

A Bayesian Monte Carlo Approach to Global Illumination

Jonathan Brouillat¹, Christian Bouville¹, Brad Loos², Charles Hansen² and Kadi Bouatouch¹

¹University of Rennes 1 / INRIA Rennes, France

²School of Computing, Salt Lake City, UT, USA

Abstract

Most Monte Carlo rendering algorithms rely on importance sampling to reduce the variance of estimates. Importance sampling is efficient when the proposal sample distribution is well-suited to the form of the integrand but fails otherwise. The main reason is that the sample location information is not exploited. All sample values are given the same importance regardless of their proximity to one another. Two samples falling in a similar location will have equal importance whereas they are likely to contain redundant information. The Bayesian approach we propose in this paper uses both the location and value of the data to infer an integral value based on a prior probabilistic model of the integrand. The Bayesian estimate depends only on the sample values and locations, and not how these samples have been chosen. We show how this theory can be applied to the final gathering problem and present results that clearly demonstrate the benefits of Bayesian Monte Carlo.

Keywords: global illumination, Bayesian Monte Carlo, photon mapping

ACM CCS: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism

1. Introduction

Monte Carlo integration has now become an essential tool in the field of global illumination. It allows accurate simulation of light transport mechanisms and leads to photorealistic images. However, its computational cost remains high due to the huge number of ray-traced samples that need to be collected before reaching an acceptable noise level. This problem has given rise to an extensive literature of which many of the proposed solutions are based on importance sampling and/or (to a less extent) control variates. Both techniques try to exploit some prior knowledge of the integrand so as to build an approximating function. This approximating function is then used either to optimize samples distribution (importance sampling) or to extract a known deterministic part of the integrand so as to reduce its variance.

In [O'H87], O'Hagan raises two fundamental objections to the importance sampling procedure. First, the estimator depends on the arbitrary choice of the sampling density. Therefore, the same set of observed integrand sample values will lead to different estimates depending on the chosen proposal distribution, which violates the Likelihood Principle [Bir62].

Briefly, the Likelihood Principle states that 'in the inference about θ , after x is observed, all relevant experimental information is contained in the likelihood function ($p(x|\theta)$) for the observed x '. Consequently, the rules governing the data collection process are irrelevant for computing the estimate. Even though the choice of the sampling density is guided by prior information we have on the integrand, the objection still holds as it remains some freedom on the choice of the sampling density.

The second objection is that Monte Carlo procedures ignore the sample locations, only the sample values of the integrand are used. Therefore, two samples falling on the same or close location will have equal importance in the estimation process, whereas the second sample brings no extra information. Of course, such occurrence can be made unlikely with stratification or even impossible with quasi-Monte Carlo (QMC) but nonetheless, the fact remains that classical Monte Carlo wastes important information.

To avoid these inconsistencies and allow better estimates, O'Hagan turns the problem of evaluating the integral into a Bayesian inference problem [O'H91]. To this end, he

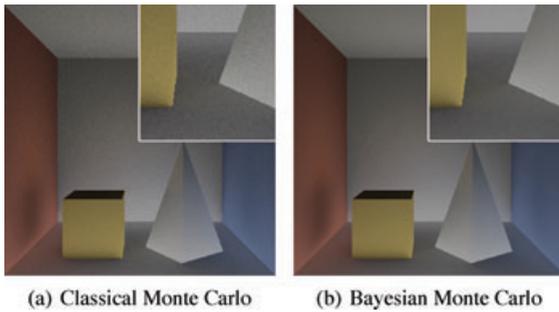


Figure 1: *Cornell Box rendering (Indirect only). $n = 256$. To obtain the same quality, Monte Carlo integration would require more than 900 rays.*

assumes a Gaussian Process (GP) behavior for the integrand which leads to a new form of quadrature called ‘Bayes-Hermite quadrature’. His results have been further generalized and applied in various domains related to Machine Learning theory under the generic name of ‘Bayesian Monte Carlo’ (BMC) [PHR06, KNKS08].

The goal of this paper is to show how the BMC method can be applied to global illumination models while trying to obtain the best compromise between computational complexity and effective noise reduction. However, as BMC significantly departs from conventional MC, it would be too long to analyze and compare in details all theoretical aspects and implementation options in this paper. Besides, many research directions need to be investigated before assessing the full potential of this new approach. Therefore, our presentation of the theoretical background only covers the minimum necessary to understand the application to global illumination and, although only one implementation method is described, we will briefly present alternative approaches.

We will first present a brief introduction to the theoretical background of BMC integration and will discuss in general the important implementation issues and options. Then, we will present a first attempt to apply BMC to compute the integral involved in the final gathering phase of a photon mapping algorithm. We show how a GP model can be built in this particular case and how the quadrature coefficients can be derived from this model. We then address the problem of optimal sampling and hyperparameters estimation, and present our solutions to these problems. As we will see, we have obtained very promising results despite suboptimal hyperparameters (Figure 1).

2. Related Work

Although the theory of Bayesian Quadrature has been known since 1991 [O’H91], it is only recently that it has been put into practice (e.g. [PHR06, KNKS08]) and to the authors’ knowl-

edge, no research in this direction has yet been reported in the computer graphics literature. In the field of global illumination, most research works are based on importance sampling and although this variance reduction technique can be applied to many cases (e.g. [CAE08]), its efficiency is questionable when prior information does not allow to find a well-focused proposal distribution. This is particularly noticeable when diffuse reflection is involved since the cosine distribution law gives rise to rather scattered sampling directions.

