Realistic Image Synthesis

- Metropolis Algorithms -

Philipp Slusallek
Karol Myszkowski
Gurprit Singh

Motivation

- **Given**
  - An arbitrary $\mathbf{f}(\boldsymbol{x}) \to \mathbb{R}, \boldsymbol{x} \in \boldsymbol{\Omega}$
  - With an integral $I(f) = \int_{\Omega} f(x) d\Omega$
  - Ideally we could use the sampling distribution with zero variance
    - $f_{pdf} = f/I(f)$

- **Can we define an algorithm that**
  - Can generate samples $X = \{x_i\},\ \ x_i \sim f_{pdf}$
  - Without needing to compute $I(f)$ or f_{pdf}?

Introduction

- **The Metropolis-Hastings Algorithm**
  - Introduced in 1953 by Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller.
  - Initially designed for the Boltzmann distribution, and was later generalized and formalized by W.K. Hastings in 1970.
- **Main Features**
  - Allows to sample from probability distributions that are only known point-wise – and this, even if it is only up to a constant.
  - The theory behind it is related to Markov chains

Notation and Reminders

- **σ-Algebra**
  - χ: Set of states
  - $\mathcal{B}(\chi)$: σ-algebra over χ
    - $\chi \in \mathcal{B}(\chi)$
    - $\mathcal{B}(\chi)$ is stable under complementation
    - $\mathcal{B}(\chi)$ is stable under countable union.
  - **Informally:** σ-algebras have the properties you would expect for performing algebra on sets

- **Measures**
  - μ is a measure over $\mathcal{B}(\chi)$ iff:
    - $\mu(\emptyset) = 0$
    - $\forall B \in \mathcal{B}(\chi)$: $\mu(B) \geq 0$
    - For all countable collections of disjoint sets $\{E_i\}(i = 1 \rightarrow \infty)$,

$$\mu\left(\sum_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{\infty} \mu(E_k)$$

  - **Informally:** Measure functions have the properties you would expect for measuring sets

Background

- **Transition Kernel**
  - A transition kernel is a function $K(x, A)$ defined on $\chi \times \mathcal{B}(\chi)$ s.t.
    - $\forall \mathrm{x} \in \mathcal{B}(\chi), \quad \mathrm{K}(\mathrm{x}, \cdot)$ is a probability measure
    - $\forall \mathrm{A} \in \mathcal{B}(\chi), \quad \mathrm{K}(\cdot, A)$ is measurable
  - **Informally:** $K\,(x, A)$ is the probability of transitioning from x into the set of states A

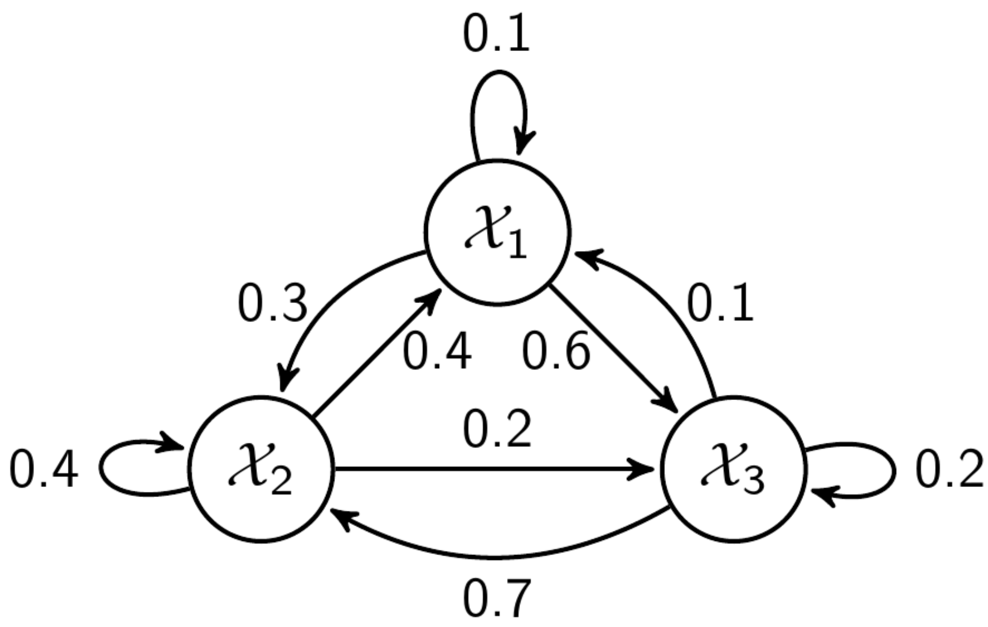
- **Example**
  - If $\chi = \{\chi_1, \dots, \chi_k\}$, the transition kernel is the following matrix:

