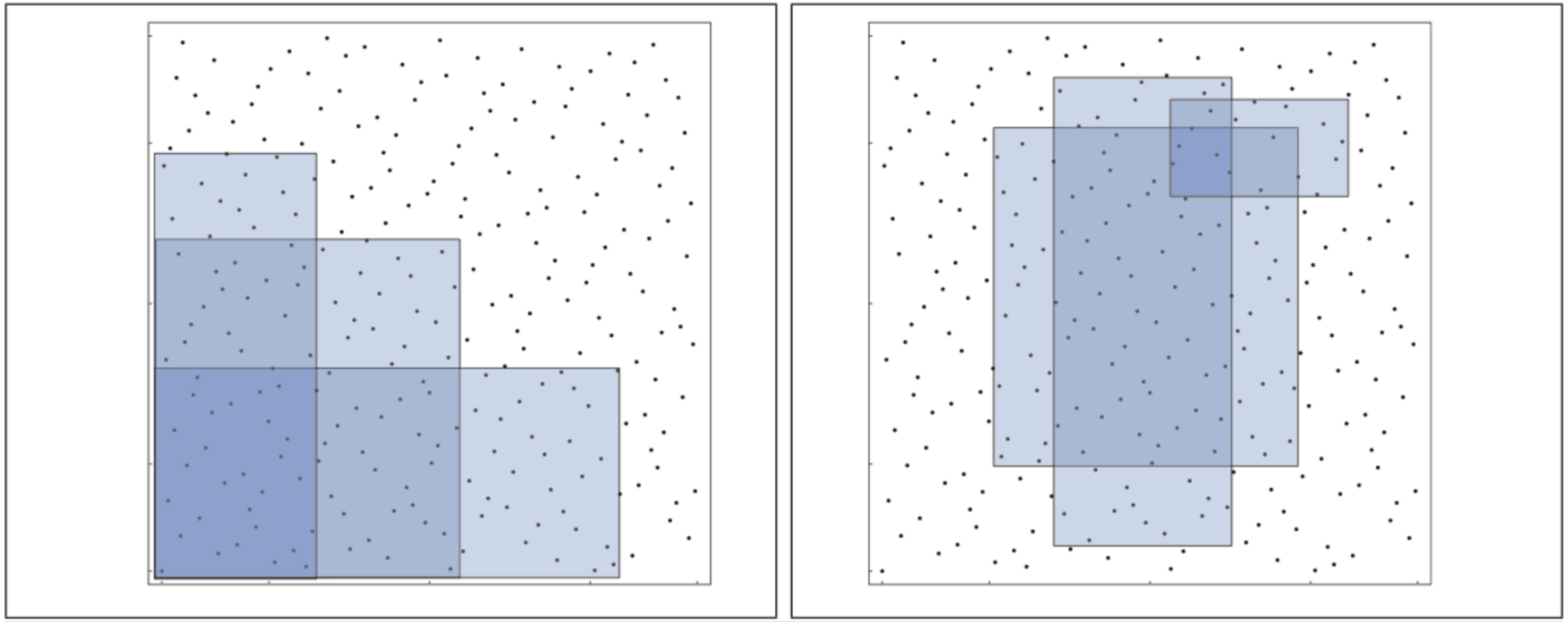


Figure 2.2: Star Discrepancy and Extreme Discrepancy. Visualization of the discrepancy concepts—case s=2—introduced in Definition 2.2. The star discrepancy based on axis-aligned 2-dimensional subareas of $\mathbf{I}^2$ attached at the origin, and the extreme discrepancy based on the choice of arbitrary 2-dimensional subvolumes of $\mathbf{I}^2$..

Slide from Philipp Slusallek

Discrepancy

DEFINITION 2.1 (Discrepancy) *Let* $\mathbf{P} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ *with* $\mathbf{x}_i \in \mathbf{I}^s, i = 1, \ldots, N$ *be a point set. The* discrepancy *of* $\mathbf{P}$*, denoted as* $D_N(\mathbf{P})$*, is a measure for the deviation of a point set from its ideal distribution. The discrepancy of* $\mathbf{P}$ *is defined as*

$$D_N(\mathbf{P}) \equiv D_N(\mathbf{P}, \mathcal{B}) \stackrel{\text{def}}{=} \sup_{\mathbf{B} \in \mathcal{B}} \left| \frac{\#(\mathbf{P} \cap \mathbf{B})}{N} - \mu^s(\mathbf{B}) \right|,$$

where $\mathcal{B}$ *corresponds to a Lebesgue measurable family of subsets of* $\mathbf{I}^s$*,* $\#$ *corresponds to the counting measure over* $\mathcal{B}$ *with respect to* $\mathbf{P}$*,* μ^s *is, as usual, the Lebesgue measure and* $\mathbf{B}$ *refers to a non empty subset of* $\mathcal{B}$*.*

Slide from Philipp Slusallek

Fourier Analysis: Samples Quality Measure

Advance Sampling Strategies: May 21, 2019