Supplementary Material:
A Framework for Transient Rendering

Adrian Jarabo[1] Julio Marco[1] Adolfo Muñoz[1] Raul Buisan[1] Wojciech Jarosz[2] Diego Gutierrez[1]
[1]Universidad de Zaragoza [2]Disney Research Zurich

Contents This is the supplementary material for the paper *A Framework for Transient Rendering* [Jarabo et al. 2014]. It includes:

- The present document.
- A document including pseudocode of transient versions of the main algorithms used.
- The main Supplementary Video.

This document includes the following sections:

A Background

Here we introduce our notation and review the classic (steady state) path integral formulation, as well as the photon mapping algorithms and its progressive variant. The former will serve as the basis for the definition of *transient path integral* (Section 3), while the second closely relates with the *path reuse technique* described in Section 4, which as we will see, its crucial for transient rendering.

A.1 Path Integral

In the path integral formulation [Veach 1997; Pauly et al. 2000], the image pixel intensity I is computed as an integral over the space of light transport paths Ω:

$$I = \int_\Omega f(\overline{\mathbf{x}})\, \mathrm{d}\mu(\overline{\mathbf{x}}), \tag{S.1}$$

where $\overline{\mathbf{x}} = \mathbf{x}_0 \dots \mathbf{x}_k$ represents the spatial coordinates of the $k+1$ vertices of a length-k path with $k \geq 1$ segments. Vertex $\mathbf{x}_0$ lies on a light source, $\mathbf{x}_k$ lies on the camera sensor, and $\mathbf{x}_1 \dots \mathbf{x}_{k-1}$ are intermediate scattering vertices. The differential measure $\mathrm{d}\mu(\overline{\mathbf{x}})$ denotes area integration for surfaces vertices and volume integration for media vertices. The path contribution function $f(\overline{\mathbf{x}})$ is the product of the emitted radiance L_e, path throughput $\mathfrak{T}$, and sensor importance W_e:

$$f(\overline{\mathbf{x}}) = L_\mathrm{e}(\mathbf{x}_0 \to \mathbf{x}_1)\, \mathfrak{T}(\overline{\mathbf{x}})\, W_\mathrm{e}(\mathbf{x}_{k-1} \to \mathbf{x}_k). \tag{S.2}$$

The path throughput is itself the product of the scattering function ρ for the inner path vertices and the geometry G and visibility V terms for path segments:

$$\mathfrak{T}(\overline{\mathbf{x}}) = \left[\prod_{i=1}^{k-1} \rho(\mathbf{x}_i)\right]\left[\prod_{i=0}^{k-1} G(\mathbf{x}_i, \mathbf{x}_{i+1}) V(\mathbf{x}_i, \mathbf{x}_{i+1})\right]. \tag{S.3}$$

For a path segment $\mathbf{xy}$, we have $G(\mathbf{x},\mathbf{y}) = \frac{D(\mathbf{x}\to\mathbf{y})D(\mathbf{y}\to\mathbf{x})}{\|\mathbf{x}-\mathbf{y}\|^2}$, where $D(\mathbf{x}\to\mathbf{y}) = |n_\mathbf{x}\cdot\omega_\mathbf{xy}|$ if $\mathbf{x}$ is on a surface and $D(\mathbf{x}\to\mathbf{y}) = 1$ if $\mathbf{x}$ is in a medium, and likewise for $D(\mathbf{y}\to\mathbf{x})$. Here $n_\mathbf{x}$ is the surface normal at $\mathbf{x}$ and $\omega_\mathbf{xy}$ is a unit-length vector from $\mathbf{x}$ to $\mathbf{y}$. We assume that V is a *fractional* visibility function accounting for transmittance within media in addition to binary visibility for opaque objects. For path segment $\mathbf{xy}$, it is given by $V(\mathbf{x},\mathbf{y}) = \exp\left(-\int_0^{\|\mathbf{x}-\mathbf{y}\|} \sigma_t(\mathbf{x}+t\omega_\mathbf{xy})\,\mathrm{d}t\right)$ where σ_t is the extinction

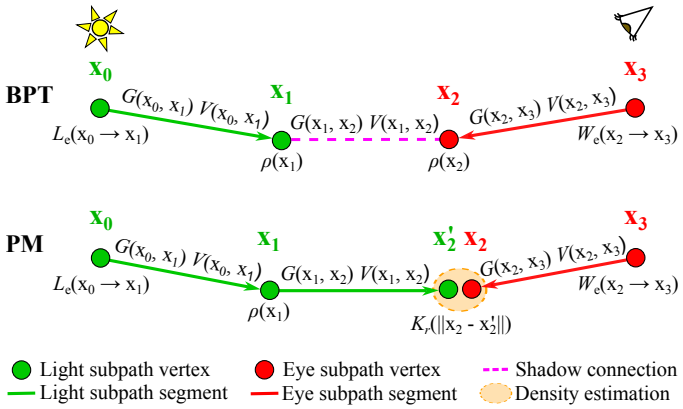

Figure S.1: *Schematic description of bidirectional path tracing (BPT, top) and photon mapping (PM, bottom). In both algorithms, an eye and a light subpath are traced from the eye and the light source respectively; these two subpaths are then connected to form a full path, via deterministic shadow connection in the case of bidirectional path tracing, and via an additional random segment and density estimation in photon mapping (Figure after Georgiev et al. [2012]).*

coefficient of the medium or a delta function at the boundary of opaque objects. The scattering kernel at each vertex is defined as:

$$\rho(\mathbf{x}_i) = \begin{cases} \rho_\mathrm{s}(\mathbf{x}_{i-1} \to \mathbf{x}_i \to \mathbf{x}_{i+1}) & \mathbf{x}_i \text{ on surface,} \\ \rho_\mathrm{p}(\mathbf{x}_{i-1} \to \mathbf{x}_i \to \mathbf{x}_{i+1})\sigma_s(\mathbf{x}_i) & \mathbf{x}_i \text{ in medium,} \end{cases} \tag{S.4}$$

where σ_s is the scattering coefficient in the medium, and ρ_s and ρ_p are the surface BSDF and phase function respectively.

Monte Carlo solutions approximate the path integral as a Monte Carlo estimator:

$$\langle I \rangle = \frac{1}{n} \sum_{j=1}^{n} \frac{f(\overline{\mathbf{x}}_j)}{p(\overline{\mathbf{x}}_j)}, \tag{S.5}$$

that averages the contribution of n random paths $\overline{\mathbf{x}}_j$, sampled with a probability distribution function (pdf) $p(\overline{\mathbf{x}}_j)$, which is given by the combined probability density of each of the vertex locations $p(\overline{\mathbf{x}}) = p(\mathbf{x}_0 ... \mathbf{x}_k)$. The probability density of the path is determined by the sampling technique used to obtain the path: for example, *bidirectional path tracing* (BPT) [Lafortune and Willems 1993; Veach and Guibas 1994] independently generates a subpath $\overline{\mathbf{x}}_w$ from the eye with pdf $p(\overline{\mathbf{x}}_w)$ and a subpath $\overline{\mathbf{x}}_l$ from the light with pdf $p(\overline{\mathbf{x}}_l)$. These are then (optionally) connected using a shadow ray to build the full path $\overline{\mathbf{x}}$ with pdf $p(\overline{\mathbf{x}}) = p(\overline{\mathbf{x}}_l)p(\overline{\mathbf{x}}_w)$ (see Figure S.1, top).

A.2 Photon Mapping

Photon mapping (PM) [Jensen 2001] is an efficient and robust two-pass global illumination algorithm. In the first pass (*photon tracing*), light subpaths are traced from the light sources, and their vertices hitting a diffuse surface are stored in a data structure called the photon map, which represents incoming flux.

In the second pass (*radiance estimation*, see Figure S.1,bottom), PM estimates the reflected radiance at a point $\mathbf{x}$ in direction ω_o using density estimation from the M nearest photons as:

$$\widehat{L_\mathrm{o}}(\mathbf{x}, \omega_\mathrm{o}) = \frac{1}{M} \sum_{j=1}^{M} K_R(\|\mathbf{x} - \mathbf{x}_j\|)\gamma_j \tag{S.6}$$

where K_R is the spatial smoothing kernel with bandwidth R, and $\mathbf{x}_j$ and γ_j are the position and contribution of photon j. Note that γ_j is a function of the light subpath of the photon $\overline{\mathbf{x}}_{l,j}$, and is computed as: $\gamma_j = \rho(\omega_j \to \mathbf{x} \to \omega_\mathrm{o}) L_\mathrm{e}(\mathbf{x}_{0,j} \to \mathbf{x}_{1,j}) \, \mathfrak{T}(\overline{\mathbf{x}}_{l,j}) / p(\overline{\mathbf{x}}_{l,j})$, with ω_j the incoming direction of the photon j. PM successfully handles difficult light paths, at the price of introducing bias. This means that the estimated radiance in Equation S.6 can be written as $\widehat{L_\mathrm{o}} = L_\mathrm{o} + \epsilon$, where L_o is the actual radiance and ϵ represents the error introduced by the density estimation. An important property of PM is that, although biased, it is consistent,

meaning that the bias vanishes in the limit using an infinite number of photons M and a kernel K_R with differential bandwidth $\mathrm{d}R$. This is obviously not a practical solution, so several improvements have been proposed over the last few years [Hachisuka et al. 2013].

Progressive Photon Mapping Progressive photon mapping (PPM) [Hachisuka et al. 2008; Hachisuka and Jensen 2009; Knaus and Zwicker 2011] is a multipass variant of photon mapping that allows handling difficult light paths without having to store an infinite number of photons in the photon map to eliminate bias. The method progressively traces photons, estimates radiance and discards photons, updating the results at each step while ensuring convergence in the limit. Since a new photon map is computed at each iteration, this is equivalent to performing different independent samples on the radiance. The approximated pixel measure $\langle I_n \rangle$ in pass n is computed as:

$$\langle I_n \rangle = \frac{1}{n} \sum_{j=1}^{n} \Psi_j \widehat{L_\mathrm{o}}(\mathbf{x}_j, \omega_{\mathrm{o},j}) \tag{S.7}$$

where $\Psi_j = \mathfrak{T}(\overline{\mathbf{x}}_{w,j})/p(\overline{\mathbf{x}}_{w,j}) W_\mathrm{e}(\mathbf{x}_{k,j} \to \mathbf{x}_{k-1,j})$ is the eye subpath contribution to $\mathbf{x}$.