$$K = \begin{pmatrix} P(X_n = \chi_1 | X_{n-1} = \chi_1) & \cdots & P(X_n = \chi_k | X_{n-1} = \chi_1) \\ \vdots & \ddots & \vdots \\ P(X_n = \chi_1 | X_{n-1} = \chi_k) & \cdots & P(X_n = \chi_k | X_{n-1} = \chi_k) \end{pmatrix}$$

  - Note that each row sums up to 1 since $\forall x, \sum_y P(y|x) = 1$

Background

- **Example**



$$K = \begin{pmatrix} 0.1 & 0.3 & 0.6 \\ 0.4 & 0.4 & 0.2 \\ 0.1 & 0.7 & 0.2 \end{pmatrix}$$

Background

- **Continuous Case**
  - If χ is continuous, we have:

$$P(X_n \in A|x) = \int_A K(x,y)dy$$

Background

- **Homogeneous Markov Chain**
  - An homogeneous Markov chain is a sequence (X_n) of random variables s.t.

$$\forall k\colon\ P(x_{k+1} \in A | x_0 x_1, \cdots, x_k) = P(x_{k+1} \in A | x_k) = \int_A K(x_k, x) dx$$

  - **Informally:** Each state in the chain only depends on the previous one

  - This definition implies that the construction of the chain is determined by an initial state x_0, and a transition kernel

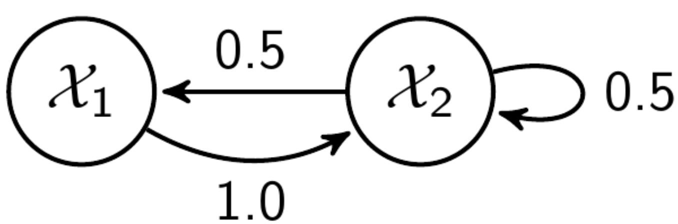
Background

- **Irreducibility**
  - The Markov chain (X_n) with transition kernel K is ϕ-irreducible iff:

$$\forall A \in \mathcal{B}(\chi) \text{ with } \phi(A) > 0, \exists n \text{ s.t. } K^n(x, A) > 0 \quad \forall x \in \chi$$

  - **Informally:** All states can communicate in a finite number of steps

- **Example**



$$K = \begin{pmatrix} 0.0 & 1.0 \\ 0.5 & 0.5 \end{pmatrix}$$

Background

- **Detailed Balance**
  - A Markov chain with transition kernel K satisfies the detailed balance condition if there exists a function f s.t.

$$\forall(x,y), K(x,y)f(x) = K(y,x)f(y)$$

  - Informally: Going from state x to state y has the same probability as going from y to x

Background

- **Stationary Distribution**
  - A probability measure π is a stationary distribution for the transition kernel K iff

$$\forall B \in \mathcal{B}(\chi), \qquad \pi(B) = \int_B K(x, B)\pi(x)dx$$

  - **Informally:** A transition leaves a stationary distribution unchanged
  - Under the condition of irreducibility, this distribution is unique up to a multiplicative constant.

