Title

Towards Real-Time Monte Carlo for Biomedicine

ISBN

9783319914350

Authors

Zhao, Shuang

Kong, Rong

Spanier, Jerome

Publication Date

2018

DOI

10.1007/978-3-319-91436-7_25

Towards Real-Time Monte Carlo for Biomedicine

Shuang Zhao, Rong Kong and Jerome Spanier

Abstract Monte Carlo methods provide the "gold standard" computational technique for solving biomedical problems but their use is hindered by the slow convergence of the sample means. An exponential increase in the convergence rate can be obtained by adaptively modifying the sampling and weighting strategy employed. However, if the radiance is represented globally by a truncated expansion of basis functions, or locally by a region-wise constant or low degree polynomial, a bias is introduced by the truncation and/or the number of subregions. The sheer number of expansion coefficients or geometric subdivisions created by the biased representation then partly or entirely offsets the geometric acceleration of the convergence rate. As well, the (unknown amount of) bias is unacceptable for a gold standard numerical method. We introduce a new unbiased estimator of the solution of radiative transfer equation (RTE) that constrains the radiance to obey the transport equation. We provide numerical evidence of the superiority of this Transport-Constrained Unbiased Radiance Estimator (T-CURE) in various transport problems and indicate its promise for general heterogeneous problems.

Keywords Monte Carlo simulations · Transport-constrained radiance estimators

S. Zhao
Donald Bren School of Information and Computer Sciences,
University of California @ Irvine, 3019 Donald Bren Hall, Irvine, CA 92697, USA
e-mail: shz@ics.uci.edu

R. Kong
Hyundai Capital America, 3161 Michelson Drive, Irvine, CA 92612, USA
e-mail: kongr413@yahoo.com

J. Spanier (✉)
Beckman Laser Institute and Medical Clinic, University of California @ Irvine, 1002 Health Sciences Road, Irvine, CA 92612, USA
e-mail: jspanier@uci.edu

A. B. Owen and P. W. Glynn (eds.), *Monte Carlo and Quasi-Monte Carlo Methods*, Springer Proceedings in Mathematics & Statistics 241,
https://doi.org/10.1007/978-3-319-91436-7_25

1 Introduction

Monte Carlo simulation has provided the "gold standard" numerical method for solving biomedical problems for the past thirty years [22]. Nevertheless, its slow convergence (at the rate $N^{-1/2}$ where N equals sample size) inhibits use of Monte Carlo on a routine basis. Instead, diffusion-based numerical methods are often used because of their superior speed of execution, even though they may provide very poor descriptions of the radiant light field in many situations. Consequently, there has been a lot of interest in accelerating the convergence of Monte Carlo simulations, especially within the biomedical community, where accuracy is a primary focus.

Conventional density function estimation methods [6, 16, 17] are widely used with success where photorealism—not image perfection—is the goal. Such methods have revolutionized the rendering of scenes for electronic games and movies [7]. Density estimation methods avoid the need to represent the radiance in a functional expansion, but they introduce the need for "smoothing parameters" which also causes a bias in the density estimator. This precludes convergence to the exact solution and is unacceptable as a gold standard method for biomedicine or biology. The question we then asked was: Can any of these ideas be used in such a way that the radiance it produces actually satisfies the governing radiative transport equation? If so, might that produce a candidate to serve as a gold standard for biomedical simulations? That investigation has led to the publication [12] and to this paper.

2 Radiative Transport Fundamentals

Before proceeding with this line of thinking we want to establish our notation and clarify our goals.

The rigorous transport of light in tissue usually begins with the integro-differential form of the equation which is then transformed to the integral form [18] of the RTE:

$$L(\boldsymbol{P}) = \int_{\Gamma} K(\boldsymbol{P}' \to \boldsymbol{P})\, L(\boldsymbol{P}')\, \mathrm{d}\boldsymbol{\omega}' \mathrm{d}\rho + S(\boldsymbol{P}), \tag{1}$$

where $\boldsymbol{P} := (\boldsymbol{r}, \boldsymbol{\omega})$, $\boldsymbol{P}' := (\boldsymbol{r}', \boldsymbol{\omega}')$, $\boldsymbol{r}' := \boldsymbol{r} - \rho\boldsymbol{\omega}$ and

$$K(\boldsymbol{P}' \to \boldsymbol{P}) := \frac{\mu_s(\boldsymbol{r}')}{\mu_t(\boldsymbol{r}')}\, f(\boldsymbol{r}';\ \boldsymbol{\omega}' \to \boldsymbol{\omega})\, T(\boldsymbol{r}' \to \boldsymbol{r};\ \boldsymbol{\omega}), \tag{2}$$

$$T(\boldsymbol{r}' \to \boldsymbol{r};\ \boldsymbol{\omega}) := \mu_t(\boldsymbol{r}')\, \mathrm{e}^{-\int_0^{\|\boldsymbol{r}-\boldsymbol{r}'\|} \mu_t(\boldsymbol{r}-\tau\boldsymbol{\omega})\mathrm{d}\tau}, \tag{3}$$

with source function

$$S(P) := \mathrm{e}^{-\int_0^R \mu_t(\boldsymbol{r}-\tau\boldsymbol{\omega})\mathrm{d}\tau}\, Q_0(\boldsymbol{r}-R\boldsymbol{\omega},\boldsymbol{\omega}) + \int_0^R \mathrm{e}^{-\int_0^\rho \mu_t(\boldsymbol{r}-\tau\boldsymbol{\omega})\mathrm{d}\tau}\, Q(\boldsymbol{r}-\rho\boldsymbol{\omega},\boldsymbol{\omega})\,\mathrm{d}\rho. \tag{4}$$

Appendix 7 provides details on how Eq. (1) arises from the integro-differential equation (24) in the time-independent case.

To complete the mathematical description,

- $\Gamma := V \times \mathbb{S}^2$ ($V \subseteq \mathbb{R}^3$) denotes the *phase space* of vectors $(\boldsymbol{r},\boldsymbol{\omega})$;
- μ_s and μ_a are respectively the *scattering* and *absorption* coefficients;
- $\mu_t := \mu_s + \mu_a$ is the *total attenuation* coefficient;
- f is the *single-scattering phase function* (that scatters photons from direction $\boldsymbol{\omega}'$ to $\boldsymbol{\omega}$ at location $\boldsymbol{r}'$);
- L denotes *photon radiance*.

2.1 Role of Eqs. (1)–(4) in Generating Samples

We now indicate how the Eqs. (1)–(4) play a role in generating photon biographies; i.e., samples drawn from our sample space $\mathbb{B}$.