Knaus and Zwicker [Knaus and Zwicker 2011] showed that this estimation is consistent if both the variance $\mathrm{Var}[\epsilon_n]$ and expected value $\mathrm{E}[\epsilon_n]$ of the error ϵ_n vanish as $n \to \infty$:

$$\mathrm{Var}[\epsilon_n] \to 0 \quad \Longrightarrow \quad \mathrm{Var}[\langle I_n \rangle] \to 0 \tag{S.8}$$

$$\mathrm{E}[\epsilon_n] \to 0 \quad \Longrightarrow \quad \mathrm{E}[\langle I_n \rangle] \to I. \tag{S.9}$$

To accomplish this, the bias $\mathrm{E}[\epsilon_j]$ is reduced in each iteration j by progressively reducing the bandwidth R of K_R, while allowing the $\mathrm{Var}[\epsilon_j]$ at each iteration to increase as:

$$\frac{\mathrm{Var}[\epsilon_{j+1}]}{\mathrm{Var}[\epsilon_j]} = \frac{R_j^2}{R_{j+1}^2} = \frac{j+1}{j+\alpha} \tag{S.10}$$

Note that this radius redution is valid for surfaces, for volumetric density estimation of photon points in media it is a function of R^3. The user parameter $\alpha \in (0, 1)$ controls how much the variance is allowed to increase in each iteration. This value determines the trade-off between the reduction of bias and radiance [Knaus and Zwicker 2011], and its choice has a dramatic effect on the convergence rate of the algorithm, as shown by Kaplanyan and Dachsbacher [2013].

B Progressive Temporal Density Estimation

In this section we analyze the behavior and convergence of the error of the kernel-based temporal density estimation described in Section 4. We first analyze the variance and expected error introduced by using the density estimation kernel with a fixed bandwidth (Section B.1). Based on these results, we then analyze the error and convergence rate of the progressive density estimation scheme, which allows to obtain a consistent estimation in the limit (Section B.2). Finally, we derive the parameters yielding optimal convergence with respect to the AMSE for the progressive approach (Section B.3).

B.1 Variance and expected error of density estimation

Following the recent probabilistic framework for the progressive photon mapping algorithm [Knaus and Zwicker 2011], here we analyze the variance and expected value of the error introduced by the temporal kernel-based density estimate at each iteration. This error ϵ is defined as the difference between the estimated pixel value $\langle I \rangle$ and the actual value I, at sensor point $\mathbf{x}$ and time t. Using the temporal kernel $K_\mathcal{T}$, with bandwidth $\mathcal{T}$, we have:

$$\epsilon = \frac{1}{n} \sum_{j=1}^{n} K_\mathcal{T}(\|t - t_j^-\|) \widehat{I}_j - I. \tag{S.11}$$

Variance In order to compute the variance of the error $\mathrm{Var}[\epsilon]$ we need to make a set of assumptions: First, we assume that the samples' probability density is constant within the kernel $K_\mathcal{T}$ in the temporal domain. We denote this probability as $p_\mathcal{T}(t)$. We also assume that the samples' time t_j and pixel measurement $\widehat{I}_j$ are independent samples of the random variables T and

$\widehat{I}$, respectively, where T has probability density $p_{\mathcal{T}}(t)$. Finally, we assume that the random variables T and $\widehat{I}$ are mutually independent. We thus model $\text{Var}[\epsilon]$ as:

$$\begin{aligned}\text{Var}[\epsilon] &= \text{Var}[\frac{1}{n}\sum_{j=1}^{n} K_{\mathcal{T}}(\|t-T\|)\widehat{I} - I] \\ &= \frac{1}{n}(\text{Var}[K_{\mathcal{T}}] + \text{E}[K_{\mathcal{T}}]^2) \\ &\quad (\text{Var}[\widehat{I}] + \text{E}[\widehat{I}]^2) - \frac{1}{M}\text{E}[K_{\mathcal{T}}]^2\text{E}[\widehat{I}]^2. \end{aligned} \tag{S.12}$$

Here $\text{E}[K_{\mathcal{T}}] = p_{\mathcal{T}}(t)$, while the variance introduced by the temporal kernel $\text{Var}[K_{\mathcal{T}}]$ has the form:

$$\text{Var}[K_{\mathcal{T}}] = \int_{\Omega_{\mathcal{T}}} K_{\mathcal{T}}(\|t-T\|)^2 p_{\mathcal{T}}(T)\,\mathrm{d}T - p_{\mathcal{T}}(t)^2 \tag{S.13}$$

where $\Omega_{\mathcal{T}}$is the area where $K_{\mathcal{T}}$ is compactly supported, with constant density of samples $p_{\mathcal{T}}(t)$. We express $K_{\mathcal{T}}$ as a canonical kernel $k_{\mathcal{T}}$ with unit integral such that $K_{\mathcal{T}}(\xi) = k_{\mathcal{T}}(\xi/\mathcal{T})\mathcal{T}^{-1}$, and perform the change of variable $\psi = (\xi - t)/\mathcal{T}$ and $\mathrm{d}\xi = \mathcal{T}\,\mathrm{d}\psi$:

$$\int_{\Omega_{\mathcal{T}}} K_{\mathcal{T}}(\|\xi - t\|)^2\,\mathrm{d}\xi = \int_{\mathbb{R}} \frac{1}{\mathcal{T}} k_{\mathcal{T}}(\psi)^2\,\mathrm{d}\psi, \tag{S.14}$$

which substituted in (S.13) allows us to define $\text{Var}[K_{\mathcal{T}}]$ as:

$$\text{Var}[K_{\mathcal{T}}] = \frac{p_{\mathcal{T}}(t)}{\mathcal{T}} \int_{\mathbb{R}} k_{\mathcal{T}}(\psi)^2\,\mathrm{d}\psi - p_{\mathcal{T}}(t)^2. \tag{S.15}$$

These transformations allow us to express (S.12) as:

$$\begin{aligned}\text{Var}[\epsilon] &= \frac{1}{M}(\text{Var}[\widehat{I}] + \text{E}[\widehat{I}]^2) \\ &\quad (\frac{p_{\mathcal{T}}(t)}{\mathcal{T}} \int_{\mathbb{R}} k_{\mathcal{T}}(\psi)^2\,\mathrm{d}\psi - p_{\mathcal{T}}(t)^2 + p_{\mathcal{T}}(t)^2) \\ &\quad -\frac{1}{M} p_{\mathcal{T}}(t)^2 \text{E}[\widehat{I}]^2 \\ &= \frac{1}{M}(\text{Var}[\widehat{I}] + \text{E}[\widehat{I}]^2)(\frac{p_{\mathcal{T}}(t)}{\mathcal{T}}\mathcal{C}_{\mathcal{T}}) \end{aligned} \tag{S.16}$$

where $\mathcal{C}_{\mathcal{T}}$ is a kernel-dependent constant. The last term can be neglected by assuming that the kernels cover small areas in their respective domains, which effectively means that $\mathcal{C}_{\mathcal{T}} \gg p_{\mathcal{T}}(t)$. Equation (S.16) shows that in transient density estimation, the variance $\text{Var}[\epsilon]$ is inversely proportional to $\mathcal{T}$.

Bias Bias at each iteration j is defined as the expected value of the error $\text{E}[\epsilon_j]$:

$$\begin{aligned}\text{E}[\epsilon_j] &= \text{E}[\frac{1}{M}\sum_{i=1}^{M} K_{\mathcal{T}}(\|t-T\|)\widehat{I} - I] \\ &= \text{E}[K_{\mathcal{T}}(\|t-T\|)]\text{E}[\widehat{I}] - I. \end{aligned} \tag{S.17}$$

As with the variance, we need to assume that the samples time and contribution can be interpreted as independent identically distributed random samples from the random variables T (with probability density $p_{\mathcal{T}}(t)$) and $\widehat{I}$ respectively.

The expected value of $K_{\mathcal{T}}(\|t-T\|)$ is described as:

$$\begin{aligned}\text{E}[K_{\mathcal{T}}] &= \int_{\Omega_{\mathcal{T}}} K_{\mathcal{T}}(\|t-T\|) p_{\mathcal{T}}(T)\,\mathrm{d}T \\ &= \int_{\mathbb{R}} \frac{1}{\mathcal{T}} k_{\mathcal{T}}(\|t-T\|/\mathcal{T}) p_{\mathcal{T}}(T)\,\mathrm{d}T. \end{aligned} \tag{S.18}$$

The second form of the equation is obtained by transforming $K_\mathcal{T}$ into a unit canonical form of the kernel $k_\mathcal{T}$ such that $K_\mathcal{T}(\xi) = k_\mathcal{T}(\xi/\mathcal{T})1/\mathcal{T}$. Assuming a locally uniform distribution $p_\mathcal{T}(\xi)$ (similar to modeling Var[ϵ]) is too restrictive to model the expected error accurately, since it leads to zero bias in the limit. Therefore, following previous work [Knaus and Zwicker 2011], we use a Taylor expansion of $p_\mathcal{T}(\xi)$:

$$p_\mathcal{T}(\xi) = p_\mathcal{T}(t) + (\xi - t)\bigtriangledown p_\mathcal{T}(t) + O(\|\xi - t\|^2). \tag{S.19}$$

We plug this expression into (S.18), and apply the changes of variable $\psi = (T - t)/\mathcal{T}$ and $\mathrm{d}T = \mathcal{T}\,\mathrm{d}\psi$, to get:

$$\begin{aligned}
\mathrm{E}[K_\mathcal{T}] &= \frac{1}{\mathcal{T}}\int_\mathbb{R} k_\mathcal{T}(\psi)(p_\mathcal{T}(t) + \mathcal{T}\psi\bigtriangledown p_\mathcal{T}(t) + O(\|\mathcal{T}\psi\|^2))\mathcal{T}\,\mathrm{d}\psi \\
&= p_\mathcal{T}(t)\int_\mathbb{R} k_\mathcal{T}(\psi)\,\mathrm{d}\psi + \mathcal{T}\bigtriangledown p_\mathcal{T}(t)\int_\mathbb{R} k_\mathcal{T}(\psi)\psi\,\mathrm{d}\psi \\
&\quad + \mathcal{T}^2\int_\mathbb{R} k_\mathcal{T}(\psi)O(\|\psi\|^2)\,\mathrm{d}\psi \\
&\approx p_\mathcal{T}(t) + \mathcal{T}^2\int_\mathbb{R} k_\mathcal{T}(\psi)O(\|\psi\|^2)\,\mathrm{d}\psi \\
&= p_\mathcal{T}(t) + \mathcal{T}^2\mathcal{C}^{ii}_\mathcal{T}.
\end{aligned} \tag{S.20}$$

This means that bias due to radiance estimation in the temporal domain is inversely proportional to $\mathcal{T}^2$ and a constant $\mathcal{C}^{ii}_\mathcal{T}$ dependent on the high-order derivatives of the probability densities. We apply the last approximation by observing that in most common scenarios $\int_\mathbb{R} k_\mathcal{T}(\psi)\psi\,\mathrm{d}\psi \approx 0$. Given this approximation, and using $I = p_\mathcal{T}(t)\mathrm{E}[\widehat{I}]$, we can formulate the expected value $\mathrm{E}[K_\mathcal{T}]$ as:

$$\begin{aligned}
\mathrm{E}[\epsilon_j] &\approx (p_\mathcal{T}(t) + \mathcal{T}^2\mathcal{C}^{ii}_\mathcal{T})\mathrm{E}[\widehat{I}] - p_\mathcal{T}(t)\mathrm{E}[\widehat{I}] \\
&= \mathrm{E}[\widehat{I}]\mathcal{T}^2\mathcal{C}^{ii}_\mathcal{T}.
\end{aligned} \tag{S.21}$$

B.2 Variance and expected error of the pixel estimate

Here we derive the variance and expected error of the pixel estimate $\langle I_n\rangle$ after the n-th pass of the progressive algorithm, as modeled in Equation (7). The sampled variable is the time instant t_j where the estimation is being computed. As previously, we assume that they are independent identically distributed random samples.