Background

- **Theorem**
  - If a Markov chain with transition kernel K satisfies the with the pdf π, then π is the *stationary distribution* of the chain
- **Proof:**
  - Using the fact that

$$K(x,y)\pi(x) = K(y,x)\pi(y)$$

$$\begin{aligned}\int_Y K(y,B)\pi(y)dy &= \int_Y \int_B K(y,x)\pi(y)dxdy \\ &= \int_Y \int_B K(x,y)\pi(x)dxdy \\ &= \int_Y \pi(x) \int_B K(x,y)dy\,dx\end{aligned}$$

$$= \int_Y \pi(x)dx = \pi(Y)$$

Metropolis Sampling

- **Problem**
  - Sampling $X \sim f(x)$
    - When f can be *inversed analytically*, use inversion
    - When f is known *up to a constant*, use rejection sampling
    - When f is only known *point-wise & up to a constant*, what can we do?

Metropolis Sampling

- **The Metropolis-Hastings algorithm**
  - **Idea:** Construct an homogeneous Markov chain that converges to
  - the target distribution $f(x)$. Here, g is a function s.t. $g \sim f$.

Start from an initial state x_0, and $t = 0$
loop
  Choose a proposal sample $y_t \sim q(y \mid x_t)$
  Compute $a = \min\left(1, \frac{q(x_t|y_t)g(y_t)}{q(y_t|x_t)g(x_t)}\right)$
  Sample $u \sim U(0,1)$ // Uniform random number
  if u ≤ a **then**
    $x_{t+1} \leftarrow y_t$ // accept and use sample
  else
    $x_{t+1} \leftarrow x_t$ // reject and do nothing
  end if
  $t \leftarrow t+1$
end loop

Metropolis Sampling

- **Proposal distribution**
  - How to design the proposal distribution q?
  - Freedom in the choice of q as long as it follows some properties to ensure convergence
  - The two following conditions form a sufficient convergence criterion:
    - **Non-zero rejection probability**
      $P\left[f(X_t)q(Y_t\,|X_t) \leq f(Y_t)q(X_t|Y_t)\right] < 1$
    - **Strong irreducibility**
      $\forall(x,y), q(y|x) > 0$
  - When these conditions are met, the chain converges to the stationary distribution of the chain
  - But choosing a good q can be difficult

Metropolis Sampling

- **Convergence**
  - We can prove that
    - The kernel associated with the Markov chain generated by the algorithm satisfies the detailed balance with the target function f
    - This implies that f is a stationary distribution of the chain
    - Under the sufficient convergence conditions, the chain then converges to the distribution f

- **Key Messages**
  - The Metropolis-Hastings algorithm generates a Markov chain that converges to the distribution f
  - There is freedom in the choice of the proposal q as long as the convergence is ensured
  - The function f needs only be known point-wise & up to a constant

Metropolis Light Transport [Veach 1997]

- **Metropolis Light Transport (MLT)**
  - It works as a Markov chain process, which – in its steady state distribution – provides optimal importance sampling for any kind of function automatically!
  - But only over multiple samples (after warm-up phase)