Figure 1 depicts a hypothetical photon biography that is launched from the light source at the left, makes collisions at the locations $\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{r}_3$ and $\boldsymbol{r}_4$, then exits the tissue at the detector on the right. If we assume that there is no internal volumetric source, ($Q \equiv 0$ in Eq. (4)), but there is a nonzero source of light Q_0 on the boundary, then the launch position and direction $\boldsymbol{P}_0 = (\boldsymbol{r}_0, \boldsymbol{\omega}_0)$ are drawn by sampling Q_0, while the first collision location $\boldsymbol{r}_1$ is drawn by sampling the exponential probability density function with exponent $\int_0^{\|\boldsymbol{r}_0-\boldsymbol{r}_1\|} \mu_t(\boldsymbol{r}_0-\tau\boldsymbol{\omega}_0)\,\mathrm{d}\tau$ (see Eq. (3)). Provided that the photon is scattered at $\boldsymbol{r}_1$ (with probability μ_s/μ_t), the direction $\boldsymbol{\omega}_0$ is scattered into the direction $\boldsymbol{\omega}_1$ by sampling from the single-scattering phase function $f(\boldsymbol{r}_1;\ \boldsymbol{\omega}_0 \to \boldsymbol{\omega}_1)$. This process of locating successive collision points and unit directions continues until the photon biography $\boldsymbol{P}_0 = (\boldsymbol{r}_0,\boldsymbol{\omega}_0)$, $\boldsymbol{P}_1 = (\boldsymbol{r}_1,\boldsymbol{\omega}_1), \ldots$ either terminates by absorption (with probability $1-\mu_s/\mu_t$) at some collision point or escapes from the tissue, either at the detector or elsewhere on ∂V.

2.2 Equivalence Between Physical/Analytic and Stochastic Models

The **physical/analytic RTE model** consists of the equations of radiative transport in tissue, together with a linear functional L_i of the solution L of the RTE for each detector:

$$L_i = \int_\Gamma d_i(\boldsymbol{r},\boldsymbol{\omega}) L(\boldsymbol{r},\boldsymbol{\omega})\,\mathrm{d}\boldsymbol{r}\,\mathrm{d}\boldsymbol{\omega}, \quad (i = 1,\ldots,d).$$

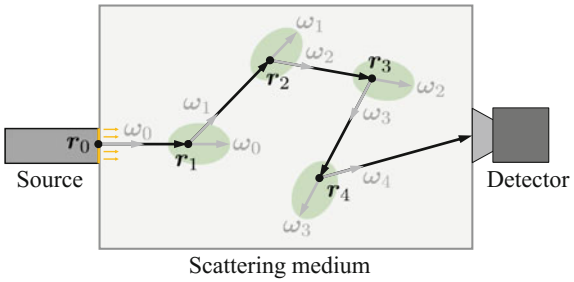

Fig. 1 Illustration of a photon biography comprising four collision points at $\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{r}_3, \boldsymbol{r}_4$. When entering a collision $\boldsymbol{r}_i$, the direction $\boldsymbol{\omega}_{i-1}$ of the photon changes to $\boldsymbol{\omega}_i$ according to the single-scattering phase function at $\boldsymbol{r}_i$ (indicated as green ellipses)

The **stochastic/probabilistic RTE model** used to characterize the Monte Carlo solution of this system consists of a probability measure space $\mathbb{B}$, a set of measurable subsets $\digamma$ of $\mathbb{B}$, and a probability measure $\mathbb{M}$ on $\mathbb{B}$ together with d random variables $\Xi_i : \mathbb{B} \to \mathbb{R}$ for $i = 1, \ldots, d$, each of which describes the contribution (tally) of any photon biography $\bar{b}$ to the detector L_i. Here,

- $\mathbb{B}$ is the *sample space* of all possible photon biographies $\bar{b}$ (that are termed as *light transport paths* in computer graphics [20], as illustrated in Fig. 1);
- $\Xi_i(\bar{b})$ is the *tally/score* associated with biography $\bar{b}$ for detector i;
- $\mathbb{E}[\Xi_i]$ is the expected value of Ξ_i with respect to $\mathbb{M}$;
- V is the physical domain of the phase space Γ.

If the measure $\mathbb{M}$ on $\mathbb{B}$ is induced by the analog simulation (by launching photons according to the physical source Q, transporting them from $\boldsymbol{r}'$ to $\boldsymbol{r}$ along $\boldsymbol{\omega}$ by sampling T, absorbing them at $\boldsymbol{r}$ with probability $1 - \mu_s/\mu_t$, scattering them at $\boldsymbol{r}$ with probability μ_s/μ_t and changing their direction from $\boldsymbol{\omega}'$ to $\boldsymbol{\omega}$ by sampling f), it is the case that for $i = 1, 2, \ldots, d$:

$$\mathbb{E}[\Xi_i] = \int_{\mathbb{B}} \Xi_i(\bar{b}) \, \mathrm{d}\mathbb{M}(\bar{b}) = \int_{\Gamma} d_i(\boldsymbol{r}, \boldsymbol{\omega}) L(\boldsymbol{r}, \boldsymbol{\omega}) \, \mathrm{d}\boldsymbol{r} \, \mathrm{d}\boldsymbol{\omega} = L_i. \tag{5}$$

The equality (5) establishes that the probabilistic model (the left-hand side) and the analytic model (the right-hand side) both represent the quantities L_i being estimated.

The same equality shows that each Ξ_i is a **theoretically unbiased** estimator of L_i for every N:

$$\frac{1}{N} \sum_{j=1}^{N} \Xi_i(\bar{b}_j) \xrightarrow{N \to \infty} \int_{\mathbb{B}} \Xi_i(\bar{b}) \, \mathrm{d}\mathbb{M}(\bar{b}) = L_i. \tag{6}$$

The symbol Ξ is reserved here for a random variable on the space $\mathbb{B}$ of all possible photon biographies $\bar{b}$ (i.e., the sample space), N is the total number of photons

released from the source, and $\mathbb{M}$ denotes the probability measure induced on $\mathbb{B}$ by the process associated with generating $\bar{b} \in \mathbb{B}$. That is, the measure $\mathbb{M}$ is constructed from the choice of probability density functions used to launch, transport, scatter and absorb each biography $\bar{b}$, whether these are analog or not.

3 Our Goal and Current Status

Our previous research (see [8–11, 14]) on adaptive Monte Carlo algorithms for radiative transport problems resulted in the development of several geometrically convergent algorithms for global transport solutions L. By geometric convergence, we mean

$$E_s < \lambda E_{s-1} < \lambda^s E_0, \qquad (0 < \lambda < 1),$$

where s is the stage number and E_s is the sth stage error; e.g.,

$$E_s = \left\| L(\boldsymbol{P}) - \tilde{L}^s(\boldsymbol{P}) \right\|_\infty,$$

and $\tilde{L}^s(\boldsymbol{P})$ is an approximation obtained in the sth stage to $L(\boldsymbol{P})$, the solution of the radiative transport equation (RTE). The geometric convergence means that the *rate of convergence* of the approximate solution $\tilde{L}^s(\boldsymbol{P})$ to the solution $L(\boldsymbol{P})$ is exponentially greater than the central limit theorem-constrained rate of non-adaptive methods. However, taking into account *both* variance *and* time, our true goal for adaptive methods is to exponentially increase the computational efficiency

$$\text{Eff} := \frac{1}{\text{Var} \times T},$$

when compared with non-adaptive Monte Carlo, where Var is the estimator variance, and T denotes total computation time.