Variance Assuming the random variable ϵ_j, we model the variance of the estimator Var[$\langle I_n\rangle$] as [Knaus and Zwicker 2011]:

$$\mathrm{Var}[\langle I_n\rangle] = \frac{1}{n}\mathrm{Var}[I] + \frac{1}{n^2}\sum_{j=1}^{n}\mathrm{Var}[\epsilon_j]. \tag{S.22}$$

The first term is the usual Monte Carlo estimator, which vanishes with $O(n^{-1})$. The other three terms, however, are functions of the error ϵ_j. Var[ϵ_n] is the variance of the average error, modeled as:

$$\mathrm{Var}[\epsilon_n] = \sum_{j=1}^{n}\frac{1}{n}\mathrm{Var}[\epsilon_j] = \frac{1}{n^2}\sum_{j=1}^{n}\mathrm{Var}[\epsilon_j]. \tag{S.23}$$

In oder to achieve consistency, we allow the variance of the expected error to increase at each iteration by a factor (S.10):

$$\frac{\mathrm{Var}[\epsilon_{j+1}]}{\mathrm{Var}[\epsilon_j]} = \left(\frac{j+1}{j+\alpha}\right). \tag{S.24}$$

We can model Var[ϵ_n] as a function of the variance at the first iteration Var[ϵ_1] as:

$$\mathrm{Var}[\epsilon_n] = \frac{\mathrm{Var}[\epsilon_1]}{n^2}(1 + \sum_{j=2}^{n} j\alpha\mathrm{B}(\alpha, j)), \tag{S.25}$$

where $\mathrm{B}(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$ is the Beta function, and $\Gamma(n) = (n-1)!$ is the Gamma function. Using the approximation proposed by Kaplanyan and Dachsbacher [2013], we get:

$$\mathrm{Var}[\epsilon_n] \approx \frac{\mathrm{Var}[\epsilon_1]}{(2-\alpha)n^{\alpha}} = O(n^{-\alpha}). \tag{S.26}$$

Finally, using this formulation of the variance of the average error $\mathrm{Var}[\epsilon_n]$ and asypmtotic simplifications, we can formulate $\mathrm{Var}[\langle I_n \rangle]$ (S.22) as:

$$\begin{aligned} \mathrm{Var}[\langle I_n \rangle] &\approx \frac{1}{n}\mathrm{Var}[I] + \mathrm{Var}[\epsilon_n] \\ &\approx \frac{1}{n}\mathrm{Var}[I] + \frac{\mathrm{Var}[\epsilon_1]}{(2-\alpha)n^{\alpha}} \\ &= O(n^{-1}) + O(n^{-\alpha}) = O(n^{-\alpha}). \end{aligned} \tag{S.27}$$

Expected error The expected value of of the estimator $\mathrm{E}[\langle I_n \rangle]$ is modeled as:

$$\begin{aligned} \mathrm{E}[\langle I_n \rangle] &= \mathrm{E}[\frac{1}{n}\sum_{j=1}^{n}(I_j + \epsilon_j)] \\ &= \frac{1}{n}\sum_{j=1}^{n}\mathrm{E}[I_j] + \frac{1}{n}\sum_{j=1}^{n}\mathrm{E}[\epsilon_j] \\ &= I + \mathrm{E}[\epsilon_n] \end{aligned} \tag{S.28}$$

where $\mathrm{E}[\epsilon_n]$ is the bias of the estimator after n steps:

$$\mathrm{E}[\epsilon_n] = \frac{1}{n}\sum_{j=1}^{n}\mathrm{E}[\epsilon_j], \tag{S.29}$$

and $\mathrm{E}[\epsilon_j]$ is the expected value of the error at each pass, described in Equation (S.21).

Following Equation (S.10) we compute $\mathcal{T}_j$ as a function of its initial value $\mathcal{T}_1$ as:

$$\begin{aligned} \mathcal{T}_j &= \mathcal{T}_1 \prod_{k=1}^{j-1}\left(\frac{k+\alpha}{k+1}\right) \\ &= \mathcal{T}_1 \left(\frac{(\alpha+1)_{j-1}}{\Gamma(j-1)}\right) \\ &= \mathcal{T}_1 \left(\frac{\Gamma(\alpha+j)}{\Gamma(\alpha+1)\Gamma(j+1)}\right) \\ &= \mathcal{T}_1 \left(\frac{\Gamma(\alpha+j)}{j\cdot\alpha\cdot\Gamma(\alpha)\Gamma(j)}\right) \\ &= \mathcal{T}_1 (j\cdot\alpha\cdot\mathrm{B}(\alpha,j))^{-1} \end{aligned} \tag{S.30}$$

where $\Gamma(n)$ is the Gamma function, $\mathrm{B}(x,y)$ is the Beta function, and $(x)_n$ is the Pochhammer symbol $(x)_n = x(x-1)(x-2)..(x-n+1)$.

Using (S.30) we can express $\mathrm{E}[\epsilon_j]$ (S.21) as a function of the initial kernel bandwidths:

$$\mathrm{E}[\epsilon_j] = \mathcal{T}_1^2 (j\cdot\alpha\cdot\mathrm{B}(\alpha,j))^{-2}\mathcal{C}_{\mathcal{T}}^{ii} \tag{S.31}$$

As noted by Knaus and Zwicker [2011], we can use Stirling's formula to get the asymptotic approximation of $\mathcal{T}_j$, which allows us to express $\mathrm{E}[\epsilon_j]$ in asymptotic notation:

$$\mathrm{E}[\epsilon_j] = \mathcal{C}_{\mathcal{T}}^{ii}\mathcal{T}_1^2\Theta(j^{1-\alpha})^{-2} \tag{S.32}$$

Finally, we use $\sum_{j=1}^{n}\Theta(j^x) = nO(n^x)$ to plug Equation (S.32) into Equation (S.29) to get the asymptotic behavior of $\mathrm{E}[\epsilon_n]$:

$$\begin{aligned}\mathrm{E}[\epsilon_n] &= \frac{1}{n}\mathcal{C}_{\mathcal{T}}^{ii}\mathcal{T}_1^2 nO(n^{1-\alpha})^{-2} \\ &= O(n^{1-\alpha})^{-2}, \end{aligned} \tag{S.33}$$

Inserting this last equality into Equation (S.28) allows us to compute the asymptotic form of $\mathrm{E}[\langle I_n\rangle]$ as:

$$\mathrm{E}[\langle I_n\rangle] = O(n^{1-\alpha})^{-2}. \tag{S.34}$$

B.3 Minimizing the Asymptotic Mean Squared Error

As shown above, the convergence rates of both the variance $\mathrm{Var}[\langle I_n\rangle]$ and the expected error $\mathrm{E}[\langle I_n\rangle]$ depend on the value assigned to the parameter α. Ideally, we want to obtain the parameter that allows reducing faster the total error. We measure the total error using the asymptotic mean squared error (AMSE), defined in Equation (8); using the obtained values for variance $\mathrm{Var}[\langle I_n\rangle]$ and bias $\mathrm{E}[\langle I_n\rangle]$, it becomes:

$$AMSE(\langle I_n\rangle) = O(n^{-\alpha}) + O(n^{1-\alpha})^{-4}. \tag{S.35}$$

By finding the derivate of Equation (S.35) and equating to zero we get the optimal parameter $\alpha = 4/5$, which leads to the optimal convergence rate of the AMSE of our transient density estimate:

$$AMSE(\langle I_n\rangle) = O(n^{-\frac{4}{5}}) + O(n^{1-\frac{4}{5}})^{-4} = O(n^{-\frac{4}{5}}) \tag{S.36}$$

C Transient Progressive Photon Mapping

Here we describe our transient formulation of progressive photon mapping (PPM) [Hachisuka et al. 2008; Hachisuka and Jensen 2009]. We first give an overview of the algorithm, describing the spatio-temporal smoothing kernel used, the progressive approach followed to vanish error in the limit, and the behaviour of the algorithm (Section C.1). Then, we analyze the variance and expected error introduced by the spatio-temporal density estimation for a given iteration j (Section C.2), and use these results to derive the asymptotic convergence rate after n iterations (Section C.3). Finally, we detail the derivation of the optimal parameters for higher asymptotic convergence with respect to the AMSE (Section C.4).

C.1 Algorithm

To include the temporal domain in a photon mapping framework, we need to take into account the time delays described in Equation (4), while adding a temporal smoothing kernel $K_{\mathcal{T}}$ in the density estimation, similar to Cammarano and Jensen [2002]. This results in an approximation of the radiance $\widehat{L_{\mathrm{o}}}(\mathbf{x}, t)$ as:

$$\widehat{L_{\mathrm{o}}}(\mathbf{x}, t) = \frac{1}{M}\sum_{i=1}^{M} K(\|\mathbf{x} - \mathbf{x}_i\|, \|t - t_i^-\|)\gamma_i, \tag{S.37}$$

where $K(\mathbf{x}, t) = K_R(\mathbf{x}) \cdot K_{\mathcal{T}}(t)$ and K_R is the spatial smoothing kernels with bandwidth R. We decompose K into K_R and $K_{\mathcal{T}}$ since they operate in separate domains, which might result into different kernel types; this is a typical approach in multivariate density estimation [Scott 1992].

In our transient framework, Equation (S.7) now becomes:

$$\langle I_n\rangle = \frac{1}{n}\sum_{j=1}^{n}\Psi_j\widehat{L_{\mathrm{o}}}(\mathbf{x}_j, t_j). \tag{S.38}$$

where $\Psi_j = \mathfrak{T}(\overline{\mathbf{x}}_{w,j})/p(\overline{\mathbf{x}}_{w,j})W_{\mathrm{e}}(\mathbf{x}_{k,j} \to \mathbf{x}_{k-1,j})$ is the importance of the measurement, which is the contribution of the eye subpath $\overline{\mathbf{x}}_{w,j}$. Unfortunately, although variance is reduced, using this estimator introduces bias in the final solution due to the expected error ϵ_j in each iteration, which depends on the size of the kernel estimation (as seen in Section A). This means that the estimate of the pixel $\langle I_n\rangle$ is the sum of the actual value of the pixel I and the expected error $\mathrm{E}[\epsilon_n]$:

$$\mathrm{E}[\langle I_n\rangle] = I + \mathrm{E}[\epsilon_n] = \Psi_j L_{\mathrm{o}} + \frac{1}{n}\sum_{j=1}^{n}\Psi_j\mathrm{E}[\epsilon_j]. \tag{S.39}$$

Section C.2 derives the expression for the expected error $\mathrm{E}[\epsilon_j]$ for the case of transient rendering.