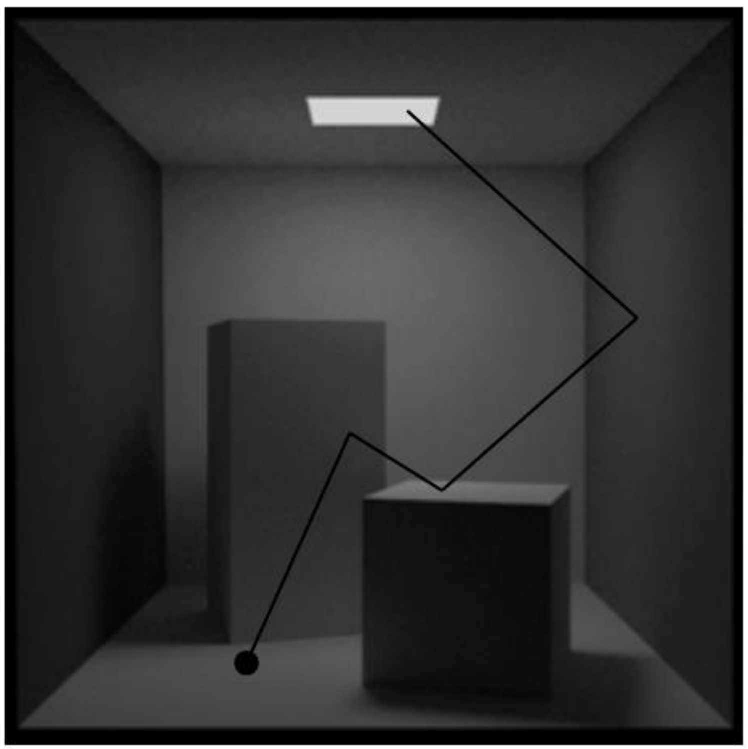
- **There are two main variants of MLT**
  - **Veach-type Metropolis:** It works in path-space, which means that it behaves differently based on the type of the sampled path (caustics, etc)
    - While this is a brilliant algorithm, it is fairly difficult to implement.
  - **Primary Sample Space Metropolis Light Transport (PSSMLT):** A simpler solution that retains the robustness of the original algorithm [Kelemen & Szirmay-Kalos, 2002]

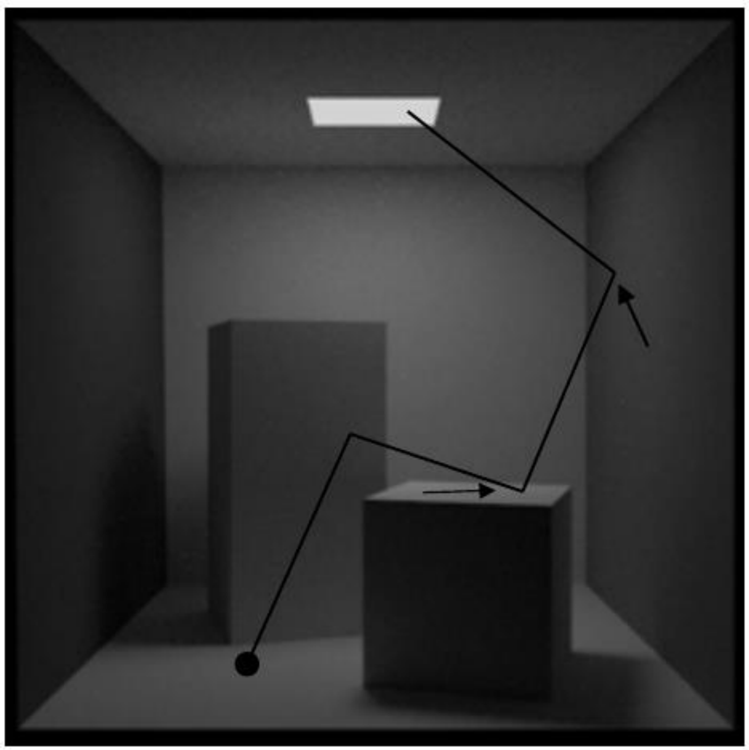
Metropolis Light Transport

- **General Idea**
  - Generate paths
  - Once a valid path is found, use it
  - Then *mutate* it to generate a new valid path
- **Advantages:**
  - Path re-use
  - Local exploration
  - **Insight:** Once you found a hard-to-find light distribution, try to stay in the neighborhood as it likely is also a “good” path

Metropolis Light Transport [Veach 1997]

- **Veach-Style MLT**
  - Once a path is found by PT/BPT/… mutate the path locally
  - Finding good local mutation and probabilities can be hard