The control variates method has been proposed to supplement importance sampling by introducing a correlated function which approximates the integrand with a constant difference term [LW94]. In multiple importance sampling (MIS), the control variate estimate is built with the same component functions as the ones used in the mixture probability density functions [OZ00, FCH*06, Vea97]. When control variates is used without importance sampling, finding the optimal mixture coefficients amounts to a least square fitting of the integrand as shown in [HLO04]. There are some similarities with BMC with regard to this point since BMC also involves a regression step. However, as discussed in [RW06], Bayesian regression differs in that it is a non-parametric method, which means (in brief) that the basis functions of the feature space and their associated weights do not appear explicitly. Consequently, the computational complexity does not depend on the dimension of the feature space but on the size of the sample set only. Actually, the dimension of the feature space implicitly involved in a Bayesian regression may be infinite.

3. Issues at Stake

In [RG02], Rasmussen and Ghahramani have shown that BMC can significantly outperform any classical importance sampling method if appropriate prior knowledge is used. However, such ideal conditions are rarely met in the context of global illumination. The main reason is that the GP prior used in the BMC method is very effective in reducing the estimator variance when the integrand is a smooth function of the input variables. This smooth integrand behavior is rare in global illumination models mainly because of the sharp changes of illumination at object boundaries [DLAW01]. We will see that the features of illumination statistics can be taken into account in the GP model so as to alleviate the effects of discontinuities. Surprisingly, we have already obtained good results with a very simple GP model and largely suboptimal hyperparameters, which raises hopes of wide margin for improvement.

4. Background Theory

4.1. Notation

A generic point in the D dimensional unit cube $[0, 1]^D$ is denoted by $\mathbf{x} = (x^1, \dots, x^D)^t$ while a random point in the same space is denoted $\mathbf{X} = (X^1, \dots, X^D)^t$ and a point used

in an integration rule is \mathbf{X}_i . In conventional Monte Carlo methods, \mathbf{X}_i is a sample of a random vector \mathbf{X} drawn from a given density.

4.2. Bayesian quadrature equations

We consider the problem of computing the integral I of the product of a function $f(\mathbf{x})$ with a weight function $p(\mathbf{x})$, both functions defined on the D dimensional unit cube $[0, 1]^D$:

$$I = \int f(\mathbf{x})p(\mathbf{x})d\mathbf{x} \quad \text{with } \mathbf{x} \in [0, 1]^D. \quad (1)$$

Here and elsewhere, integrals without explicit range are understood to be over $[0, 1]^D$. Typically, $p(\mathbf{x})$ is known under its analytical form whereas $f(\mathbf{x})$ can only be evaluated numerically or through computer simulation. $p(\mathbf{x})$ is not required to be a probability density function, however we will assume that it is normalized:

$$\int p(\mathbf{x})d\mathbf{x} = 1.$$

In Bayesian Monte Carlo (BMC) the problem of evaluating the integral in Equation (1) is turned into a Bayesian inference problem, which allows avoiding the inconsistencies discussed in Section 1 and leads to a better estimate. As proposed by O’Hagan in [O’H91], BMC is based on the following reasoning: $f(\mathbf{x})$ is considered as random simply because it is unknown (and thus uncertain) before its evaluation. This view may seem to be somewhat inappropriate but it is totally consistent with the Bayesian approach that all forms of uncertainty can be modeled by probabilities. For this purpose, we will put a prior on $f(\mathbf{x})$ by using a GP. Detailed presentation of GPs is beyond the scope of this paper, a comprehensive introduction to GPs can be found in [RW06]. Formally, a GP is a collection of random variables, any finite number of which have a joint Gaussian distribution. A GP is completely defined by its mean function $\bar{f}(\mathbf{x})$ and its covariance function $k(\mathbf{x}, \mathbf{x}')$ which depends only on the input \mathbf{x} , i.e. the sample locations:

$$\begin{aligned} \bar{f}(\mathbf{x}) &= E[f(\mathbf{x})] \\ k(\mathbf{x}, \mathbf{x}') &= E[(f(\mathbf{x}) - \bar{f}(\mathbf{x}))(f(\mathbf{x}') - \bar{f}(\mathbf{x}'))] \end{aligned} \quad (2)$$

and will be denoted by

$$f(\mathbf{x}) \sim \mathcal{GP}[\bar{f}(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')].$$

The choice of the covariance function $k(\mathbf{x}, \mathbf{x}')$ allows us to introduce prior knowledge on the smoothness properties of the integrand. However, this choice is not arbitrary as the covariance function must be a symmetric ($k(\mathbf{x}, \mathbf{x}') = k(\mathbf{x}', \mathbf{x})$) and positive semidefinite kernel (see [RW06] p. 80 for details). A covariance function is stationary when it is invariant to translation, i.e. $k(\mathbf{x}, \mathbf{x}') = k(\mathbf{x} - \mathbf{x}')$. Moreover, the covariance function is isotropic when it is a function only of $r = |\mathbf{x} - \mathbf{x}'|$. In this case, these are also known as *radial basis functions* (RBFs). Furthermore, Bochner’s theorem states

that a stationary covariance function is positive semidefinite if and only if its Fourier transform is non-negative [RW06].

A common choice of the stationary covariance function is the squared exponential (SE) covariance function:

$$k(\mathbf{x}_1, \mathbf{x}_2) = w_0^2 \exp \left\{ -\frac{1}{2} \sum_{d=1}^D \left(\frac{x_1^{(d)} - x_2^{(d)}}{w_d} \right)^2 \right\}, \quad (3)$$

where the w_i ’s are the hyperparameters of the model. With this function, close function samples are highly correlated whereas $k(\mathbf{x}_1, \mathbf{x}_2) \approx 0$ for distant samples, which means that the function values $f(\mathbf{x}_1)$ and $f(\mathbf{x}_2)$ are almost independent. Let us note that $w_0^2 = k(\mathbf{x}, \mathbf{x}) = \text{Var}[f(\mathbf{x})]$. The other hyperparameters w_1, \dots, w_D are the lengthscales of the individual input dimensions. The SE covariance function is thus isotropic when all the lengthscales are equal. A GP having a SE covariance function has mean square derivatives of all orders and is thus very smooth. This strong smoothness might not be appropriate for illumination models. The Matérn class [RW06] of covariance functions have been recommended by Stein when such smoothness is unrealistic. Although we have already obtained good results with SE covariance function in the application described below, we will consider the Matérn class option in future works.