Eliminating bias. To eliminate bias, a transient *progressive* approach can be used. As discussed in Section A, progressive photon mapping uses a number of photon tracing passes, each providing an increasingly accurate solution. It can be shown (see Section C.2 for the full derivation) that the variance of the density estimation in transient PPM is inversely proportional to $R^2\mathcal{T}$:

$$\frac{\text{Var}[\epsilon_{j+1}]}{\text{Var}[\epsilon_j]} = \frac{R_j^2\mathcal{T}_j}{R_{j+1}^2\mathcal{T}_{j+1}} = \frac{j+1}{j+\alpha}. \tag{S.40}$$

Equation (S.40) shows that transient PPM involves progressively reducing two parameters, R and $\mathcal{T}$, referring to the spatial and temporal domains. We thus split the variance scaling factor $(j+1)/(j+\alpha)$ into two, one for each kernel bandwidth. This yields:

$$\frac{\mathcal{T}_{j+1}}{\mathcal{T}_j} = \left(\frac{j+\alpha}{j+1}\right)^{\beta_\mathcal{T}}, \qquad \frac{R_{j+1}^2}{R_j^2} = \left(\frac{j+\alpha}{j+1}\right)^{\beta_R} \tag{S.41}$$

where $\beta_\mathcal{T}$ and β_R are scalars in the range $[0,1]$ which control how much each term is to be scaled separately, with $\beta_\mathcal{T}+\beta_R = 1$.

Error analysis. For our analysis on the error of the estimate, we use the AMSE metric (8) again. The progressive approach described above ensures that error tends to zero when $n \to \infty$. However, the two terms of the AMSE metric have different convergence rates, dependent on the parameters α, $\beta_\mathcal{T}$ and β_R. We model the variance of the estimate $\text{Var}[\langle I_n \rangle]$ as [Knaus and Zwicker 2011]:

$$\text{Var}[\langle I_n \rangle] = \frac{1}{n}\text{Var}[\Psi L_\text{o}] + \frac{1}{n^2}\sum_{j=1}^{n}\text{Var}[\Psi\epsilon_j]. \tag{S.42}$$

The first term is a Monte Carlo estimator, with known convergence rate $O(n^{-1})$, while the second is the variance of the average error $\text{Var}[\epsilon_n]$, which converges with $O(n^{-\alpha})$. The detailed derivation of this result appears in Appendix C.3:

$$\text{Var}[\langle I_n \rangle] = O(n^{-1}) + O(n^{-\alpha}) = O(n^{-\alpha}). \tag{S.43}$$

Therefore, variance for transient PPM converges at the same rate as standard PPM, which for $\alpha = 1$ converges at the same rate as a Monte Carlo estimator. The parameters $\beta_\mathcal{T}$ and β_R have no influence on the convergence rate.

On the other hand, the expected average error in transient PPM $\text{E}[\epsilon_n]$ is (see Section C.3):

$$\text{E}[\langle I_n \rangle] = \text{E}[\frac{1}{n}\sum_{j=1}^{n}\Psi_j(L_{\text{o},j} + \epsilon_j)] = I + \text{E}[\frac{1}{n}\sum_{j=1}^{n}\Psi_j\epsilon_j]. \tag{S.44}$$

while its convergence rate is:

$$\text{E}[\langle I_n \rangle] = O(n^{1-\alpha})^{-2\beta_\mathcal{T}} + O(n^{1-\alpha})^{\beta_\mathcal{T}-1} + O(n^{1-\alpha})^{-(1+\beta_\mathcal{T})}. \tag{S.45}$$

In this case, the choice of $\beta_\mathcal{T}$ (and in consequence β_R) has a crucial effect on the asymptotic bias reduction. By observing the behavior in the limit, we find that the optimal parameter is $\beta_\mathcal{T} = 1/3$, which means that $\text{E}[\langle I_n \rangle]$ converges with $O(n^{-\frac{2(1-\alpha)}{3}})$.

Given these convergence rates, and using the AMSE metric given by Equation (8), we can now obtain the optimal parameter α for progressive transient radiance estimation, $\alpha = 4/7$ (see Section C.4). This means that the AMSE vanishes asymptotically with order $AMSE = O(n^{-\frac{4}{7}})$. This shows that including the temporal domain leads to a slower convergence rate than standard PPM ($O(n^{-2/3})$ [Kaplanyan and Dachsbacher 2013]); since the additional temporal kernel effectively increases bias, reducing it requires increasing the variance more at each step. However, note that in order to use standard PPM we would need to combine it with e.g. the histogram; this leads to a much slower convergence of standard PPM in *transient rendering*.

C.2 Variance and Expected Error of Transient Radiance Estimation

Following again the recent probabilistic framework for the progressive photon mapping algorithm [Knaus and Zwicker 2011], we first analyze the variance and expected value of the error introduced by the radiance estimate at each iteration. This error ϵ is defined as the difference between the estimated radiance $\widehat{L_\text{o}}$ and the actual radiance L_o, at point $\mathbf{x}$ and time t. Using the spatial and temporal kernels K_R and $K_\mathcal{T}$, with bandwidths R and $\mathcal{T}$ respectively, we have:

$$\epsilon(\mathbf{x}, R, t, \mathcal{T}) = \frac{1}{M}\sum_{i=1}^{M} K_R(\|\mathbf{x}-\mathbf{x}_i\|)K_\mathcal{T}(\|t-t_i^-\|)\gamma_i - L_\text{o}(\mathbf{x}, t). \tag{S.46}$$

Variance In order to compute the variance of the error $\mathrm{Var}[\epsilon]$ we need to make a set of assumptions: First, we assume that the photons' probability density is constant within the kernel K_R in the spatial domain [Knaus and Zwicker 2011], and within $K_\mathcal{T}$ in the temporal domain. We denote these probabilities as $p_R(\mathbf{x})$ and $p_\mathcal{T}(t)$ respectively. We also assume that the photons' position $\mathbf{x}_i$, time t_i and energy contribution γ_i are independent samples of the random variables X, T and γ, respectively, where X and T have probability densities $p_R(\mathbf{x})$ and $p_\mathcal{T}(t)$. Finally, we assume that the random variables X, T and γ are mutually independent. We thus model $\mathrm{Var}[\epsilon]$ as:

$$\begin{aligned}\mathrm{Var}[\epsilon] &= \mathrm{Var}[\frac{1}{M}\sum_{i=1}^{M} K_R(\|\mathbf{x}-X\|)K_\mathcal{T}(\|t-T\|)\gamma - L_\mathrm{o}(\mathbf{x},t)] \\ &= \frac{1}{M}(\mathrm{Var}[K_R]+\mathrm{E}[K_R]^2)(\mathrm{Var}[K_\mathcal{T}]+\mathrm{E}[K_\mathcal{T}]^2) \\ &\quad (\mathrm{Var}[\gamma]+\mathrm{E}[\gamma]^2) - \frac{1}{M}\mathrm{E}[K_R]^2\mathrm{E}[K_\mathcal{T}]^2\mathrm{E}[\gamma]^2. \end{aligned} \tag{S.47}$$

Here $\mathrm{E}[K_R] = p_R(\mathbf{x})$ and $\mathrm{E}[K_\mathcal{T}] = p_\mathcal{T}(t)$, the variance introduced by the temporal kernel $\mathrm{Var}[K_\mathcal{T}]$ is modeled by Equation (S.13), and $\mathrm{Var}[K_R]$ is derived analogously to Equation (S.13) (see Section B.1) as:

$$\mathrm{Var}[K_R] = \frac{p_R(\mathbf{x})}{R^2}\int_{\mathbb{R}^2} k_R(\psi)^2\,\mathrm{d}\psi - p_R(\mathbf{x})^2. \tag{S.48}$$

These transformations allow us to express (S.47) as:

$$\begin{aligned}\mathrm{Var}[\epsilon] &= \frac{1}{M}(\mathrm{Var}[\gamma]+\mathrm{E}[\gamma]^2) \\ &\quad (\frac{p_R(\mathbf{x})}{R^2}\int_{\mathbb{R}^2} k_R(\psi)^2\,\mathrm{d}\psi - p_R(\mathbf{x})^2 + p_R(\mathbf{x})^2) \\ &\quad (\frac{p_\mathcal{T}(t)}{\mathcal{T}}\int_{\mathbb{R}} k_\mathcal{T}(\psi)^2\,\mathrm{d}\psi - p_\mathcal{T}(t)^2 + p_\mathcal{T}(t)^2) \\ &\quad -\frac{1}{M}p_R(\mathbf{x})^2 p_\mathcal{T}(t)^2\mathrm{E}[\gamma]^2 \\ &\approx \frac{1}{M}(\mathrm{Var}[\gamma]+\mathrm{E}[\gamma]^2)(\frac{p_R(\mathbf{x})}{R^2}\mathcal{C}_R)(\frac{p_\mathcal{T}(t)}{\mathcal{T}}\mathcal{C}_\mathcal{T}) \end{aligned} \tag{S.49}$$

where $\mathcal{C}_R$ and $\mathcal{C}_\mathcal{T}$ are kernel-dependent constants. The last term can be neglected by assuming that the kernels cover small areas in their respective domains, which effectively means that $\mathcal{C}_R \gg p_R(\mathbf{x})$ and $\mathcal{C}_\mathcal{T} \gg p_\mathcal{T}(t)$. Equation (S.49) shows that for transient density estimation, the variance $\mathrm{Var}[\epsilon]$ is inversely proportional to $R^2\mathcal{T}$. We show in Section C.3 how this fact affects the shrinking formulation for progressive estimation.

Bias Bias at each iteration j is defined as the expected value of the error $\mathrm{E}[\epsilon_j]$:

$$\begin{aligned}\mathrm{E}[\epsilon_j] &= \mathrm{E}[\frac{1}{M}\sum_{i=1}^{M} K_R(\|\mathbf{x}-X\|)K_\mathcal{T}(\|t-T\|)\gamma - L_\mathrm{o}(\mathbf{x},t)] \\ &= \mathrm{E}[K_R(\|\mathbf{x}-X\|)]\mathrm{E}[K_\mathcal{T}(\|t-T\|)]\mathrm{E}[\gamma] - L_\mathrm{o}(\mathbf{x},t). \end{aligned} \tag{S.50}$$

As with the variance, we need to assume that the photons' position, time and energy contribution can be interpreted as independent identically distributed random samples from the random variables X (with probability density $p_R(\mathbf{x})$), T (with probability density $p_R(\mathbf{x})$) and γ respectively.