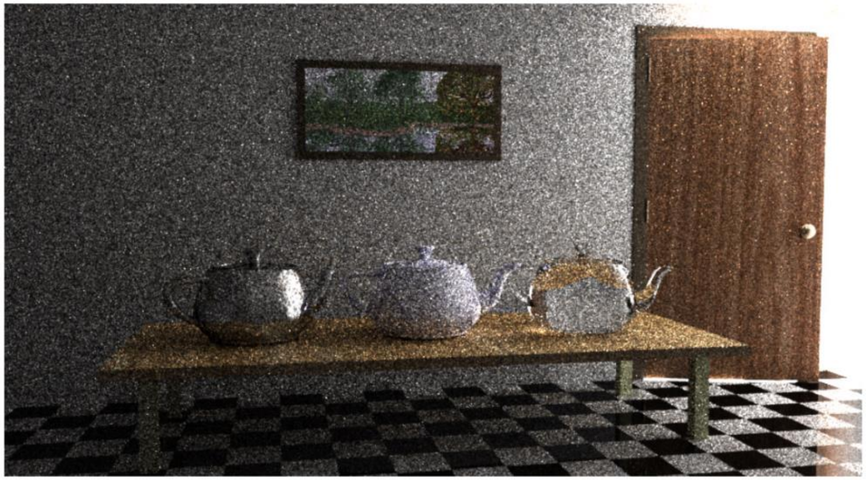


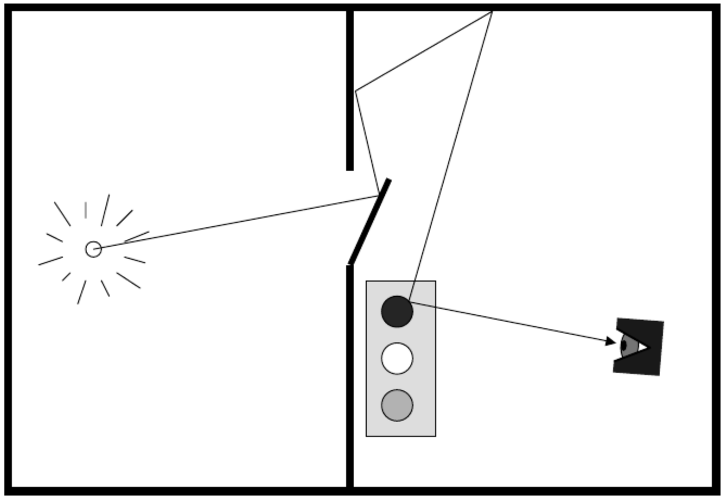


Metropolis Light Transport

- **Primary Sample Space Metropolis Light Transport (PSSMLT)**
  - General Idea: Every decision along a path is mapped into a n-dimensional *primary sample space* $[0,1]^n$
    - This includes decision to split the path, Russian roulette, etc.
  - “Mutations” are then essentially small perturbations in PSS