Now that we have specified the GP prior that models our beliefs about $f(\mathbf{x})$, let us suppose that we are provided with a set of noisy samples $\mathcal{D} = \{(\mathbf{X}_i, Y_i) \mid i = 1, \dots, n\}$ where Y_i is the observed value of f at point \mathbf{X}_i :

$$Y_i = f(\mathbf{X}_i) + \varepsilon_i$$

and the ε_i are samples of an independent, identically distributed Gaussian distribution with zero mean and variance σ^2 . Then, in application of the Bayes’ rule, the posterior process is also a GP with mean and covariance given by [RW06]:

$$\begin{aligned} E[f(\mathbf{x}) \mid \mathcal{D}] &= \bar{f}(\mathbf{x}) + \mathbf{k}(\mathbf{x})' Q^{-1} (\mathbf{Y} - \bar{\mathbf{F}}) \\ \text{Cov}[f(\mathbf{x}), f(\mathbf{x}') \mid \mathcal{D}] &= k(\mathbf{x}, \mathbf{x}') - \mathbf{k}(\mathbf{x}')' Q^{-1} \mathbf{k}(\mathbf{x}) \end{aligned} \quad (4)$$

where

$$\begin{aligned} \mathbf{k}(\mathbf{x}) &= (k(\mathbf{X}_1, \mathbf{x}), \dots, k(\mathbf{X}_n, \mathbf{x}))' \\ Q &= (K + \sigma^2 I_n) \\ K_{i,j} &= k(\mathbf{X}_i, \mathbf{X}_j) \quad \text{with } (i, j) \in [1, n]^2 \\ \mathbf{Y} &= (Y_1, \dots, Y_n)' \\ \bar{\mathbf{F}} &= (\bar{f}(\mathbf{X}_1), \dots, \bar{f}(\mathbf{X}_n))' \end{aligned}$$

where I_n is the identity matrix. Equation (4) gives the expected value (or the mean prediction) of f for an unseen input \mathbf{x} given the observed data \mathcal{D} . This particular form of regression is called *Bayesian regression*. When the covariance function is a RBF, Bayesian Regression is not to be confused with RBF non-linear regression (see [Her04, RW06] for more details).

From this, we can derive the posterior distribution of the integral I given by Equation (1). As integration is a linear operation, the distribution of I is Gaussian with mean and variance [O'H91]:

$$E(I | \mathcal{D}) = \bar{I} + \mathbf{z}' \mathbf{Q}^{-1} (\mathbf{Y} - \bar{\mathbf{F}}) \quad (5)$$

$$\text{Var}(I | \mathcal{D}) = \bar{V} - \mathbf{z}' \mathbf{Q}^{-1} \mathbf{z}, \quad (6)$$

where

$$\begin{aligned} \bar{I} &= \int \bar{f}(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} \\ \mathbf{z} &= \int \mathbf{k}(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} \\ \bar{V} &= \int \int k(\mathbf{x}, \mathbf{x}') p(\mathbf{x}) p(\mathbf{x}') d\mathbf{x} d\mathbf{x}' \end{aligned} \quad (7)$$

The estimate of I , given our GP prior and the observed data \mathcal{D} , is then $\hat{I}_{BMC} = E(I | \mathcal{D})$. Note that \bar{I} and \bar{V} are respectively the prior mean and variance of I , i.e. the \hat{I}_{BMC} estimate and its variance when $n = 0$ in Equations (5) and (6).

The \mathbf{z} vector can also be expressed in terms of the function $f_z(\mathbf{x}')$ given by:

$$f_z(\mathbf{x}') = \int k(\mathbf{x}, \mathbf{x}') p(\mathbf{x}) d\mathbf{x}. \quad (8)$$

Then, we have

$$\mathbf{z} = (f_z(\mathbf{X}_1), \dots, f_z(\mathbf{X}_n))'$$

and Equation (7) can be rewritten as

$$\bar{V} = \int f_z(\mathbf{x}) p(\mathbf{x}) d\mathbf{x}. \quad (9)$$

In Equation (6), the estimate $\text{Var}(I | \mathcal{D})$ is based only on prior expectation since it depends on sampling points \mathbf{X}_i but not on the sample values Y_i .

Equation (5) can also be rewritten as

$$E(I | \mathcal{D}) = \bar{I}_0 + \mathbf{c}' \mathbf{Y} \quad (10)$$

where

$$\bar{I}_0 = \bar{I} - \mathbf{c}' \bar{\mathbf{F}} \quad (11)$$

$$\mathbf{c} = \mathbf{Q}^{-1} \mathbf{z}. \quad (12)$$

Equation (10) expresses a quadrature rule in which \mathbf{c} is the vector of quadrature coefficients. This form of quadrature is called *Bayesian quadrature* and *Bayes-Hermite quadrature* when $p(\mathbf{x})$ is a Gaussian distribution. In this latter case or when $p(\mathbf{x})$ is uniform, the quadrature coefficients can be computed through a closed-form solution [O'H91].

4.3. Overview of the BMC method

The first problem that we have to solve in implementing the BMC method is how to choose the mean function and the covariance function of the GP model associated with $f(\mathbf{x})$. The first point is addressed in the following section and we have already discussed in the above section the general properties of covariance functions and their impact on the smoothness assumption. The problem of hyperparameters selection is specifically addressed in Section 4.3.3. Given a set of samples \mathcal{D} , we will then need to compute the vector of quadrature coefficients \mathbf{c} with Equation (12) and the I estimate with Equations (11) and (10). The main computing task is represented by Equation (12) which involves the inversion of the \mathbf{Q} matrix and the computation of the \mathbf{z} vector using Equation (8). We will see in the following how to solve this problem for our application. Let us recall that with the BMC method, samples can be drawn from arbitrary distributions but the sampling strategy has a significant effect on the variance as discussed in Sections 4.3.2 and 5.4.