We have demonstrated geometric convergence using both correlated sampling and importance sampling as the stage-to-stage variance reduction mechanisms. Our algorithms, as well as others developed at Los Alamos [1–3], also achieve geometric convergence but each faces implementation challenges and limitations. For example, for Sequential Correlated Sampling (SCS), the evaluation of the residual (i.e., the RTE equation error) and its use in generating a distributed source for each new adaptive stage creates unavoidable new sources of approximation errors. However, SCS is fast and very robust because each adaptive stage produces a correction to the estimate of the solution obtained from all of the previous stages. For Adaptive Importance Sampling (AIS), there is both a cost and loss of precision involved in sampling from the complex importance-modified expressions that result from altering the kernel K at each adaptive stage. On the plus side, AIS is very powerful and seems to produce the most rapid error reduction per adaptive stage of those adaptive methods we know.

Table 1 Comparison of the convergence characteristics of GWAS and AIS when each is used to estimate the solution of a 1D bidirectional RTE for which an exact solution is known

Method	S	Est.	$\|R\|$	σ^2	t	Rel. Eff
Exact	–	0.8964537768861041	–	–	–	∞
GWAS	60	0.8964537768857454	5.95×10^{-13}	4.29×10^{-21}	203,940	1.142×10^{15}
AIS	20	0.8964537768868207	5.36×10^{-12}	6.08×10^{-19}	743,100	2.212×10^{12}

In [19] we introduced a new adaptive Monte Carlo method—Generalized Weighted Analog Sampling (GWAS)—for the solution of RTEs. The idea behind GWAS is to combine the power of importance sampling with strategies that loosen the restrictions associated with sampling from importance-modified transport kernels. In this way, we hoped to combine rapid error reduction with fast algorithm execution in order to exponentially increase the computational efficiency. The price we pay for the flexibility of GWAS is that it biased. The fact that GWAS is biased (though asymptotically unbiased) greatly complicates the proof that it produces geometrically convergent estimates of RTE solutions. However, we have recently proved a new theorem that establishes that geometric convergence does obtain for GWAS [13]. As well, GWAS is able to provide increased computational efficiency compared with AIS, as we showed in [13]. The complete numerical results are provided in this recent publication, but here we repeat the table that summarizes this behavior.

We note that GWAS has a much higher computational efficiency than AIS because of its speed of execution. Note, too, that even though the variance of GWAS is more than 100 times as large as that of AIS, the efficiency of GWAS is more than 5,000 times that of AIS.

Each of these three adaptive methods:

1. generates biographies in stages, each of which consists of the same number of biographies;
2. applies variance reduction (correlated sampling, importance sampling, and GWAS) in each stage, linking stage s output to stage $(s + 1)$ input in an intrinsic way;
3. makes use of an analytic representation of the radiance.

Detailed examination of the behavior of these three adaptive algorithms reveals that *the need to represent the RTE solution by means of a* ***formula*** *introduces* ***bias*** *in its adaptive estimates*. This, in turn, prevents each algorithm from achieving unlimited precision as the number of adaptive stages tends to infinity. Thus, even though we don't *need* unlimited precision in order to make our adaptive algorithms useful, introducing an unknown amount of bias in our estimates falls short of our goal to create a new gold standard simulation tool for adoption by the biomedical community. In fact, the central obstacle to creating a real-time transport-rigorous Monte Carlo simulator is estimator bias (Table 1).

4 The Role of Bias in Estimating Radiance

The **theoretical bias** of an estimator $\Xi(\bar{b})$ of a linear functional

$$I = \int_{V \times \mathbb{S}^2} d(\boldsymbol{r}, \boldsymbol{\omega})\, L(\boldsymbol{r}, \boldsymbol{\omega})\, \mathrm{d}\boldsymbol{r}\, \mathrm{d}\boldsymbol{\omega},$$

of the radiance $L(\boldsymbol{r}, \boldsymbol{\omega})$ defined on the sample space $\mathbb{B}$ is

$$\mathbb{E}[\Xi(\bar{b})] - I = \int_{\mathbb{B}} \Xi(\bar{b})\, \mathrm{d}\mathbb{M}(\bar{b}) - I.$$

Theoretical bias introduces a component of *systematic* error in the mean integrated square error:

$$\begin{aligned} \mathrm{MISE} &= \mathbb{E}\left[\int_{\mathbb{B}} (\Xi(\bar{b}) - I)^2\, \mathrm{d}\mathbb{M}(\bar{b})\right] \\ &= \int_{\mathbb{B}} \left(\mathbb{E}[\Xi(\bar{b})] - I + \Xi(\bar{b}) - \mathbb{E}[\Xi(\bar{b})]\right)^2 \mathrm{d}\mathbb{M}(\bar{b}) \\ &= \int_{\mathbb{B}} \mathrm{BIAS}^2[\Xi(\bar{b})]\, \mathrm{d}\mathbb{M}(\bar{b}) + \int_{\mathbb{B}} \mathrm{Var}[\Xi(\bar{b})]\, \mathrm{d}\mathbb{M}(\bar{b}). \end{aligned} \tag{7}$$

Here $\mathbb{E}[]$ is the expected value operator and $\bar{b} \in \mathbb{B}$ is a photon biography (i.e., a sample). In contrast with theoretical bias, we will say that **computational bias** results from the accumulation of small errors due to the computer's limited precision. Computational bias is unavoidable in most cases. However, with sufficient care, computational sources of error can be controlled and estimated, whereas the source of error from theoretical bias is largely unknown and therefore much more difficult to estimate and to control.

The first term of the right hand side of (7) is the integral of the squared bias, while the second term is the integrated variance. Thus, for biased estimators it is necessary to control *both* the bias *and* the variance to exhibit geometric convergence.

Our approach to the avoidance of biased RTE estimators was to see whether the biased estimators often used in the graphics community could be improved or modified sufficiently to serve as the engine of a "gold standard" RTE solver. One of the conventional tools used for achieving realistic-looking scenes rapidly is **kernel density estimation** [17].