Metropolis Light Transport



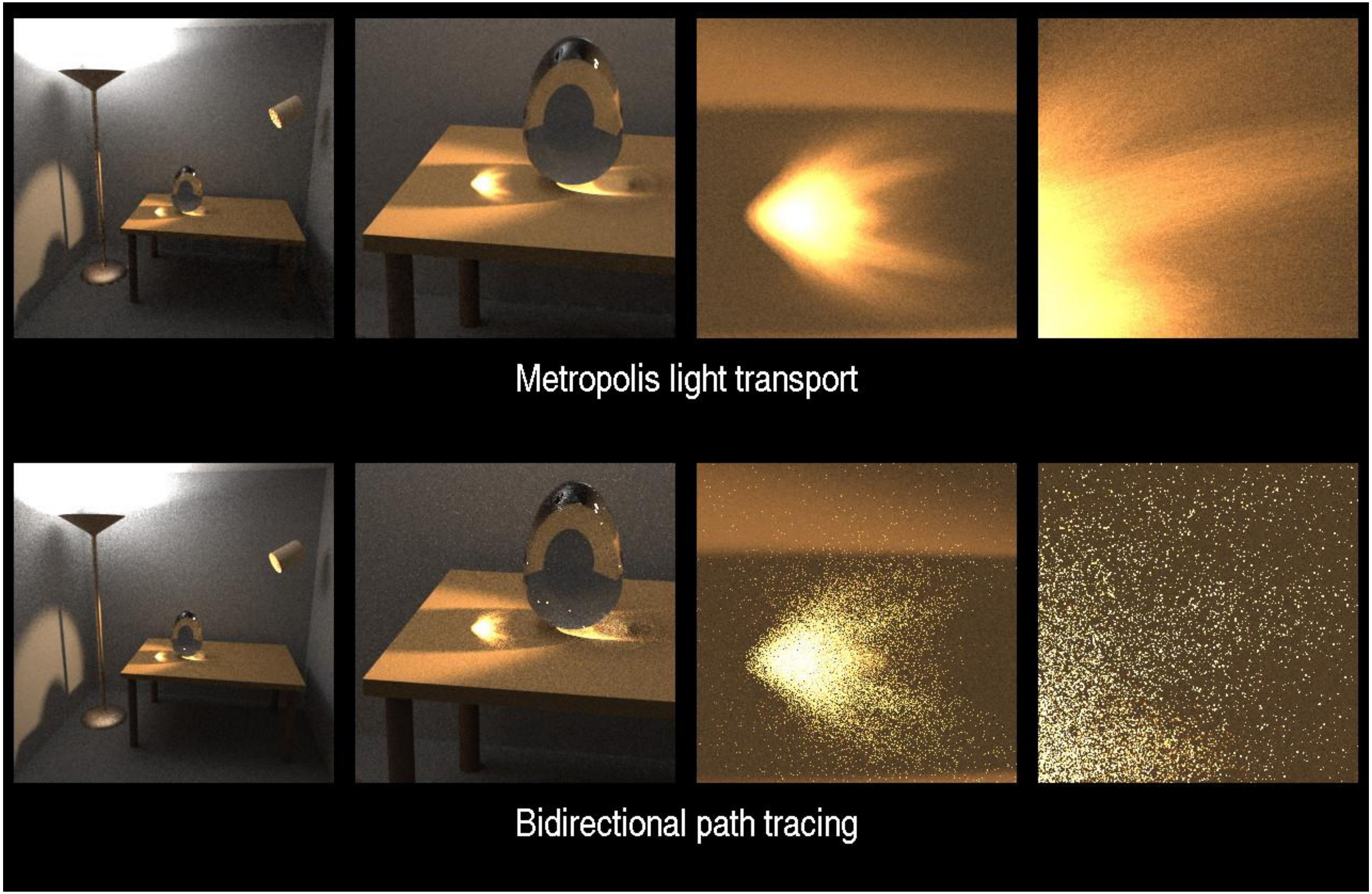


Also see: http://www.youtube.com/watch?v=AFJihgfocno

Metropolis Light Transport

- **Metropolis Light Transport**
  - It is the path tracing / bidirectional path tracing algorithm equipped with a smart sampling method, therefore...
    - Unbiased: Yes.
    - Consistent: Yes.
  - It is a *very robust* algorithm which is able to handle a variety of difficult light transport situations.
  - It is *not the easiest* algorithm to implement
  - It tends to be on the *slower* side
  - It can *converge unevenly*, getting stuck in difficult areas
  - Often *not helpful for easier scenes*, because a path tracer will outperform it, computing more samples per pixel in unit time

Example



Comparison BPT vs Metropolis

(a) Bidirectional path tracing with 40 samples per pixel.