The expected value of $K_\mathcal{T}(\|t-T\|)$ is described in Section B.1, and modeled using Equation (S.20), while the expected value of the spatial kernel $\mathrm{E}[K_R]$ is derived in [Knaus and Zwicker 2011] as:

$$\mathrm{E}[K_R] \approx p_R(\mathbf{x}) + R^2\int_{\mathbb{R}^2} k_R(\psi)O(\|\psi\|^2)\,\mathrm{d}\psi = p_R(\mathbf{x}) + R^2\mathcal{C}_R^{ii}. \tag{S.51}$$

Using Equations (S.20) and (S.51), and $L_\mathrm{o}(\mathbf{x},t) = p_R(\mathbf{x})p_\mathcal{T}(t)\mathrm{E}[\gamma]$ we get the expected value of the error $\mathrm{E}[\epsilon_j]$ for iteration j:

$$\begin{aligned}\mathrm{E}[\epsilon_j] &\approx (p_R(\mathbf{x}) + R^2\mathcal{C}_R^{ii})(p_\mathcal{T}(t) + \mathcal{T}^2\mathcal{C}_\mathcal{T}^{ii})\mathrm{E}[\gamma] \\ &\quad -p_R(\mathbf{x})p_\mathcal{T}(t)\mathrm{E}[\gamma] \\ &= \mathrm{E}[\gamma](p_R(\mathbf{x})\mathcal{T}^2\mathcal{C}_\mathcal{T}^{ii} + p_\mathcal{T}(t)R^2\mathcal{C}_R^{ii} + \mathcal{T}^2\mathcal{C}_\mathcal{T}^{ii}R^2\mathcal{C}_R^{ii}) \end{aligned} \tag{S.52}$$

C.3 Variance and expected error of the pixel estimate

Here we derive the variance and expected error of the pixel estimate $\langle I_n \rangle$ after n steps of the progressive algorithm, as modeled in Equation (S.39). The samples are the hit position $\mathbf{x}_j$ and time instant t_j where the radiance estimation is being computed. As previously, we assume that they are independent identically distributed random samples.

Variance Assuming that the random variables Ψ and ϵ_j are independent, we model the variance of the estimator $\text{Var}[\langle I_n \rangle]$ as [Knaus and Zwicker 2011]:

$$\begin{aligned}
\text{Var}[\langle I_n \rangle] &= \frac{1}{n}\text{Var}[\Psi L_\text{o}] + \frac{1}{n^2}\sum_{j=1}^{n}\text{Var}[\Psi\epsilon_j] \\
&= \frac{1}{n}\text{Var}[\Psi L_\text{o}] + \text{Var}[\Psi]\frac{1}{n^2}\sum_{j=1}^{n}\text{Var}[\epsilon_j] \\
&\quad + \text{E}[\Psi]^2\frac{1}{n^2}\sum_{j=1}^{n}\text{Var}[\epsilon_j] + \text{Var}[\Psi]\frac{1}{n^2}\sum_{j=1}^{n}\text{E}[\epsilon_j]^2.
\end{aligned} \tag{S.53}$$

The first term is the usual Monte Carlo estimator, which vanishes with $O(n^{-1})$. The other three terms, however, are functions of the error ϵ_j. $\text{Var}[\epsilon_n]$ is the variance of the average error, modeled as:

$$\text{Var}[\epsilon_n] = \sum_{j=1}^{n}\frac{1}{n}\text{Var}[\epsilon_j] = \frac{1}{n^2}\sum_{j=1}^{n}\text{Var}[\epsilon_j]. \tag{S.54}$$

As described before, we allow the variance of the expected error to increase at each iteration by a factor:

$$\frac{\text{Var}[\epsilon_{j+1}]}{\text{Var}[\epsilon_j]} = \left(\frac{j+1}{j+\alpha}\right)^{\beta_\mathcal{T}} \cdot \left(\frac{j+1}{j+\alpha}\right)^{\beta_R} = \frac{j+1}{j+\alpha}. \tag{S.55}$$

Following [Knaus and Zwicker 2011], we can model $\text{Var}[\epsilon_n]$ as a function of the variance at the first iteration $\text{Var}[\epsilon_1]$ as:

$$\text{Var}[\epsilon_n] = \frac{\text{Var}[\epsilon_1]}{n^2}(1 + \sum_{j=2}^{n} j\alpha\text{B}(\alpha, j)), \tag{S.56}$$

where $\text{B}(x, y)$ is the Beta function. This value can be approximated as [Kaplanyan and Dachsbacher 2013]:

$$\text{Var}[\epsilon_n] \approx \frac{\text{Var}[\epsilon_1]}{(2-\alpha)n^\alpha} = O(n^{-\alpha}). \tag{S.57}$$

Finally, using this formulation of the variance of the average error $\text{Var}[\epsilon_n]$ and asypmtotic simplifications, we can formulate $\text{Var}[\langle I_n \rangle]$ (S.54) as:

$$\begin{aligned}
\text{Var}[\langle I_n \rangle] &\approx \frac{1}{n}\text{Var}[\Psi L_\text{o}] + \text{E}[\Psi]^2\text{Var}[\epsilon_n] \\
&\approx \frac{1}{n}\text{Var}[\Psi L_\text{o}] + \frac{\text{Var}[\epsilon_1]}{(2-\alpha)n^\alpha} \\
&= O(n^{-1}) + O(n^{-\alpha}) = O(n^{-\alpha}).
\end{aligned} \tag{S.58}$$

Expected error The expected value of of the estimator $\text{E}[\langle I_n \rangle]$ is modeled as:

$$\begin{aligned}
\text{E}[\langle I_n \rangle] &= \text{E}[\frac{1}{n}\sum_{j=1}^{n}\Psi_j(L_{\text{o},j} + \epsilon_j)] \\
&= \frac{1}{n}\sum_{j=1}^{n}\text{E}[\Psi_j L_{\text{o},j}] + \frac{1}{n}\sum_{j=1}^{n}\text{E}[\Psi_j]\text{E}[\epsilon_j] \\
&= I_n + \text{E}[\Psi]\text{E}[\epsilon_n]
\end{aligned} \tag{S.59}$$

where $\mathrm{E}[\epsilon_n]$ is the bias of the estimator after n iterations (S.29), and $\mathrm{E}[\epsilon_j]$ is the expected value of the error at each pass (S.53). Following Equation (S.41) and (S.30) we compute $\mathcal{T}_j$ as a function of its initial value $\mathcal{T}_1$ as:

$$\mathcal{T}_j = \mathcal{T}_1(j\alpha \mathrm{B}(\alpha, j))^{-\beta_\mathcal{T}} \tag{S.60}$$

where $\mathrm{B}(x, y)$ is the Beta function. Analogously, we compute R_j^2 as a function of its initial value R_1^2:

$$R_j^2 = R_1^2(j\alpha \mathrm{B}(\alpha, j))^{-\beta_R}. \tag{S.61}$$

Using (S.60) and (S.61) we can express $\mathrm{E}[\epsilon_j]$ as a function of the initial kernel bandwidths:

$$\begin{aligned} \mathrm{E}[\epsilon_j] &= \mathrm{E}[\gamma] p_R(\mathbf{x}) \mathcal{T}_1^2 (j\alpha \mathrm{B}(\alpha, j))^{-2\beta_\mathcal{T}} \mathcal{C}_\mathcal{T}^{ii} \\ &\quad + \mathrm{E}[\gamma] p_\mathcal{T}(t) R_1^2 (j\alpha \mathrm{B}(\alpha, j))^{-\beta_R} \mathcal{C}_R^{ii} \\ &\quad + \mathrm{E}[\gamma] \mathcal{T}_1^2 R_1^2 (j\alpha \mathrm{B}(\alpha, j))^{-(2\beta_\mathcal{T} + \beta_R)} \mathcal{C}_\mathcal{T}^{ii} \mathcal{C}_R^{ii}, \end{aligned} \tag{S.62}$$

which using Stirling's formula to approximate $\mathcal{T}j$ and Rj allows us to express $\mathrm{E}[\epsilon_j]$ as:

$$\begin{aligned} \mathrm{E}[\epsilon_j] &= \mathrm{E}[\gamma] p_R(\mathbf{x}) \mathcal{C}_\mathcal{T}^{ii} \mathcal{T}_1^2 \Theta(j^{1-\alpha})^{-2\beta_\mathcal{T}} \\ &\quad + \mathrm{E}[\gamma] p_\mathcal{T}(t) \mathcal{C}_R^{ii} R_1^2 \Theta(j^{1-\alpha})^{-\beta_R} \\ &\quad + \mathrm{E}[\gamma] \mathcal{C}_\mathcal{T}^{ii} \mathcal{C}_R^{ii} \mathcal{T}_1^2 R_1^2 \Theta(j^{1-\alpha})^{-(2\beta_\mathcal{T} + \beta_R)}. \end{aligned} \tag{S.63}$$

Finally, we use $\sum_{j=1}^{n} \Theta(j^x) = nO(n^x)$ to plug Equation (S.63) into Equation (S.29) to get the asymptotic behavior of $\mathrm{E}[\epsilon_n]$ in transient progressive photon mapping:

$$\begin{aligned} \mathrm{E}[\epsilon_n] &= \frac{\mathrm{E}[\gamma]}{n} p_R(\mathbf{x}) \mathcal{C}_\mathcal{T}^{ii} \mathcal{T}_1^2 n O(n^{1-\alpha})^{-2\beta_\mathcal{T}} \\ &\quad + \frac{\mathrm{E}[\gamma]}{n} p_\mathcal{T}(t) \mathcal{C}_R^{ii} R_1^2 n O(n^{1-\alpha})^{-\beta_R} \\ &\quad + \frac{\mathrm{E}[\gamma]}{n} \mathcal{C}_\mathcal{T}^{ii} \mathcal{C}_R^{ii} \mathcal{T}_1^2 R_1^2 n O(n^{1-\alpha})^{-(2\beta_\mathcal{T} + \beta_R)} \\ &= O(n^{1-\alpha})^{-2\beta_\mathcal{T}} + O(n^{1-\alpha})^{-\beta_R} + O(n^{1-\alpha})^{-(2\beta_\mathcal{T} + \beta_R)}, \end{aligned} \tag{S.64}$$

which, by using the equality $\beta_R = 1 - \beta_\mathcal{T}$, becomes:

$$\mathrm{E}[\epsilon_n] = O(n^{1-\alpha})^{-2\beta_\mathcal{T}} + O(n^{1-\alpha})^{\beta_\mathcal{T} - 1} + O(n^{1-\alpha})^{-(1+\beta_\mathcal{T})}. \tag{S.65}$$

Inserting this last equality into Equation (S.59) allows us to compute the asymptotic form of $\mathrm{E}[\langle I_n \rangle]$ as:

$$\mathrm{E}[\langle I_n \rangle] = O(n^{1-\alpha})^{-2\beta_\mathcal{T}} + O(n^{1-\alpha})^{\beta_\mathcal{T} - 1} + O(n^{1-\alpha})^{-(1+\beta_\mathcal{T})}. \tag{S.66}$$

where the error vanishes in the limit since we impose that $\mathrm{E}[\epsilon_j] > \mathrm{E}[\epsilon_j]$. This consistency has a convergence rate dependent on both α (the common PPM parameter), and $\beta_\mathcal{T}$, which balances the shrinking rate of the spatial and temporal kernels.