Kernel density estimation is a non-parametric method (i.e., no assumptions are made about the unknown underlying pdf) for recovering an unknown probability density function $f(x)$ by drawing samples $x_1, x_2, \ldots, x_n$ that are distributed according to $f(x)$. The kernel estimator with kernel k (satisfying $\int_{-\infty}^{\infty} k(x)\, \mathrm{d}x = 1$) produces the estimate $\tilde{f}(x)$ of the pdf $f(x)$ according to

$$\tilde{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} k\left(\frac{x - X_i}{h}\right). \tag{8}$$

where $X_1, \ldots, X_n$ are samples drawn independently from $f(x)$ and h is the window width (or smoothing parameter) that controls the influence of the kernel k near each sample point X_i. The kernel k can be chosen in a variety of ways: for example, as a standard Gaussian density

$$k_G(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2},$$

if x ranges over the entire real line, or as the Epanechnikov density

$$k_E(x) = \begin{cases} 6\left(\frac{1}{4} - x^2\right) & x^2 < \frac{1}{4}; \\ 0 & \text{otherwise}, \end{cases}$$

or in various other ways. Kernel Density Estimation is consistent: i.e., $\tilde{f}_h(x)$ converges to $f(x)$ as the number of samples n increases without limit, provided that the smoothing parameter h tends to 0 in such a way that the product hn tends to ∞. This last requirement means roughly that there are sufficiently many samples within the support sets of the kernel as the smoothing parameter is reduced.

Instead of applying conventional (unconstrained) density estimation to the RTE, we took the approach of constraining our method; i.e., to relate the RTE solution expansion directly to the random walks actually generated, treating these as the "samples" of the Monte Carlo simulation. Indeed, the sample space $\mathbb{B}$ is defined in this way [18], and the new Transport-Constrained Unbiased Radiance Estimator (T-CURE) does exactly this: it describes the expected contribution to the RTE solution at *any* point of phase space from *each* collision point of *every* photon biography. This creates an *unbiased representation* of the global RTE solution for all sample sizes that requires no smoothing parameters nor any special treatment of boundaries.

In the following section we will show how the T-CURE estimator can be derived as an extension of the conventional collision estimator [18].

4.1 T-CURE

We return to the integral equation characterized by Eq. (1), together with Eqs. (2), (3) and (4). The scattering integrals appearing on the right hand side of Eq. (1) are functions defined on the problem phase space Γ that are closely related to the RTE solution itself. We establish an unbiased estimator for those functions (hence, the RTE solutions) that extends the "conventional" collision estimators [18] to produce estimates of the entire RTE solution.

First, however, we consider the problem of computing the inner product of L, the solution of Eq. (1), and some $S^* : \Gamma \to \mathbb{R}$:

$$I = \langle S^*, L \rangle := \int_\Gamma S^*(\boldsymbol{r}, \boldsymbol{\omega}) L(\boldsymbol{r}, \boldsymbol{\omega}) \, \mathrm{d}\boldsymbol{\omega} \, \mathrm{d}\boldsymbol{r}. \tag{9}$$

Instead of estimating Eq. (9) directly, one can equally well solve its dual problem which leads to the same answer I:

$$I = \langle S, L^* \rangle := \int_\Gamma S(\boldsymbol{r}, \boldsymbol{\omega}) L^*(\boldsymbol{r}, \boldsymbol{\omega}) \, \mathrm{d}\boldsymbol{\omega} \, \mathrm{d}\boldsymbol{r}, \tag{10}$$

where L^* is the solution to the adjoint integral equation:

$$L^*(\boldsymbol{P}) = \int_\Gamma K^*(\boldsymbol{P} \to \boldsymbol{P}') \, L^*(\boldsymbol{P}') \, \mathrm{d}\rho \, \mathrm{d}\boldsymbol{\omega}' + S^*(\boldsymbol{P}), \tag{11}$$

with K^* being the *adjoint* of K satisfying $K^*(\boldsymbol{P} \to \boldsymbol{P}') = K(\boldsymbol{P}' \to \boldsymbol{P})$.

The conventional collision estimators for Eqs. (9) and (10) are

$$\eta(\bar{b}) := \frac{S(\boldsymbol{P}_1)}{p_1(\boldsymbol{P}_1)} \sum_{i=1}^{k} S^*(\boldsymbol{P}_i) \quad \text{and} \quad \eta^*(\bar{b}) := \frac{S^*(\boldsymbol{P}_1)}{p_1^*(\boldsymbol{P}_1)} \sum_{i=1}^{k} S(\boldsymbol{P}_i), \tag{12}$$

where $\bar{b} := (\boldsymbol{P}_1, \boldsymbol{P}_2, \ldots, \boldsymbol{P}_k)$ is a biography consisting of collision points $\boldsymbol{P}_i = (\boldsymbol{r}_i, \boldsymbol{\omega}_i)$ and k is the number of collisions made by $\bar{b}$ in the interior of the physical domain V of Γ. For the estimator η in Eq. (12), $\bar{b}$ is created using the following random walk process. The first collision $\boldsymbol{P}_1$ is drawn from a pre-defined density p_1 (which is normally selected to be proportional to S). At each collision $\boldsymbol{P}_i$, the next state, which can either be the next collision $\boldsymbol{P}_{i+1}$ or the termination of the random walk (leading to $k = i$), is determined using K^*. Similarly, when η^* is used, the photon biography $\bar{b}$ should be generated in the adjoint manner by sampling $\boldsymbol{P}_1$ based on S^* and the next state at collision $\boldsymbol{P}_i$ using K.

These conventional collision estimators (12) can be highly inefficient when S or S^* are concentrated in small subsets of Γ and vanishes everywhere else. This, unfortunately, is usually the case in biomedicine since many applications involve optical sources and/or detectors with small physical sizes (e.g., lasers and optical fibers). In particular, when the support of S^* is small, $S^*(\boldsymbol{P}_i)$ will be zero with high probability, making the estimator η inefficient. On the other hand, when S vanishes almost everywhere in the phase space, η will offer limited efficiency. Further, S and S^* can contain delta functions, making direct evaluations of $S^*(\boldsymbol{P}_i)$ in η and $S(\boldsymbol{P}_i)$ in η^* problematic.

To ease this problem, we introduce T-CURE estimators as an extension of the conventional collision estimators (12). For notational convenience, define operators $\mathscr{K}$ and $\mathscr{K}^*$ as functionals on $h : \Gamma \to \mathbb{R}$ as

$$(\mathscr{K}h)(\boldsymbol{P}) := \int_{\Gamma} K(\boldsymbol{P}' \to \boldsymbol{P})\, h(\boldsymbol{P}')\, \mathrm{d}\boldsymbol{\omega}'\, \mathrm{d}\rho, \tag{13}$$

$$(\mathscr{K}^*h)(\boldsymbol{P}) := \int_{\Gamma} K^*(\boldsymbol{P} \to \boldsymbol{P}')\, h(\boldsymbol{P}')\, \mathrm{d}\rho\, \mathrm{d}\boldsymbol{\omega}', \tag{14}$$

where the use of ρ is the same as in Eqs. (1)–(4). Then, Eqs. (1) and (11) respectively simplify to

$$L = \mathscr{K}L + S, \tag{15}$$

$$L^* = \mathscr{K}^*L^* + S^*. \tag{16}$$

It follows that

$$\langle S^*, L\rangle = \langle S^*,\ \mathscr{K}L + S\rangle = \langle S^*, H\rangle + \langle S^*, S\rangle, \tag{17}$$

$$\langle S, L^*\rangle = \langle S,\ \mathscr{K}^*L^* + S^*\rangle = \langle S, H^*\rangle + \langle S, S^*\rangle, \tag{18}$$

with $H := \mathscr{K}L$ and $H^* := \mathscr{K}^*L^*$. In Eqs. (17) and (18), $\langle S^*, S\rangle$ contains only known quantities and can be easily evaluated. Thus, we now focus on estimating the remaining terms $\langle S^*, H\rangle$ and $\langle S, H^*\rangle$.