4.3.1. Finding a mean function

The GP prior assumes that we know a mean function $\bar{f}(\mathbf{x})$. This mean function does not need to be an accurate estimate of the statistical mean at any point \mathbf{x} . A rough approximation of $f(\mathbf{x})$ is often sufficient since the generated bias is negligible with the sample sets sizes used in practice. In the context of global illumination, well-known CG techniques can be used to determine such a function (e.g. radiance cache, spherical harmonics, etc.). Depending on how well $\bar{f}(\mathbf{x})$ can model the discontinuities of $f(\mathbf{x})$, the high frequency components of the difference $f(\mathbf{x}) - \bar{f}(\mathbf{x})$ can be reduced, which allows the smoothing kernel of the GP model to more faithfully fit the observed data statistics. Note also that there are some similarities with the approximating function used in the control variates method and techniques such as the 5D tree proposed in [LW95] can be used as well.

Another alternative offered by the GP theory is to infer $\bar{f}(\mathbf{x})$ from the observed data \mathcal{D} as proposed in [RW06]. To this end, we write $\bar{f}(\mathbf{x})$ as:

$$\bar{f}(\mathbf{x}) = \mathbf{h}(\mathbf{x})' \boldsymbol{\beta},$$

where $\mathbf{h}(\mathbf{x})$ is a given set of m basis functions (such as spherical harmonic basis) and $\boldsymbol{\beta}$ are unknown parameters. The derivation of the vector of coefficients $\boldsymbol{\beta}$ can be found in [RW06]. This solution results in an additional variance term but this term decreases rapidly as n increases. For the application described in this paper, we have chosen a constant mean function. We will see that this simple solution gives good results with the sample set sizes we use. However, it is probable that a more refined model of the mean function will be necessary with adaptive techniques allowing smaller sample set sizes (cf. discussion in Section 6.4).

4.3.2. Optimal sampling

One important advantage of BMC over importance sampling is that samples do not need to be drawn from predefined distributions. However, some distributions will be more efficient than others. The space filling designs of quasi-Monte Carlo (QMC) methods [Nie92, HLO04] can be used but their optimality criteria correspond to special cases of Bayesian optimal designs and are mostly based on intuitive considerations. A more appropriate technique consists in choosing the sampling points $\mathbf{X}_1, \dots, \mathbf{X}_n$ to minimize the expected variance $\text{Var}(I | \mathcal{D})$ [O'H91, Min00]. We have used this method in our work (cf. 5.4). Note however that this design will be based on prior expectation only.

4.3.3. Adaptation of hyperparameters

The GP prior defined above assumes that we know the hyperparameters of the model. In the case of the SE covariance function (Equation (3)), the vector of hyperparameters is

$$\vartheta = (\sigma, w_0, w_1, \dots, w_D)^t,$$

where σ^2 is the variance of the additive noise. The choice of these hyperparameters must reflect our beliefs on the variability and the coherence of the function $f(\mathbf{x})$ to be integrated. In [RW06], Rasmussen suggests a maximum likelihood solution that consists in finding ϑ that maximizes the marginal likelihood:

$$\hat{\vartheta} = \underset{\vartheta}{\text{argmax}} [p(\mathcal{D} | \vartheta)].$$

However, this solution is costly because it requires that the hyperparameters be computed at each integral evaluation as the marginal likelihood maximization must be performed independently for each sample set \mathcal{D} . Another solution is to find the hyperparameters that best fit the real covariance function built from a set of training data. This can be done at a global level for the whole scene or a specific view but this choice may be locally suboptimal. We have chosen this last solution in the application presented in this paper. However a simplified version of the maximum likelihood solution will be investigated in future works.

5. Application to Final Gathering

5.1. The irradiance integral

We have chosen final gathering as a case study to show the benefit of BMC because it is typically a case for which basic Monte Carlo integration methods work poorly. As regards importance sampling, the only adequate proposal distribution that can be extracted from the analytical part of the integrand is the cosine distribution which leads to highly scattered samples. This situation does not allow us to faithfully represent the incident radiance. This problem is solved by the BMC approach which uses a GP to model the radiance

variations between sample points, and consequently yields a better estimate of the integral.

Of course, in the Monte Carlo method, more adequate distributions can be built from observed data as in [CAE08]. As for BMC, equivalent techniques (such as active learning) exist but they are difficult to implement in this context.

Briefly, the final gathering is the final step of the photon mapping method. It consists in computing the irradiance of visible surfaces by casting rays from these surfaces so as to capture the distribution of incident light. The irradiance at a given point P of a surface S is given by the integral:

$$E_s(P) = \int_{\Omega_{2\pi}} L(\theta, \phi) \cos \theta \, d\Omega, \quad (13)$$

where θ and ϕ define the direction of a ray of incident light, $d\Omega$ is the elementary solid angle, $L(\theta, \phi)$ is the radiance of the surface intersected by this ray and $\Omega_{2\pi}$ is the solid angle representing the hemisphere centered at P and located above the considered surface. The couple (θ, ϕ) defines the spherical coordinates of a point η on the unit sphere centered at P . The origin of this coordinate system is P and its z axis coincides with the normal N to the surface at point P . By expanding $d\Omega$ in terms of (θ, ϕ) , we obtain:

$$E_s(P) = \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\frac{\pi}{2}} L(\theta, \phi) \cos \theta \sin \theta \, d\theta d\phi. \quad (14)$$

To obtain an integral in the same form as Equation (1), we rewrite Equation (14) as

$$E_s(P) = \pi \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\frac{\pi}{2}} L(\theta, \phi) p(\theta, \phi) \, d\theta d\phi, \quad (15)$$

where

$$p(\theta, \phi) = \frac{\cos \theta \sin \theta}{\pi} = \frac{\sin 2\theta}{2\pi} \quad (16)$$

is a normalized weight function.