C.4 Minimizing the Mean Squared Error

Here we obtain the parameters α and $\beta_\mathcal{T}$ that allows optimal consistency in terms of ASME (8) in transient progressive photon mapping. Using the expressions obtained for the variance $\mathrm{Var}[\langle I_n \rangle]$ and expected error $\mathrm{Var}[\langle I_n \rangle]$, we model AMSE as:

$$\begin{aligned} AMSE(\langle I_n \rangle) &= O(n^{-\alpha}) + (O(n^{1-\alpha})^{-2\beta_\mathcal{T}} \\ &\quad + O(n^{1-\alpha})^{\beta_\mathcal{T} - 1} + O(n^{1-\alpha})^{-(1+\beta_\mathcal{T})})^2. \end{aligned} \tag{S.67}$$

which is a function of the parameters α and $\beta_\mathcal{T}$. Given that the variance is independent of $\beta_\mathcal{T}$, we first obtain the optimal value for this parameter that yields the highest convergence rate of the bias $\mathrm{E}[\epsilon_n]$. After differenciating Equation (S.65), applying asymptotic simplifications and equating to zero, we obtain the optimal value $\beta_\mathcal{T} = 1/3$. By plugging this value in Equation (S.65), we can express the AMSE as:

$$AMSE(\langle I_n \rangle) = O(n^{-\alpha}) + O(n^{-\frac{4}{3}(1-\alpha)}). \tag{S.68}$$

Differentiating Equation (S.68) and equating to zero we get the optimal parameter$\alpha = 4/7$, which results in the optimal convergence rate of the AMSE of our transient progressive photon mapping formulation:

$$AMSE(\langle I_n \rangle) = O(n^{-\frac{4}{7}}) + O(n^{-\frac{4}{3}(1-\frac{4}{7})}) = O(n^{-\frac{4}{7}}). \tag{S.69}$$

D Derivations for time sampling

Here we derive the different pdf shown in Section 5 and illustrate how each of those pdf lead to a uniform distribution of path samples along the temporal domain.

D.1 Sampling scattering distance in eye/light subpaths

This section describes the derivation for the pdf chosen in Section 5.1 of the paper. We aim to find a pdf $p(r)$ for a single segment of the light subpath, such as the distribution of subpath vertices along the temporal domain is uniform. We define $p\left(\cup_{i=1}^{\infty} t_i\right)$ based on the per-vertex temporal location probability distribution $p(t_i)$ for all subpath vertices:

$$p\left(\cup_{i=1}^{\infty} t_i\right) = \sum_{i=1}^{\infty} p(t_i), \tag{S.70}$$

where $p(t_i)$ is defined based as the addition on $t_i = t(\mathbf{x}_i \leftrightarrow \mathbf{x}_{i-1}) + t_{i-1}$ (propagation time plus temporal location of the previous sample). The probability distribution of an addition is the convolution of the probability distribution of the addends, so therefore

$$p(t_i) = \int_0^{t_i} p\big(t(\mathbf{x}_{i-1} \leftrightarrow \mathbf{x}_i)\big) p(t_{i-1}) dt_{i-1}, \tag{S.71}$$

$$p(t_1) = p\big(t(\mathbf{x}_0 \leftrightarrow \mathbf{x}_1)\big). \tag{S.72}$$

The probability distribution of the propagation time $p\left(t(\mathbf{x}_i \leftrightarrow \mathbf{x}_{i-1})\right)$ is related to the scattering distance pdf $p(r)$ by the change of variable $r = \frac{c}{\eta} t(\mathbf{x}_{i-1} \leftrightarrow \mathbf{x}_i)$.

This temporal distribution $p\left(\cup_{i=1}^{\infty} t_i\right)$ should be uniform in time. We are obviously not letting the system to cast paths of infinite number of interactions. As stated in the paper, we reject samples when they get out of the sensor temporal window.

For this derivation we will move our calculations to Laplace space. We note the Laplace transform of a function f as $\mathcal{L}\{f\}$, and we are interested in the following properties of the Laplace transform:

$$\mathcal{L}\{af(t)\} = a\mathcal{L}\{f\}(s) \tag{S.73}$$

$$\mathcal{L}\{f(t) + g(t)\} = \mathcal{L}\{f\}(s) + \mathcal{L}\{g\}(s) \tag{S.74}$$

$$\mathcal{L}\{f(at)\} = \frac{1}{|a|}\mathcal{L}\{f\}(\frac{s}{a}) \tag{S.75}$$

$$\mathcal{L}\{\int_0^t f(\tau) g(t-\tau) d\tau\} = \mathcal{L}\{f\}(s) \cdot \mathcal{L}\{g\}(s) \tag{S.76}$$

$$\mathcal{L}\{u(t)\} = \frac{1}{s} \tag{S.77}$$

$$\mathcal{L}\{e^{-\alpha t} \cdot u(t)\} = \frac{1}{s+\alpha} \tag{S.78}$$

where $u(t)$ is the unit step function:

$$u(t) = \begin{cases} 0 & t < 0 \\ 1 & t \geq 0 \end{cases} \tag{S.79}$$

and s is the variable in the Laplace frequency domain.

Lets prove that the exponential distribution $p(r) = u(r)\lambda e^{-\lambda r}$ leads to a uniform distribution of samples in time $p\left(\cup_{i=1}^{\infty} t_i\right)$. We start by applying the Laplace transform to $p(r)$ based on the Laplace properties defined in (S.73) and (S.78):

$$\mathcal{L}\{\lambda e^{-\lambda r}\} = \frac{\lambda}{s^{(r)} + \lambda} \tag{S.80}$$

where $s^{(r)}$ is the Laplace frequency domain representation of r. For any propagation time $t(\mathbf{x}_i \leftrightarrow \mathbf{x}_{i-1})$ we always apply the same scattering distance pdf $p(r)$ so we apply the change of variable:

$$\mathcal{L}\{p(t(\mathbf{x}_i \leftrightarrow \mathbf{x}_{i-1}))\} = \frac{c}{\eta}\frac{\eta}{c}\mathcal{L}\{p\}(\frac{\eta}{c}s^{(\leftrightarrow)}) = \frac{\lambda}{\frac{\eta}{c}s^{(\leftrightarrow)} + \lambda} \tag{S.81}$$

where $t(\mathbf{x}_i \leftrightarrow \mathbf{x}_{i-1})$ represents propagation time in this case as a change of variable from distance r, and $s^{(\leftrightarrow)}$ is the Laplace domain representation of $t(\mathbf{x}_i \leftrightarrow \mathbf{x}_{i-1})$. This is applied for the first interaction,

$$\mathcal{L}\{p(t_1)\} = \frac{c}{\eta}\frac{\eta}{c}\mathcal{L}\{p\}(\frac{\eta}{c}s_i) = \frac{\lambda}{\frac{\eta}{c}s_1^{(t)} + \lambda}, \tag{S.82}$$

where $s_i^{(t)}$ is the Laplace domain representation of t_i. The subsequent interactions are obtained by applying the convolution property (S.76) to (S.71):

$$\mathcal{L}\{p(t_i)\} = \mathcal{L}\{p\}(s^{(\leftrightarrow)}) \cdot \mathcal{L}\{p\}(s_{i-1}^{(t)}). \tag{S.83}$$

We then apply recursively (S.83) so $\mathcal{L}\{p(t_i)\}$ actually becomes a simple power:

$$\mathcal{L}\{p(t_i)\} = \left(\mathcal{L}\{p\}(s^{(\leftrightarrow)})\right)^i = \left(\frac{\lambda}{\frac{\eta}{c}s^{(\leftrightarrow)} + \lambda}\right)^i \tag{S.84}$$

We then apply the Laplace transform to (S.70) using the property (S.74):

$$\mathcal{L}\{p\left(\cup_{i=1}^{\infty} t_i\right)\} = \sum_{i=1}^{\infty}\left(\frac{\lambda}{\frac{\eta}{c}s^{(\leftrightarrow)} + \lambda}\right)^i \tag{S.85}$$

which is in fact an infinite geometric series, which has the following analytical solution:

$$\mathcal{L}\{p\left(\cup_{i=1}^{\infty} t_i\right)\} = \frac{1}{1 - \frac{\lambda}{\frac{\eta}{c}s^{(\leftrightarrow)} + \lambda}} - 1 = \frac{c}{\eta}\frac{\lambda}{s^{(\leftrightarrow)}} \tag{S.86}$$

If we then apply the anti-transform by properties (S.73) and (S.77) we get:

$$p\left(\cup_{i=1}^{\infty} t_i\right) = \frac{c}{\eta}\lambda u(t) \tag{S.87}$$

which is uniform in time. We reject all paths with total duration t_k out of the rendered time window $(0, t_e)$ so $p\left(\cup_{i=1}^{\infty} t_i\right)$ never reaches infinite time nor infinite number of bounces. Notice that this derivation is analogous for the eye subpath.

D.2 Sampling line-to-point shadow connections

This section describes the derivation of the line-to-point shadow connection pdf described in Section 5.2 of the paper.

Any sampling strategy is usually based on a pdf that can be analytically integrated to get an invertible cdf. In the case of steady-state strategies, the pdf reduces variance by approximating the integrand (in this case path contribution, radiance) as close as possible. This approach, however, cannot be directly applied to transient rendering, where the uniformity of samples along the temporal dimension becomes quite relevant for the accuracy of the result.

The inverse cdfselects a sample according to a uniformly distributed random number $\xi \in [0, 1)$. The key idea of this sampling strategy is to preserve the uniformity of such random number when changing to the temporal domain. The time to be uniformly sampled is the propagation time $\{\mathbf{x}_i, \mathbf{x}_{i+1}, \mathbf{x}_{i+2}\}$. We name the distance from $\mathbf{x}_i$ to $\mathbf{x}_{i+1}$ as r_{i+1}, and r_{i+2} is the distance from $\mathbf{x}_{i+1}$ to $\mathbf{x}_{i+2}$. Therefore, the total propagation time for this connection is

$$t = \frac{\eta}{c}\left(r_{i+2} + r_{i+1}\right). \tag{S.88}$$

We start by expressing the total propagation time as a function of r_{i+2}:

$$t_{i+2}^{-}(r_{i+2}) = \frac{\eta}{c}\left(r_{i+2} + \underbrace{\sqrt{r_{i+2}^2 - 2r_{i+2}(\mathbf{l}\cdot\omega) + \mathbf{l}\cdot\mathbf{l}}}_{r_{i+1}}\right) + t_i + \Delta t_{i+1}, \tag{S.89}$$

where $\mathbf{l} = \mathbf{x}_i - \mathbf{x}_{i+2}$. Inverting this allows us to sample path locations for a specified temporal duration:

$$r_{i+2}(t^-_{i+2}) = \frac{(t^-_{i+2} - t_i - \Delta t_{i+1})^2 - \frac{\eta^2}{c^2}(\mathbf{l}\cdot\mathbf{l})}{2\frac{\eta}{c}(t^-_{i+2} - t_i - \Delta t_{i+1}) - 2\frac{\eta^2}{c^2}(\mathbf{l}\cdot\omega)}. \tag{S.90}$$

To obtain shadow connections uniformly distributed in a time range (t_a, t_b), we use Equation (S.90) as the inverse cdf

$$cdf^{-1}(\xi) = r_{i+2}(\xi(t_b - t_a) + t_a), \tag{S.91}$$

The normalized derivative of (S.89) is then the sampling pdf:

$$p(r_{i+2}) = \frac{\eta}{c(t_b - t_a)}\left(1 + \frac{r_{i+2} - (\mathbf{l}\cdot\omega)}{\sqrt{r^2_{i+2} - 2r_{i+2}(\mathbf{l}\cdot\omega) + (\mathbf{l}\cdot\mathbf{l})}}\right), \tag{S.92}$$

Assuming that animation lies within the time range $(0, t_e)$, we set the connection time limits (t_a, t_b) to the temporal range left by the rest of the path for that connection:

$$t_a = t_i + t(\mathbf{x}_i \leftrightarrow \mathbf{x}_{i+2}) \tag{S.93}$$

$$t_b = t_e - \Delta t_k - \left(\sum_{j=i+2}^{k-1} t(\mathbf{x}_j \leftrightarrow \mathbf{x}_{j+1}) + \Delta t_j\right) \tag{S.94}$$

D.3 Angular sampling

This section describes the derivation of the angular time sampling strategy discussed on Section 5.3 of the paper.