By respectively applying $\mathscr{K}$ and $\mathscr{K}^*$ to both sides of Eqs. (15) and (16), we have

$$H = \mathscr{K}H + \mathscr{K}S, \tag{19}$$

$$H^* = \mathscr{K}^*H^* + \mathscr{K}^*S^*. \tag{20}$$

Notice that Eqs. (19) and (20) differ from Eqs. (15) and (16) only by the source terms. Therefore, $\langle S^*, H\rangle$ and $\langle S, H^*\rangle$ can be estimated using Eq. (12) with updated source terms, yielding

$$\begin{aligned} \eta_{\mathrm{NE}}(\bar{b}) &:= \frac{S(\boldsymbol{P}_1)}{p_1(\boldsymbol{P}_1)} \sum_{i=1}^{k} (\mathscr{K}^*S^*)(\boldsymbol{P}_i), \\ \eta^*_{\mathrm{NE}}(\bar{b}) &:= \frac{S^*(\boldsymbol{P}_1)}{p^*_1(\boldsymbol{P}_1)} \sum_{i=1}^{k} (\mathscr{K}S)(\boldsymbol{P}_i). \end{aligned} \tag{21}$$

Thanks to the integrals involved in $\mathscr{K}$ and $\mathscr{K}^*$, as defined in Eqs. (13) and (14), $\mathscr{K}S$ and $\mathscr{K}^*S^*$ generally have much greater supports than S and S^*, making our T-CURE estimators (21) significantly more effective than the conventional ones (12) when S and S^* vanish almost everywhere in the phase space.

Since both K and S (as well as K^* and S^*) are known, we can, in principle, evaluate these new extended next-event estimators (21) exactly. In particular, given Eqs. (13) and (14), it holds that

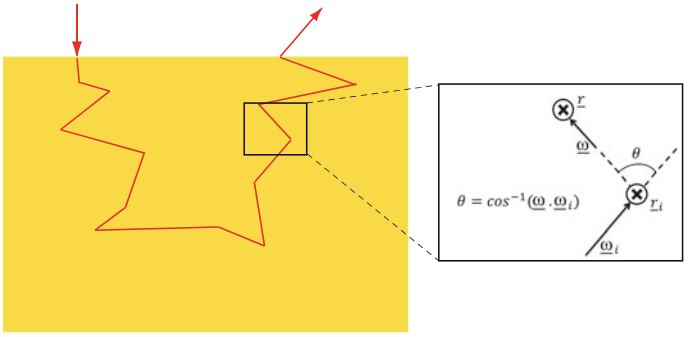

Fig. 2 Illustration of T-CURE Mechanism: Photon biography $\bar{b}$ moving in direction $\boldsymbol{\omega}_i$ collides at $\boldsymbol{r}_i$ in the blow-up at the right. Our estimator η_{NE} from Eq. (21) involves evaluating $\mathscr{K}^* S^*$ at each collision $(\boldsymbol{r}_i, \boldsymbol{\omega}_i)$. We estimate $(\mathscr{K}^* S^*)(\boldsymbol{r}_i, \boldsymbol{\omega}_i)$ by selecting some $\boldsymbol{\omega} \in \mathbb{S}^2$ and $\boldsymbol{r} \in V$ based on $K^*(\boldsymbol{r}_i, \boldsymbol{\omega}_i \to \boldsymbol{r}, \boldsymbol{\omega})\, S^*(\boldsymbol{r}_i, \boldsymbol{\omega}_i)$ but independent of $(\boldsymbol{r}_{i+1}, \boldsymbol{\omega}_{i+1})$

$$(\mathscr{K}^* S^*)(\boldsymbol{r}_i, \boldsymbol{\omega}_i) := \int_\Gamma K^*(\boldsymbol{r}_i, \boldsymbol{\omega}_i \to \boldsymbol{r}_i + \rho\boldsymbol{\omega}, \boldsymbol{\omega})\, S^*(\boldsymbol{r}, \boldsymbol{\omega})\, \mathrm{d}\rho\, \mathrm{d}\boldsymbol{\omega}, \tag{22}$$

$$(\mathscr{K} S)(\boldsymbol{r}_i, \boldsymbol{\omega}_i) := \int_\Gamma K(\boldsymbol{r}_i - \rho\boldsymbol{\omega}_i, \boldsymbol{\omega} \to \boldsymbol{r}_i, \boldsymbol{\omega}_i)\, S(\boldsymbol{r}, \boldsymbol{\omega})\, \mathrm{d}\boldsymbol{\omega}\, \mathrm{d}\rho. \tag{23}$$

In practice, the integrals in Eqs. (22) and (23) can be evaluated analytically or numerically depending on the exact forms of S and S^*. Lastly, the evaluation of the T-CURE tallies from every collision at $(\boldsymbol{r}_i, \boldsymbol{\omega}_i)$ can be accomplished either on-the-fly or by post-processing all of the biography collision points saved from the generation of the photon biographies. This is because the T-CURE tally at $(\boldsymbol{r}_i, \boldsymbol{\omega}_i)$ is independent of the process that generates the next collision at $(\boldsymbol{r}_{i+1}, \boldsymbol{\omega}_{i+1})$. See Fig. 2 for an illustration of this process.

Besides offering superior computational efficiency, our new T-CURE estimators also enjoy the following properties:

- They require the imposition of no mesh on the phase space, so they provide the basis for plotting or otherwise displaying features of the RTE solution over any desired mesh, or of several such, based on a single set of biographies.
- They can be implemented either with "on-the-fly" computation or, after generating a "smallish" number N_0 of biographies, by post-processing key data stored from the N_0 "baseline" set of biographies, or by a combination of these two methods.

The power of T-CURE estimation derives from replacing occasional contributions to reflectance from biographies that are *actually* detected (i.e., the terminal estimate of reflectance) by sums of analytic formulas over *all* the collisions of biographies, whether or not they ultimately reach the detector. Not surprisingly, this replacement often (though not always!) achieves variance reduction in estimates of reflectance.