5.2. A Gaussian Process model for incident illumination

As discussed in Section 4.3, the first step to implement the BMC method is to define the GP model associated with the 'unknown' part of the integrand of Equation (15), namely the GP mean and covariance functions associated with the incident radiance $L(\theta, \phi)$. The role and the properties of these functions has already been presented in Sections 4.2 and 4.3.1. Our choice of mean function will be discussed in Section 5.7.1 As regards the covariance function, it must be isotropic (all dimensions are equivalent) and positive definite on the unit sphere. Several functions having these properties have been proposed in the literature (e.g. [RW06]). We have obtained good results with the isotropic form of the SE covariance function defined in Equation (3) despite its strong

smoothing effect. It is given by the following equation:

$$k(r) = w_0^2 \exp\left(-\frac{r^2}{2l^2}\right), \quad (17)$$

where r is the Euclidean distance between two points η_1 and η_2 on the unit sphere, and l is the lengthscale hyperparameter. Let us call d the geodesic distance (i.e. the angle between the directions defined by η_1 and η_2). Then, we have

$$d = \arccos(\eta_1 \cdot \eta_2), \quad (18)$$

and

$$r = \sqrt{2 - 2\eta_1 \cdot \eta_2} = 2 \sin \frac{d}{2} \quad 0 \leq d \leq \pi \quad (19)$$

$$r^2 = 2(1 - \cos d).$$

Using Equation (19), the covariance function given by Equation (17) can be expressed as a function of the geodesic distance d :

$$k(d) = w_0^2 \exp\left(\frac{\cos d - 1}{l^2}\right). \quad (20)$$

Equation (18) can be expressed in terms of spherical coordinates as follows:

$$\cos d = \sin \theta_1 \sin \theta_2 \cos(\phi_2 - \phi_1) + \cos \theta_1 \cos \theta_2, \quad (21)$$

where (θ_1, ϕ_1) and (θ_2, ϕ_2) are the spherical coordinates of η_1 and η_2 respectively. The covariance function can then be expressed as a function of the spherical coordinates of η_1 and η_2 as follows:

$$k(\eta_1, \eta_2) = w_0^2 \exp\left(\frac{\sin \theta_1 \sin \theta_2 \cos(\phi_1 - \phi_2) + \cos \theta_1 \cos \theta_2 - 1}{l^2}\right). \quad (22)$$

5.3. Determination of the quadrature coefficients

Equation (22) allows us to build the covariance matrix Q for a set of sampling directions $\mathcal{D} = \{\eta_1, \dots, \eta_n\}$.

To compute the quadrature coefficients given by Equation (10), we also need to compute the coefficients $z_i = f_z(\eta_i)$ of the \mathbf{z} vector for each sample η_i of \mathcal{D} . Using Equations (16), (20) and (21), the expression of the f_z function of Equation (8), becomes

$$f_z(\eta_i) = f_z(\theta_i, \phi_i) = \frac{1}{2\pi} \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\frac{\pi}{2}} k(\eta, \eta_i) \sin 2\theta \, d\theta \, d\phi, \quad (23)$$

where (θ_i, ϕ_i) are the spherical coordinates of a sampling direction η_i . Let us observe that in Equation (21), the variable ϕ_i only appears in the $\cos(\phi - \phi_i)$ term. Since the integration domain covers a whole period of the $\cos(\phi - \phi_i)$ function,

z_i is independent of ϕ_i :

$$f_z(\theta_i, \phi_i) = f_z(\theta_i, 0) = f_z(\theta_i).$$

Moreover, since the dependence of the covariance function as regards ϕ_1 and ϕ_2 is only contained in the term $\cos(\phi_2 - \phi_1)$ (cf. Equation (20) and (21)), the covariance matrix Q is unchanged if the set of sampling directions $\mathcal{D} = \{\eta_1, \dots, \eta_n\}$ is rotated around the z axis. As the z_i coefficients depend only on θ_i , this means that the whole vector of quadrature coefficients $\mathbf{c} = Q^{-1}\mathbf{z}$ is also unchanged. This property will be very useful to generate different sample sets while keeping the same vector of quadrature coefficients.

5.4. Optimal sampling for the diffuse component

The variance function can be computed from (6). In this equation, \bar{V} can be derived from Equations (9) and (23):

$$\bar{V} = \frac{1}{2\pi} \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\frac{\pi}{2}} f_z(\theta) \sin 2\theta \, d\theta \, d\phi = \int_{\theta=0}^{\frac{\pi}{2}} f_z(\theta) \sin 2\theta \, d\theta.$$

As mentioned in Section 4.3.2, optimum sample sets for given values of l , σ , w_0 and n (the size of the sample set) can be computed by finding the samples locations that minimize the variance function of Equation (6). Compared to a uniform distribution generated by the spiral points algorithm of Rakhmanov et al [SK97], optimized sets roughly bring a 6 dB variance reduction (as computed from Equation (6)).