For the angular sampling pdf we follow a similar strategy than for a line-to-point shadow connection: we also aim to preserve the uniformity of the random number ξ when moving to the temporal domain. The propagation time between the vertices $\{\mathbf{x}_i, \mathbf{x}_{i+1}, \mathbf{x}_{i+1}\}$ is

$$t = \frac{\eta}{c}\left(r_k + r_{i+1}\right), \tag{S.95}$$

where r_{i+1} is the distance from $\mathbf{x}_i$ to $\mathbf{x}_{i+1}$ and r_k is the distance from $\mathbf{x}_{i+1}$ to $\mathbf{x}_k$. We express this propagation time as a function of θ

$$t(\theta) = \frac{\eta}{c}\left(r_{i+1} + \underbrace{\sqrt{(r_{i+1}\sin\theta)^2 + (|\mathbf{l}| - r_{i+1}\cos\theta)^2}}_{r_k}\right), \tag{S.96}$$

where $\mathbf{l} = \mathbf{x}_k - \mathbf{x}_i$. This can be simplified to:

$$t(\theta) = \frac{\eta}{c}\left(r_{i+1} + \sqrt{r^2_{i+1} + |\mathbf{l}|^2 - 2r_{i+1}|\mathbf{l}|\cos\theta}\right). \tag{S.97}$$

Inverting (S.97) allows us to choose an angle for a specified temporal duration:

$$\theta(t) = \arccos\left(\frac{|\mathbf{l}|^2 + 2t\frac{c}{\eta} - \left(t\frac{c}{\eta}\right)^2}{2|\mathbf{l}|r_{i+1}}\right) \tag{S.98}$$

We use (S.98) as the base for the inverse cdf

$$cdf^{-1}(\xi) = \theta(\xi(t(\pi) - t(0)) + t(0)), \tag{S.99}$$

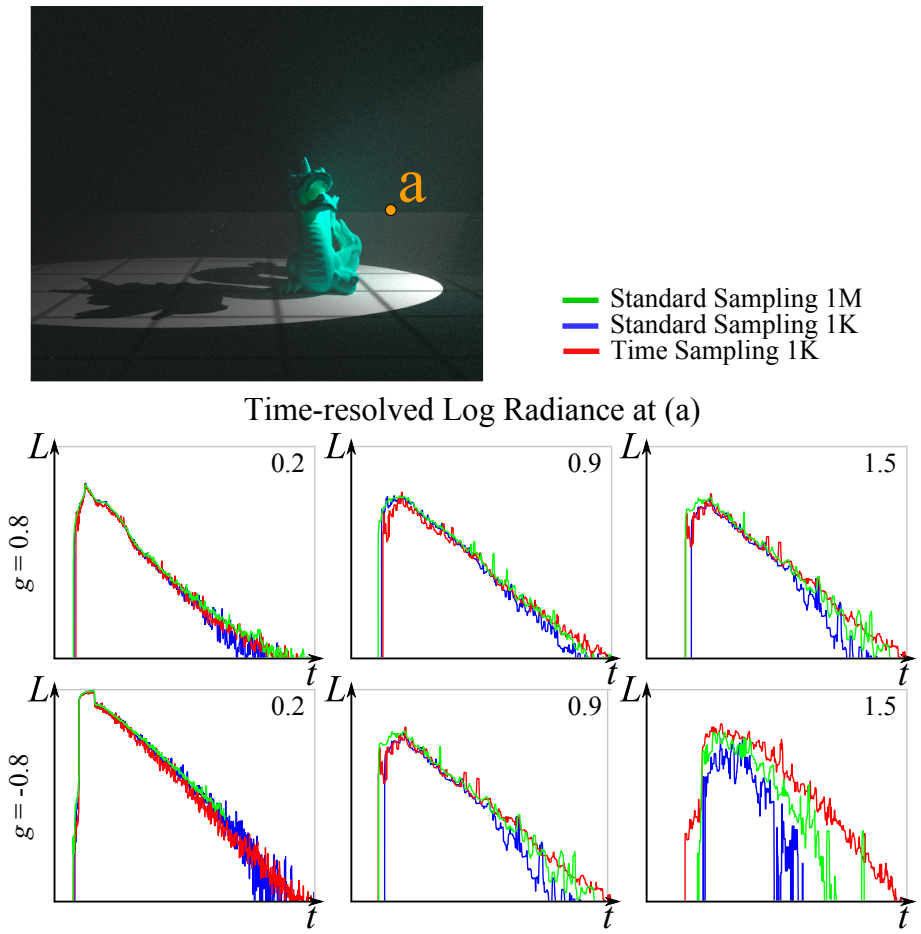

Figure S.2: *Comparison of our three time sampling strategies combined, against the standard techniques used in steady state, in the* dragon *scene accounting for multiple scattering (top). Each graph shows the time-resolved radiance (bottom) at pixel (a), for three different scattering coefficients* $\sigma_s = \{0.2, 0.9, 1.5\}$*, and absorption* $\sigma_a = 0.1$*, for anisotropic media with* $g = 0.8$ *and* $g = -.08$*. For 1K samples per pixel and frame, our combined techniques (red) feature a similar quality as standard steady state techniques with 1000 times more samples (green), while with the same number of samples, our techniques significantly outperform standard sampling (blue), especially in highly scattering media. Additionally, note that in backward scattering media our technique outperforms standard-techniques withe three orders of magnitude more samples. To emphasize the differences between sampling techniques, here we use the histogram path reuse (see Section 4). For results with isotropic media we refer to Figure 7 in the main text.*

which can be expanded into

$$cdf^{-1}(\xi) = \arccos\left(\frac{|\mathbf{l}| - 2r_{i+1}^2\xi^2 - 2\xi r_{i+1}\left(|\mathbf{l}| - 1\right)}{r_{i+1}|\mathbf{l}|}\right) \tag{S.100}$$

The normalized derivative of (S.97) is therefore our angular pdf:

$$p(\theta) = \frac{r_{i+1}\sin\theta}{2\sqrt{r_{i+1}^2 + |\mathbf{l}|^2 - 2r_{i+1}|\mathbf{l}|\cos\theta}} \tag{S.101}$$

E Additional Results

Here we include additional results showcasing the improvements of our techniques. Figure S.2 ilustrates the performance of our time-sampling techniques on non-isotropic media: our techniques significantly outperform classical radiance-based sampling

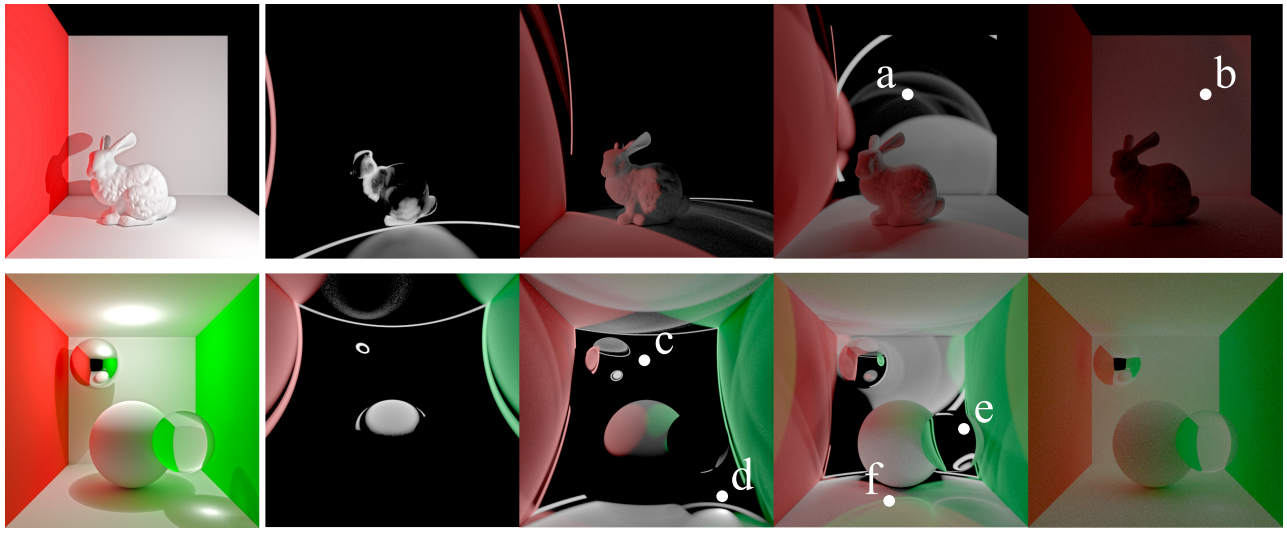

Figure S.3: *Light propagation in the absence of participating media, in the two scenes depicted in the leftmost column. In the* bunny *scene (top) we can see the curved shape of the indirect wavefronts, and how the irregular shape of the bunny model makes several different indirect wavefronts (a). In contrast the* spheres *scene we can observe the different delays due to longer optical paths in the mirror (c) and glass (e) spheres, and the primary (d) and secondary caustics (f). At longer times, when light has been reflected several times, the directionality is lost, and light is fundamentally diffuse (b).*