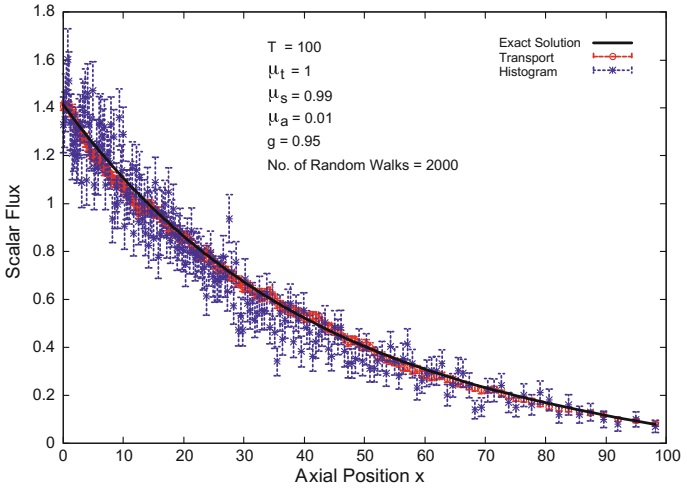

Fig. 3 Comparison of transport-constrained density estimation with both histogram and Epanechnikov density estimators. The exact solution of this 1D problem is known and constitutes the black curve, while the two more conventional density estimators are indicated in blue and green. Estimated standard deviations are shown as error bars

Previously, we have shown that T-CURE is unbiased for all sample sizes, in sharp contrast with the conventional (unconstrained) density estimators [12]. As well, T-CURE is roughly an order of magnitude more accurate than the conventional ones found in the statistical literature [6, 15, 16]. Figure 3 (from our prior work [12]) illustrates these gains in a 1-D model RTE problem that plots the (scalar) intensity of the light field against distance, x, from the light source.

5 Recent Numerical Studies: T-CURE in Multi-dimensions

The comparison of T-CURE performance with that of the histogram and the Epanechnikov density estimators was motivated by the desire to compare its computational efficiency as a constrained "density estimator" with that of the more conventional unconstrained density estimators that are widely used by the graphics community. The bidirectional problem is especially simple since it involves only one spatial dimension and two discrete scattering directions. In order to investigate how the T-CURE estimator behaves when applied to more challenging problems, we turned to multilayer tissue problems. For all of these numerical experiments we used input data typical of normal cervical tissue. The tissue was represented as two layers: a top epithelial layer and the stromal layer below. Optical data for this 5-D problem is shown in the table:

Layer	Optical data			Layer thickness
	μ_s	μ_a	g	
Epithelium	~80/cm [5]	0.12/cm	0.95	360 μm [21]
Stroma	150/cm [4]	0.15–1.2/cm[a]	0.88	∞

[a] Particular value for μ_a depends on whether there is Hb absorption at that wavelength

Intuitively, we do not expect the performance of T-CURE to degrade significantly as the dimension of the underlying phase space increases. This is because the overhead associated with T-CURE depends mainly on the *number of collisions, which is determined by the optical properties, not the dimension of the phase space*.

Next we illustrate the use of T-CURE to estimate spatially resolved reflectances in two tissue problems. Both experiments used the two-layer data for normal cervical tissue shown in the table above but ignore possible refractive index mismatches at the interfaces (Fig. 4).

Note especially the graininess of the terminal estimator and also how the information degrades as the source-detector distances increase. This graininess contrasts with the much greater smoothness of the T-CURE images, even at 10.0 cm from the source where the signal is quite low. Note also the scales along the y-axis of the images. Even though the magnitude of the signal falls by about two orders of magnitude as the source and detector radii shrink, the T-CURE 95% relative confidence interval sizes remain below 3.5% over the entire range of source-detector distances, while the terminal 95% confidence interval sizes grow from 15 to 100% as the s-d distances grow from 1.0 to 10.0 cm (Fig. 5).

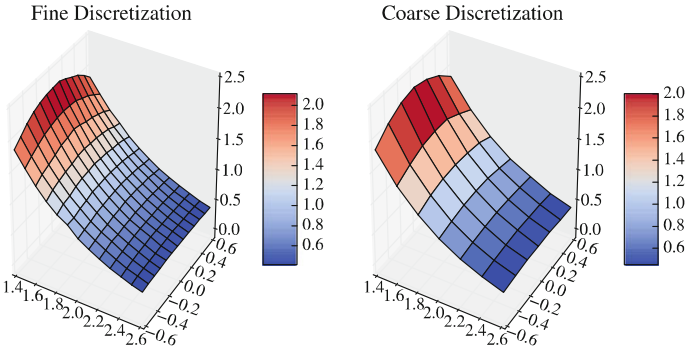

Fig. 4 5-D Cervical Tissue Problem. This example displays another advantage of T-CURE when compared to more traditional estimators: one can generate a "smallish" initial set of biographies, show the output from that, and then refine the output mesh by post-processing the initial set without generating any new biographies. For example, the plot on the right was produced by processing biographies "on the fly" (6 min), while the refined plot on the left was obtained by post-processing stored data from these biographies (few seconds)

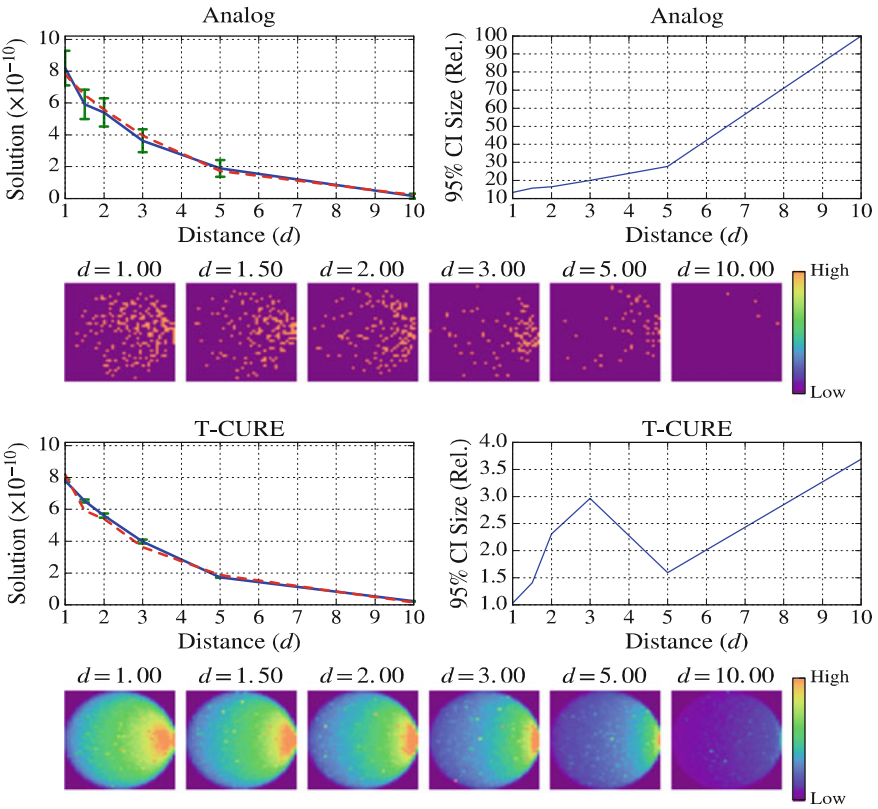

Fig. 5 In this problem, there is a single source at (0, 0, 0) and six detectors spaced at distances 1.0, 1.5, 2.0, 3.0, 5.0 and 10.0 cm from the source. Both the source and detector are discs with identically small radii $R = 0.0025$ cm. We applied both the terminal (top) and our T-CURE (bottom) estimators to this problem by letting them run for approximately 1700 s for each data point. In both experiments, the left plot shows estimation results and the right plot shows the relative size of 95% confidence intervals (obtained using twice the standard deviation of the sample data). These confidence intervals are also drawn in the left plots (as error bars). Lastly, the red dashed lines indicate the solution obtained using the other estimator, demonstrating that our analog and T-CURE estimators converge to the same answer