As expected, the minimization process becomes very computationally demanding when n reaches one hundred or more samples. To speed up the computation, we only specify the θ values as input arguments to the variance function. This choice is sensible since the distribution in θ is dominant given the respective roles of θ and ϕ in the irradiance integral. If we pursue this line of reasoning further, we can choose to act on a global θ distribution function rather than on individual θ coordinates of samples. If this distribution function is modeled by a polynomial of degree p , this will reduce the number of input variables to $p + 1$. To generate the samples locations, we use a slightly modified version of the spiral points algorithm:

```
dphi = pi*(3-sqrt(5));
phi = 0.;
dz = 1./n;
z = 1 - dz/2;
for k = 1:n
    zv = polyval(dcoeff, z); % z coord
    thetas(k) = acos(min(zv,1)); % Theta
    phis(k) = mod(phi,2*pi); % Phi
    z = z - dz;
    phi = phi + dphi;
end
```

In the standard spiral points algorithm, the z coordinate of points are uniformly distributed. In our algorithm, this linear distribution is modified by the polynomial function `polyval()`. The input variables for the optimizer are then

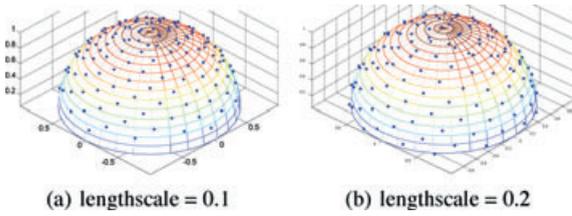


Figure 2: Optimized sample sets of 128 points.

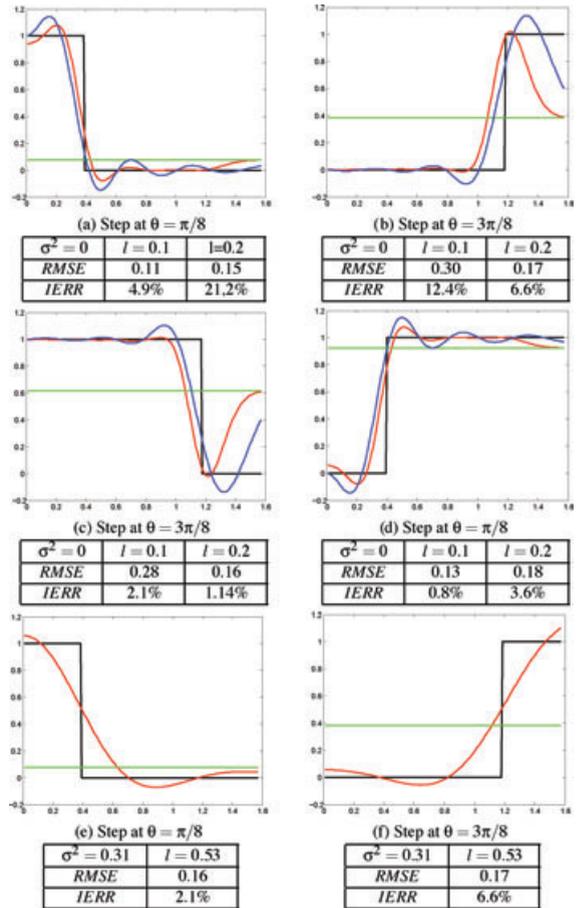
the coefficients $dcoeff$ of the polynomial. The algorithm is initialized with the uniform distribution, i.e. the monomial $y = x$. A degree 4 or 5 is sufficient in practice. We have used the Quasi-Newton line search algorithm as optimizer. The loss in variance reduction is at worst 1.8 dB compared to the direct method while the speed of computation is one hundred times faster for large sample sets. Examples of optimized sample sets are shown in Figure 2. Observe that for $l = 0.1$ (Figure 2(a)), there are fewer samples at grazing angles since with this short lengthscale, sample points need to be closer to each other and the optimizer tends to trade off sampling directions with low weight (i.e. defined by $cos(\theta)$ weight function) for high weight sampling directions. This behavior is similar to importance sampling, however, in optimized sample sets, the effect of the cosine weight function tends to be counterbalanced by the covariance function as the lengthscale increases as shown in Figure 2(b).

5.5. Behavior in case of radiance function discontinuities

As discussed in Section 4.2, the smoothness assumption for the integrand is modeled by the covariance function defined by Equation (2) and its strength is characterized by the lengthscale hyperparameter. Besides, as the power spectrum of a signal is obtained by taking the Fourier transform of its covariance function, we can also interpret the effect of lengthscale parameter in the frequency domain. Fixing a lengthscale value amounts to assuming that the bandwidth of the difference function $L(\cdot) - \bar{L}(\cdot)$ is limited. Recall however that this interpretation is valid only if the covariance function is stationary.

In this section, we analyze the behavior of the BMC estimator of the irradiance integral when the radiance function is highly discontinuous. For this purpose, we have computed the BMC estimate of the irradiance integral of Equation (13) when the radiance function is a step function in θ ($L(\theta, \phi) = s(\theta), \forall \phi$) and taking a constant mean function $\bar{L} = \frac{1}{2\pi} \int_{\Omega_{2\pi}} L(\theta, \phi) d\Omega$. We have also computed the predicted radiance function according to the regression formula of Equation (4).

The results presented in Figure 3 (a) to (d) use the two sample sets shown in Figure 2. We can observe that the



Black line Original incident radiance $L(\theta, 0)$
Green line Mean value of incident radiance
Red line Radiance prediction with $l = 0.1$ for figures (a) to (d) and $l = 0.53$ for figures (e) and (f)
Blue line Radiance prediction with $l = 0.2$
RMSE Root mean squared error of the radiance prediction
IERR Relative error of the integral estimate
 The integral value is 0.1464 in cases (a), (b), (e), (f) and 0.8536 in cases (c) and (d).

Figure 3: Behavior in case of a step radiance function.