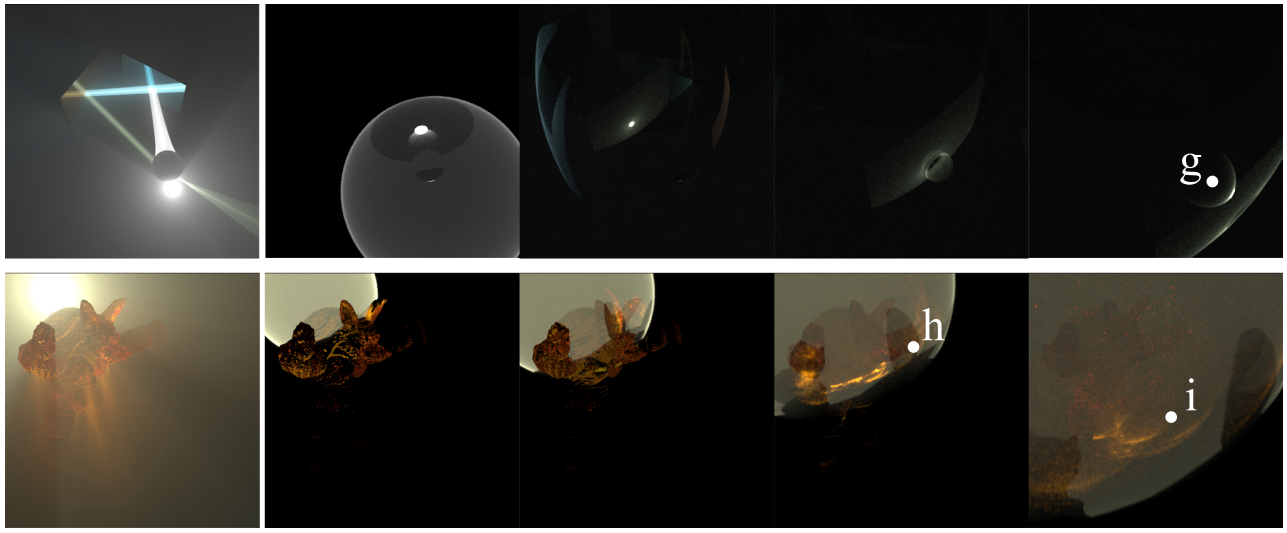

Figure S.4: *Selected frames from animations of transient light propagation in two scenes with scattering media (the steady state rendering for both is in the leftmost column) computed using* transient photon beams*. In both scenes we can see the delay produced in the caustics by the higher index-of-refraction of the crystal media (g,h), and how the different geometries generate different caustic patterns (g,i). We can also observe how the caustics are scattered and suffer extinction as they advance through the medium.*

over the full simulation domain for turbid media, while still gives comparable results for thin media. While in some cases our technique is worse for initial times, because radiance-based techniques allocates more samples at these instants, using multiple importance sampling (MIS) [Veach and Guibas 1995] to combine radiance- and our time-based sampling techniques would improve reconstruction over the full temporal domain.

We also include detailed explanations on complex time-resolved phenomena simulated with our framework, showing that by moving to transient-state several non-trivial effects appear. Note that here we place only a set of selected key-frames of the full animations; we refer to the supplementary video to visualize the full animations.

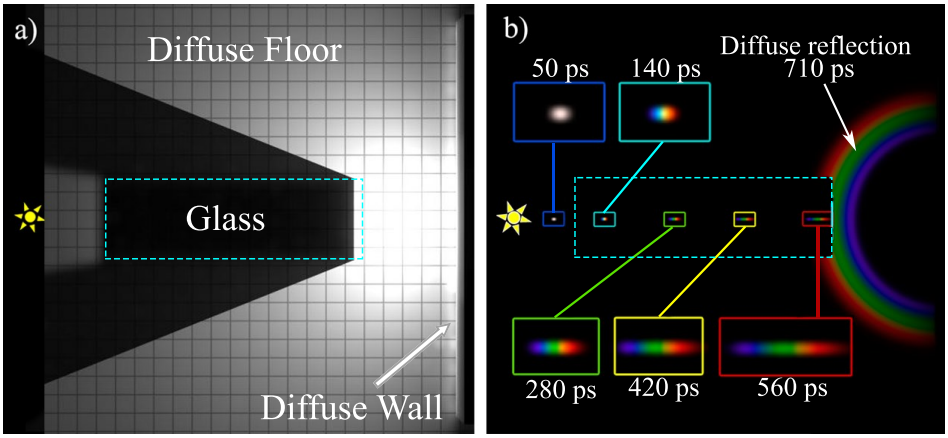

Figure S.5: *An orthogonal view of a scene with a white light pulse traversing perpendicularly a cube made of glass with index of refraction that varies linearly with the wavelength, in the range of [1.5,1.65], and hitting a wall after traversing the cube, transforming the interaction point into a virtual light source which illuminates a ground floor (b). Due to different speeds, the different wavelengths of the pulse are decomposed along the trajectory, even after the interaction. White light, as would be seen in steady state (a), therefore becomes a rainbow in the temporal domain.*

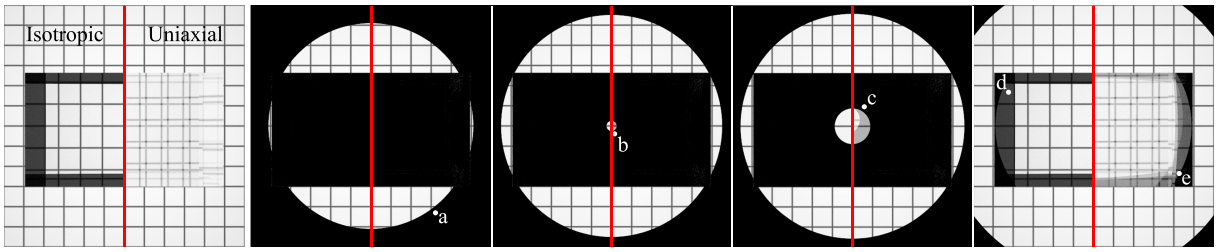

Figure S.6: *Light propagation in a piece of glass placed in front of a display with continuous emission in time; the left half of the image an isotropic crystal (with transmission governed by Snell's law) and the right half an uniaxial birefringent crystal. In both cases, the ordinary index of refraction is 1.5, and the extraordinary η_e in the birefringent is 1.65. First, direct light from the display arrives the camera (a); then, the refracted image begins to form (b). The extraordinary image appears instants later than the ordinary (c), due to higher index of refraction. Also note that ordinary refraction in the uniaxial crystal has lower energy than in the isotropic. Finally, internal reflections with longer optical paths are formed (d,e).*

Figure S.3 shows transient light propagation in scenes without and with participating media, where the different wavefronts due to surface inter-reflections can be seen. Additionally, the depicted caustics show how traversing a medium with a higher index of refraction leads to a temporal delay of the wavefront, more visible with participating media (Figure S.4).

The temporal delay due to refraction becomes particularly interesting simulating real-world glass (Figure S.5): This scene, computed using the transient progressive photon mapping described in Section 4.1 that allows us to robustly render complex paths such as caustics, shows how as light traverses the glass, its wavelength-dependent index of refraction causes chromatic dispersion even when light incomes perpendicular to the surface normal. This could be used to obtain the wavelength-dependent index of refraction of different crystals, or the power spectra of a light source by using a perpendicular incident beam. Moreover, this refraction delay is different between ordinary and extraordinary rays in birefringent crystals [Weidlich and Wilkie 2008] (Figure S.6). Figure S.7 shows an example of the effect of the delay produced at scattering events $\rho_s(\mathbf{x}_{i-1} \to \mathbf{x}_i \to \mathbf{x}_{i+1}, \Delta t_i)$, for the particular case of fluorescence. The object's material is chlorophyll, which re-emits energy at 680 nm after absorption [Gutierrez et al. 2008].

References

CAMMARANO, M., AND JENSEN, H. W. 2002. Time dependent photon mapping. In *Eurographics Workshop on Rendering '02*.

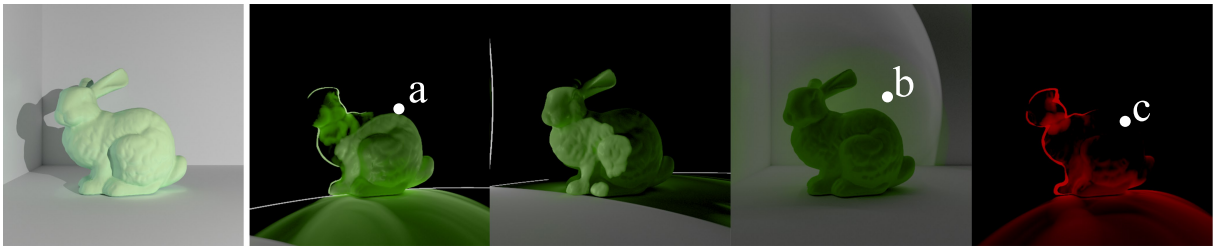

Figure S.7: *Fluorescent bunny illuminated by a pulse of light from a point light source. Light reaches the bunny, which reflects light centered in the green-ish spectra (b), and the rest is absorbed. After 10 ns, part of the absorbed light gets re-emitted at lower quantum energy, centered at 680 nm (c). Because of this re-emission, the hue of the bunny in steady state (a) is shifted towards yellow.*

GEORGIEV, I., KŘIVÁNEK, J., DAVIDOVIČ, T., AND SLUSALLEK, P. 2012. Light transport simulation with vertex connection and merging. *ACM Trans. Graph. 31*, 6.

GUTIERREZ, D., SERON, F., MUÑOZ, A., AND ANSON, O. 2008. Visualizing underwater ocean optics. *Computer Graphics Forum 27*, 2.

HACHISUKA, T., AND JENSEN, H. W. 2009. Stochastic progressive photon mapping. *ACM Trans. Graph. 28*, 5.

HACHISUKA, T., OGAKI, S., AND JENSEN, H. W. 2008. Progressive photon mapping. *ACM Trans. Graph. 27*, 5.

HACHISUKA, T., JAROSZ, W., GEORGIEV, I., KAPLANYAN, A., AND NOWROUZEZAHRAI, D. 2013. State of the art in photon density estimation. In *ACM SIGGRAPH ASIA 2013 Courses*.

JARABO, A., MARCO, J., MUÑOZ, A., BUISAN, R., JAROSZ, W., AND GUTIERREZ, D. 2014. A framework for transient rendering. *ACM Trans. Graph. 33*, 6.

JENSEN, H. W. 2001. *Realistic Image Synthesis Using Photon Mapping*. AK Peters.

KAPLANYAN, A. S., AND DACHSBACHER, C. 2013. Adaptive progressive photon mapping. *ACM Trans. Graph. 32*, 2.

KNAUS, C., AND ZWICKER, M. 2011. Progressive photon mapping: A probabilistic approach. *ACM Trans. Graph. 30*, 3.

LAFORTUNE, E. P., AND WILLEMS, Y. D. 1993. Bi-directional path tracing. In *Compugraphics '93*.

PAULY, M., KOLLIG, T., AND KELLER, A. 2000. Metropolis light transport for participating media. In *Rendering Techniques (Proc. of Eurographics Workshop on Rendering)*.

SCOTT, D. W. 1992. *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley.

VEACH, E., AND GUIBAS, L. 1994. Bidirectional estimators for light transport. In *Proc. of Eurographics Rendering Workshop*.

VEACH, E., AND GUIBAS, L. J. 1995. Optimally combining sampling techniques for Monte Carlo rendering. In *SIGGRAPH '95*.

VEACH, E. 1997. *Robust Monte Carlo methods for light transport simulation*. PhD thesis, Stanford.

WEIDLICH, A., AND WILKIE, A. 2008. Realistic rendering of birefringency in uniaxial crystals. *ACM Trans. Graph. 27*, 1.