6 Summary and Future Work

We have advanced our earlier work [12] by examining the computational efficiency of T-CURE on multidimensional, heterogeneous (multilayer) tissue problems involving 5 independent variables: 3 for position and 2 for unit direction. We observe that the advantages of T-CURE are maintained in the higher dimensional cases. We believe that this new estimator has the potential to be the computational engine of an adaptive (geometrically convergent) algorithm.

We will turn our attention next to the fully general multilayer tissue problem, including the possibility of refractive index mismatches at the layer interfaces. This introduces new challenges that call for a strategy that controls the run time consumption caused by photon biographies that can cross refractive-index-mismatched layer interfaces in both directions a very large number of times. Provided that this degradation of the computational power of the adaptive algorithm can be controlled, the resulting T-CURE-based adaptive algorithm should serve as a new gold standard Monte Carlo solver for biomedical problems.

Acknowledgements The third author gratefully acknowledges partial support from award numbers: P41RR001192 from the National Center for Research Resources and P41EB015890 from the National Institute of Biomedical Imaging and Bioengineering.

The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the National Center For Research Resources, National Institute of Biomedical Imaging and Bioengineering, or the National Institutes of Health.

7 Appendix: Transport Equations

The time independent transport equation can be written as

$$(\boldsymbol{\omega}\cdot\nabla)L+\mu_t(\boldsymbol{r})L(\boldsymbol{r},\boldsymbol{\omega})=\int_{\mathbb{S}^2}\mu_s(\boldsymbol{r})f(\boldsymbol{r}';\ \boldsymbol{\omega}'\to\boldsymbol{\omega})\,L(\boldsymbol{r},\boldsymbol{\omega}')\,\mathrm{d}\boldsymbol{\omega}'+Q(\boldsymbol{r},\boldsymbol{\omega}),\tag{24}$$

for all $r\in V\subseteq\mathbb{R}^3$ and

$$L(\boldsymbol{r},\boldsymbol{\omega})=Q_0(\boldsymbol{r},\boldsymbol{\omega}),\tag{25}$$

for all $\boldsymbol{r}\in\partial V$ and $\boldsymbol{\omega}\in\mathbb{S}^2$ satisfying $\boldsymbol{\omega}\cdot\boldsymbol{n}(\boldsymbol{r})<0$ as the boundary condition. In Eqs. (24) and (25), $\boldsymbol{r}:=(x,y,z)$ and $\boldsymbol{n}(\boldsymbol{r})$ denotes the outward unit normal on the boundary ∂V at $\boldsymbol{r}$. Note that

$$\boldsymbol{\omega}=\omega_x\boldsymbol{i}+\omega_y\boldsymbol{j}+\omega_z\boldsymbol{k}=\sin\theta\cos\phi\,\boldsymbol{i}+\sin\theta\sin\phi\,\boldsymbol{j}+\cos\theta\,\boldsymbol{k},\tag{26}$$

$$\boldsymbol{\omega}\cdot\nabla=\sin\theta\cos\phi\frac{\partial}{\partial x}+\sin\theta\sin\phi\frac{\partial}{\partial y}+\cos\theta\frac{\partial}{\partial z},\tag{27}$$

we now convert Eq. (24) to an integral equation using the method of characteristics. Consider the following characteristic system for Eq. (24):

$$\frac{\mathrm{d}x}{\mathrm{d}\rho}=-\omega_x,\qquad\frac{\mathrm{d}y}{\mathrm{d}\rho}=-\omega_y,\qquad\frac{\mathrm{d}z}{\mathrm{d}\rho}=-\omega_z.\tag{28}$$

Solving Eq. (28) produces

$$x = x_0 - \omega_x \rho, \qquad y = y_0 - \omega_y \rho, \qquad z = z_0 - \omega_z \rho. \tag{29}$$

We can then write Eq. (24) in the following form:

$$-\frac{\mathrm{d}L(\boldsymbol{r}', \boldsymbol{\omega})}{\mathrm{d}\rho} + \mu_t(\boldsymbol{r}')L(\boldsymbol{r}', \boldsymbol{\omega}) = \int_{\mathbb{S}^2} \mu_s(\boldsymbol{r}')\, f(\boldsymbol{r}';\ \boldsymbol{\omega}' \to \boldsymbol{\omega})\, L(\boldsymbol{r}',\ \boldsymbol{\omega}')\, \mathrm{d}\boldsymbol{\omega}' + Q(\boldsymbol{r}', \boldsymbol{\omega}), \tag{30}$$

where $\boldsymbol{r}' := \boldsymbol{r}_0 - \rho\boldsymbol{\omega}$, which can in turn be rewritten as:

$$\begin{aligned} &-\frac{\mathrm{d}}{\mathrm{d}\rho}\left[\mathrm{e}^{-\int_0^\rho \mu_t(\boldsymbol{r}_0 - \tau\boldsymbol{\omega})\,\mathrm{d}\tau} L(\boldsymbol{r}',\ \boldsymbol{\omega})\right] \\ &= \mathrm{e}^{-\int_0^\rho \mu_t(\boldsymbol{r}_0 - \tau\boldsymbol{\omega})\,\mathrm{d}\tau}\left[\int_{\mathbb{S}^2} \mu_s(\boldsymbol{r}')f(\boldsymbol{r}';\ \boldsymbol{\omega}' \to \boldsymbol{\omega})L(\boldsymbol{r}',\ \boldsymbol{\omega}')\,\mathrm{d}\boldsymbol{\omega}' + Q(\boldsymbol{r}', \boldsymbol{\omega})\right], \end{aligned} \tag{31}$$