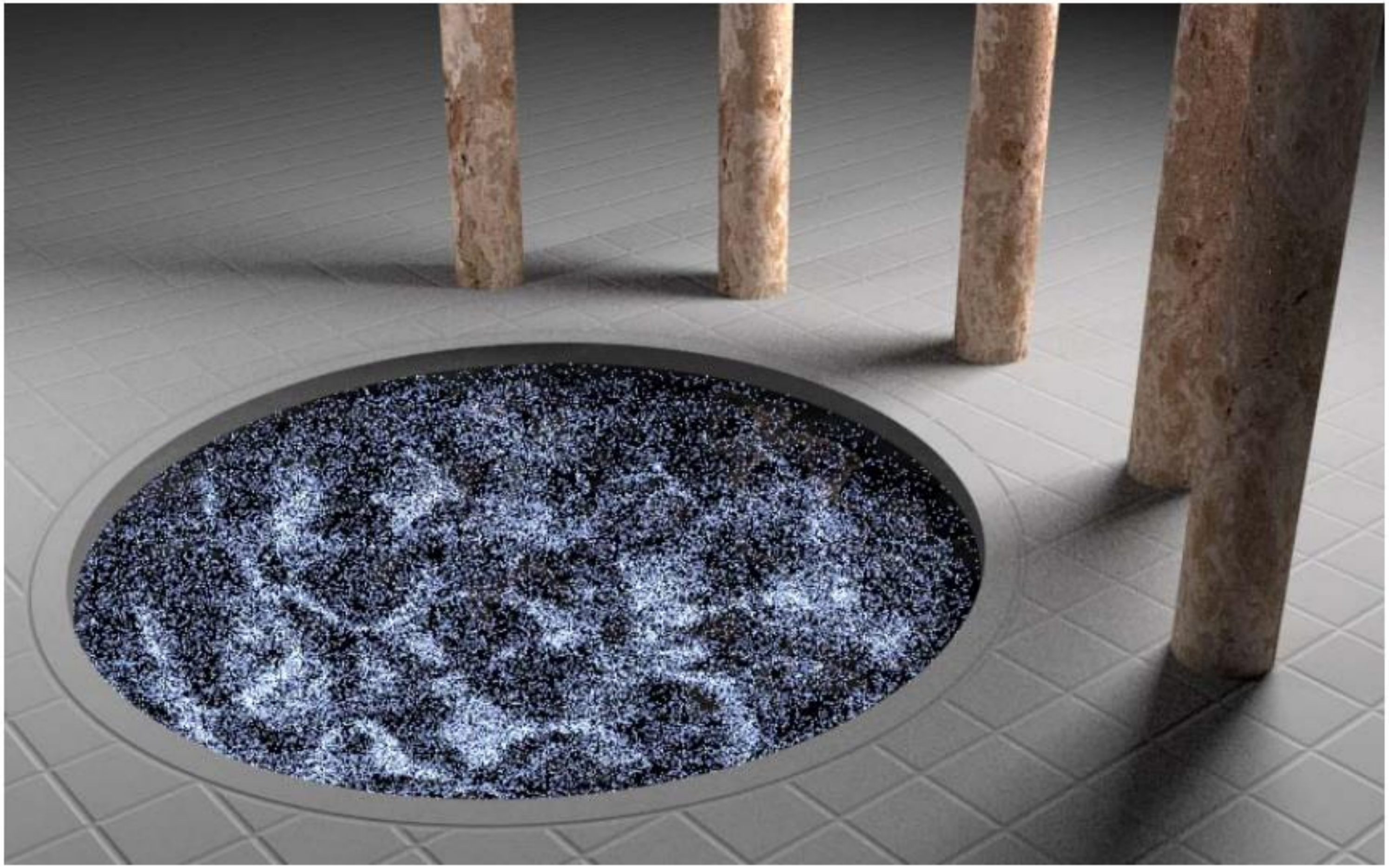
Image credit: Eric Veach

Comparison BPT vs Metropolis

(b) Metropolis light transport with an average of 250 mutations per pixel [the same computation time as (a)].