$RMSE$ of the radiance prediction is generally high whereas the error on the integral estimate remains quite low except for some extreme cases. Moreover, when comparing case (a) and (c) for $l = 0.1$, we can see that the prediction $RMSE$ is much greater for case (c) than for case (d) whereas the order is reversed as regards $IERR$. This behavior is due to the particular sample points distribution of optimized sets. As discussed in Section 5.4, the optimizing process tends to gradually allocate more points near the pole than near the equator and this is particularly well visible in Figure 2 when $l = 0.1$. Therefore, the prediction error is high at grazing angles but this has little effect on the integral estimate because

of the weight function contribution. In cases (b) and (c), observe that when θ is close to $\pi/2$, the predicted radiance approaches the mean value because the sample points are too sparse and the effect of the covariance function becomes negligible in Equation (4).

Cases (a) and (d) are the most critical ones because the step occurs near the pole where the weight function has high values. Let us note that the longest lengthscale ($l = 0.2$) corresponds roughly to a geodesic distance of $\pi/8$, which is also the length of the step function of cases (a) and (d). In these cases, the $l = 0.1$ optimized design provides much better estimate because the smoothness assumption is less strong (i.e. the assumed signal bandwidth is higher) and also because the optimizer has allocated more sample points near the pole as explained in Section 5.4. Let us note also that in both these cases, the choice of l has a much stronger impact on *IERR* than on *RMSE*. In case (b) with $l = 0.1$, the sample points are too sparse at grazing angles to provide a good estimate.

Cases (e) and (f) show the results obtained with the optimal hyperparameter values (and associated optimal sample set) used for the rendering of the Sibernik Cathedral shown in Figure 9. The relatively high value of the optimal σ hyperparameter (i.e. the standard deviation of the additive noise ε) has the effect of smoothing out high frequency components as explained in [RW06], which reduces the ringings. The estimate of the irradiance integral remains good despite this strong smoothing effect and is even improved compared to case (a). As integration is inherently a low-pass filtering operation, smoothing out high frequency components does not affect too much the integral estimate. Another interpretation of this hyperparameter settings is to consider that the high frequency components of the original signal are included in the additive noise ε .

All these observations clearly show the importance of optimizing the design (sample points location, hyperparameters values and mean function) according to the integral estimate and not the radiance function prediction.

5.6. Discussion on implementation strategies

Basically, two implementation strategies are possible. The first one consists in adapting the sampling directions and the hyperparameters values at each irradiance integral evaluation so as to obtain the least possible variance and use the least possible number of samples. This solution is costly because it will require computing the quadrature coefficients at each point P , which is $\mathcal{O}(n^3)$ in complexity due to the inversion of the Q matrix.

Another possible strategy is to precompute sample sets that closely minimize the expected variance given by Equation (6). The quadrature coefficients associated with these sets can then be precomputed for the whole scene. In this case,

a set of globally optimal hyperparameters must be selected at the scene level. This solution will be locally sub-optimal in terms of number of samples and hyperparameters values but it will avoid computing the quadrature coefficients at each integral evaluation. Another drawback of this solution is that it does not allow to take advantage of the freedom in samples distribution (recall that in Bayesian Monte Carlo, samples do not need to be drawn from a predefined distribution).

A combination of the two methods is possible. For example, we could take a precomputed optimal distribution and then supplement it with additional samples in direction of light sources. The obtained BMC estimate remains unbiased as BMC is independent of sample distribution. In the implementation proposed in this paper, we have only used precomputed sample sets, other alternatives will be studied in the future.

5.7. Implementation details

Using Equations (5) and (15), the BMC estimate \hat{E}_{BMC} of the irradiance value E at each visible point P can be written as:

$$\hat{E}_{BMC} = \bar{E} + \pi \mathbf{z}' Q^{-1}(\mathbf{Y} - \bar{\mathbf{F}}) \quad (24)$$

where

$$\begin{aligned} \bar{E} &= \pi \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\frac{\pi}{2}} \bar{f}(\eta) p(\eta) d\theta d\phi \\ \eta &= (\theta, \phi) \\ p(\eta) &= \frac{\sin(2\theta)}{2\pi}. \end{aligned}$$

For a given set of directions over the hemisphere $\mathcal{D} = \{(\eta_i, Y_i) \mid i = 1, \dots, n\}$, Y_i is the incoming radiance from direction η_i . This radiance is computed using a query to a photon map. \bar{E} can be seen as an a priori estimate of the irradiance, deduced from the prior mean \bar{f} .

We have shown that BMC integration depends on the GP model of the radiance function $L(\theta, \phi)$. In order to perform BMC integration, we need to determine the mean function \bar{f} and the hyperparameters values of the GP model w_0 , l and σ .

5.7.1. Determination of the mean function

As explained in Section 4.3.1, BMC allows us to infer \bar{f} from the observed data \mathcal{D} . The choice of \bar{f} impacts the variance of BMC estimate, but the effect of this additional term decreases rapidly as n increases and/or the value of l increases. When n is small or when the chosen value of l is not coherent with the data, the choice of \bar{f} will have a large impact on \hat{E}_{BMC} . Ultimately, when $l = 0$, \mathbf{z} is the null vector and Equation (24) gives

$$\hat{E}_{BMC} = \bar{E} = \pi \int_{\Omega} \bar{f}(\eta) p(\eta) d\eta,$$

which can be biased. As the determination of the hyperparameters may be costly and/or inaccurate, we want to choose \bar{f} such that in extreme cases, the BMC estimate \hat{E}_{BMC} is the same as the classical Monte Carlo (MC) estimate \hat{E}_{MC} (computed from Equation (13)):

$$\hat{E}_{MC} = \frac{2\pi}{n} \sum f(\eta_i) \cos(\theta_i).$$

Note that when \bar{f} is constant, $\bar{E} = \pi \bar{f}$. In order to have a simple worst-case unbiased BMC estimator, we take \bar{f} as a constant function such that $\bar{f} = \hat{E}_{MC}/\pi$. When $l = 0$, we have