We now integrate the two sides of Eq. (30) with respect to ρ from $\rho = 0$ to $\rho = R$ and replace $\boldsymbol{r}_0$ with $\boldsymbol{r}$ to produce:

$$\begin{aligned} &L(\boldsymbol{r}, \boldsymbol{\omega}) \\ &= \int_0^R \mathrm{e}^{-\int_0^\rho \mu_t(\boldsymbol{r} - \tau\boldsymbol{\omega})\,\mathrm{d}\tau}\left[\mu_s(\boldsymbol{r}')\int_{\mathbb{S}^2} f(\boldsymbol{r}';\ \boldsymbol{\omega}' \to \boldsymbol{\omega})\, L(\boldsymbol{r}', \boldsymbol{\omega}')\mathrm{d}\boldsymbol{\omega}' + Q(\boldsymbol{r}', \boldsymbol{\omega})\right]\mathrm{d}\rho \\ &\quad + \mathrm{e}^{-\int_0^R \mu_t(\boldsymbol{r} - \tau\boldsymbol{\omega})\,\mathrm{d}\tau}\, Q_0(\boldsymbol{r} - R\boldsymbol{\omega},\ \boldsymbol{\omega}). \end{aligned} \tag{32}$$

or

$$L(\boldsymbol{P}) = \int_\Gamma K(\boldsymbol{P}' \to \boldsymbol{P})\, L(\boldsymbol{P}')\, \mathrm{d}\boldsymbol{\omega}'\, \mathrm{d}\rho + S(\boldsymbol{P}), \tag{33}$$

where $\boldsymbol{P} := (\boldsymbol{r}, \boldsymbol{\omega})$, $\boldsymbol{P}' := (\boldsymbol{r}', \boldsymbol{\omega}')$,

$$\begin{aligned} K(\boldsymbol{P}' \to \boldsymbol{P}) &:= \frac{\mu_s(\boldsymbol{r}')}{\mu_t(\boldsymbol{r}')}\, f(\boldsymbol{r}'; \boldsymbol{\omega}' \to \boldsymbol{\omega})\, T(\boldsymbol{r}' \to \boldsymbol{r};\ \boldsymbol{\omega}), \\ T(\boldsymbol{r}' \to \boldsymbol{r};\ \boldsymbol{\omega}) &:= \mu_t(\boldsymbol{r}')\, \exp\left(-\int_0^{\|\boldsymbol{r} - \boldsymbol{r}'\|} \mu_t(\boldsymbol{r} - \tau\boldsymbol{\omega})\,\mathrm{d}\tau\right), \\ S(\boldsymbol{P}) &:= \mathrm{e}^{-\int_0^R \mu_t(\boldsymbol{r} - \tau\boldsymbol{\omega})\,\mathrm{d}\tau}\, Q_0(\boldsymbol{r} - R\boldsymbol{\omega},\ \boldsymbol{\omega}) \\ &\quad + \int_0^R \mathrm{e}^{-\int_0^\rho \mu_t(\boldsymbol{r} - \tau\boldsymbol{\omega})\,\mathrm{d}\tau}\, Q(\boldsymbol{r} - \rho\boldsymbol{\omega},\ \boldsymbol{\omega})\mathrm{d}\rho. \end{aligned}$$

Eqs. (32) and (33) are the integral forms of the RTE that we seek.

References

1. Baggerly, K., Cox, D., Picard, R.: Exponential convergence of adaptive importance sampling for Markov chains. J. Appl. Probab. **37**(2), 342–358 (2000)
2. Booth, T.: Exponential convergence on a continuous Monte Carlo transport problem. Nucl. Sci. Eng. **127**(3), 338–345 (1997)
3. Booth, T.: Adaptive importance sampling with a rapidly varying importance function. Nucl. Sci. Eng. **136**(3), 399–408 (2000)
4. Chang, V.T.C., Cartwright, P.S., Bean, S.M., Palmer, G.M., Bentley, R.C., Ramanujam, N.: Quantitative physiology of the precancerous cervix in vivo through optical spectroscopy. Neoplasia **11**(4), 325–332 (2009)
5. Collier, T., Arifler, D., Malpica, A., Follen, M., Richards-Kortum, R.: Determination of epithelial tissue scattering coefficient using confocal microscopy. IEEE J. Sel. Top. Quantum Electron. **9**(2), 307–313 (2003)
6. Devroye, L.: A Course in Density Estimation. Progress in Probability. Birkhauser, Boston (1987)
7. Jensen, H.W.: Realistic Image Synthesis Using Photon Mapping. AK Peters Ltd, Wellesley (2001)
8. Kong, R.: Transport problems and Monte Carlo methods. Ph.D. thesis, Claremont Graduate University (1999)
9. Kong, R., Ambrose, M., Spanier, J.: Efficient, automated Monte Carlo methods for radiation transport. J. Comput. Phys. **227**(22), 9463–9476 (2008)
10. Kong, R., Spanier, J.: A new proof of geometric convergence for general transport problems based on sequential correlated sampling methods. J. Comput. Phys. **227**(23), 9762–9777 (2008)
11. Kong, R., Spanier, J.: Geometric convergence adaptive Monte Carlo algorithms for radiative transport problems based on importance sampling methods. Nucl. Sci. Eng. **168**(3), 197–225 (2011)
12. Kong, R., Spanier, J.: Transport-constrained extensions of collision and track length estimators for solutions of radiative transport problems. J. Comput. Phys. **242**(0), 682–695 (2013). https://doi.org/10.1016/j.jcp.2013.02.023, http://www.sciencedirect.com/science/article/pii/S0021999113001423
13. Kong, R., Spanier, J.: A new proof of geometric convergence for the adaptive generalized weighted analog sampling (GWAS) method. Monte Carlo Methods Appl. **22**(3), 161–196 (2016)
14. Lai, Y., Spanier, J.: Adaptive importance sampling algorithms for transport problems. Monte Carlo and Quasi-MonteCarlo Methods 1998, pp. 273–283. Springer, Berlin (1999)
15. Rao, C.R.: Linear Statistical Inference and Its Applications. Wiley, New York (1973)
16. Scott, D.: Multivariate Density Estimation: Theory. Practice and Visualization. Wiley Series in Probability and Mathematical Statistics. Wiley, New York (1990)
17. Silverman, B.: Density Estimation for Statistics and Data Analysis. Chapman and Hall, London (1986)
18. Spanier, J., Gelbard, E.: Monte Carlo Principles and Neutron Transport Problems. Wesley, New York (1969). (reprinted by Dover Publications, Inc. 2008)
19. Spanier, J., Kong, R.: A new adaptive method for geometric convergence. Monte Carlo and Quasi-MonteCarlo Methods 2002, pp. 439–449. Springer, Berlin (2004)
20. Veach, E.: Robust Monte Carlo methods for light transport simulation. Ph.D. thesis, Stanford, CA, USA (1998). AAI9837162
21. Walker, D., Brown, B., Blackett, A., Tidy, J., Smallwood, R.: A study of the morphological parameters of cervical squamous epithelium. Physiol. Meas. **24**(1), 121 (2003)
22. Wilson, B.C., Adam, G.: A Monte Carlo model for the absorption and flux distributions of light in tissue. Med. Phys. **10**(6), 824–830 (1983)