Image credit: Eric Veach

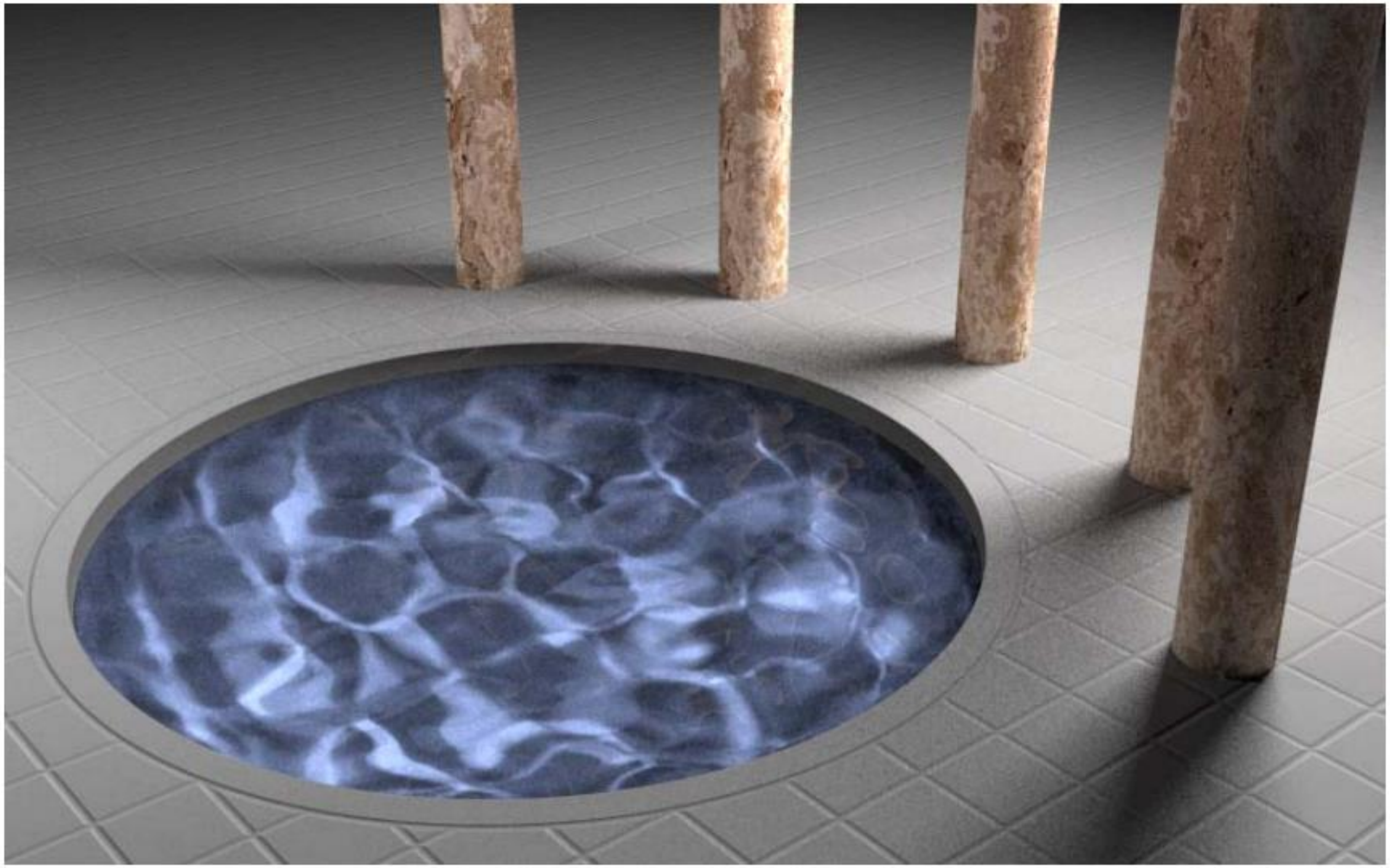
Comparison BPT vs Metropolis



(a) Path tracing with 210 samples per pixel.

Image credit: Eric Veach

Comparison BPT vs Metropolis



(b) Metropolis light transport with an average of 100 mutations per pixel [the same computation time as (a)].

Image credit: Eric Veach