$$\hat{E}_{BMC} = \bar{E} = \pi \bar{f} = \hat{E}_{MC}$$

Hence, in extremes cases, our estimate is equivalent to the classical Monte Carlo estimate. Note that the previous equations hold only for uniformly distributed samples. When using cosine importance sampling, the equation used to compute \hat{E}_{MC} is the corresponding Monte Carlo estimator:

$$\hat{E}_{MC} = \frac{\pi}{n} \sum f(\eta_i).$$

5.7.2. Determination of the hyperparameters

We chose the covariance function $k(\mathbf{x}, \mathbf{x}')$ as the square exponential function, parameterized by l and w_0 . However, the quadrature coefficients $\mathbf{c} = Q^{-1}\mathbf{z}$ do not depend on the value of w_0^2 directly but on the ratio σ^2/w_0^2 . Therefore, the result of BMC integration depends only on the values of the hyperparameters l and $\sigma^2 = \sigma^2/w_0^2$. In the following, we will consider that $w_0^2 = 1$ and $\sigma^2 = \sigma^2$.

In order to determine the impact of the choice of l and σ^2 on the BMC estimate, we have studied the behavior of BMC integration for a given point P in a classical diffuse Cornell Box scene. We computed the irradiance E at point P using MC and BMC integration. We performed 2000 integrations for each method and computed the variance of both estimators. Figure 4 shows the gain in quality obtained (in terms of variance reduction) with BMC integration compared to MC integration. Because of the choice we made for \bar{f} , when $l = 0$ the variance of the BMC estimate is the same as that of the MC estimate. We can see that the optimal values for the hyperparameters are $l \approx 0.2$ and $\sigma^2 \approx 0.3$. Note that even with significant deviations from these optimum values, BMC performs better than MC. We did the same study for several visible points in the scene, and found that BMC integration outperforms MC integration. Of course, the optimal values obtained for l and σ^2 were different at each point.

As explained in Section 4.3.3, finding the optimal hyperparameters that maximize the marginal likelihood is costly. We prefer determining the hyperparameters that best fit the covariance function estimated at P from training data because this solution can be extended at a scene level as explained below. Figure 5 shows $\tilde{k}(d)$, the computed covariance func-

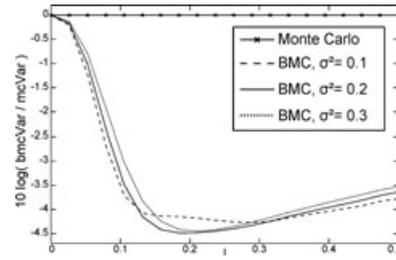


Figure 4: Gain in quality achieved by BMC for different values of l and σ^2 , at a fixed point P within the Cornell Box scene ($n = 256$).

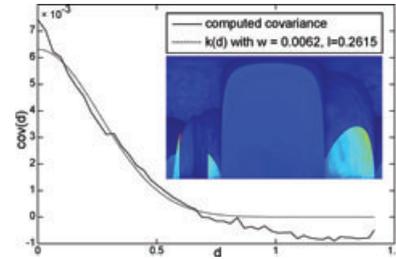


Figure 5: Covariance curve fitting and corresponding incoming radiance function in (ϕ, θ) coordinate space.

tion of the incoming radiance at point P . Using least-square fitting applied to the model given by Equation (20), we obtain $w_0^2 = 6.2 \cdot 10^{-3}$ and $l = 0.2615$. Since σ^2 is related to the divergence between the model and the actual data, we propose

$$\sigma^2 \approx \frac{\int |(\tilde{k}(x) - k(x))| dx}{\int k(x) dx},$$

$k(x)$ being the model given by Equation (20), which yields $\sigma^2 = 0.31$ for the given example. Note that these values are close to the optimal values deduced from Figure 4.

In order to evaluate the covariance function $\tilde{k}(d)$ we cast N_c pairs of rays $(\eta_{1,i}, \eta_{2,i})$ from P , for each value of d . For a given d , $\tilde{k}(d)$ is given by:

$$\begin{aligned} \tilde{k}(d) &= E[(L(P, \eta_1) - \bar{L})(L(P, \eta_2) - \bar{L})] \\ &\approx \frac{1}{N_c} \sum_{i=1}^{N_c} (L(P, \eta_{1,i}) - \bar{L})(L(P, \eta_{2,i}) - \bar{L}), \end{aligned} \quad (25)$$

where

$$d = \arccos(\eta_{1,i} \cdot \eta_{2,i}) \quad \forall i \in [1, N_c].$$

$L(P, \eta_{1,i})$ is the radiance incoming at P from the direction $\eta_{1,i}$ and \bar{L} the mean of all $L(P, \eta_{j,i})$. To compute $\tilde{k}(d)$ associated with the incoming radiance at point P (both showed in Figure 5), we used a total number of 262k rays. This number of additional rays is obviously too high if this computation had to be performed for each new point P .

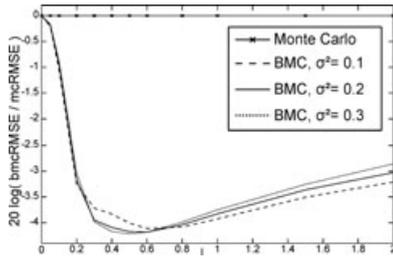


Figure 6: Gain in quality achieved by BMC with different values of \bar{l} and $\bar{\sigma}^2$, for the rendering of a Cornell Box view ($n = 64$).

To overcome this problem, we propose to determine a global value of l and σ^2 for the whole rendered image and use these global values \bar{l} and $\bar{\sigma}^2$ in each final gathering BMC integration. Based on Equation (25), we compute a global covariance function $\bar{k}(d)$:

$$\bar{k}(d) \approx \frac{1}{N} \sum_{i=1}^{N_c} (L(P_i, \eta_{1,i}) - \bar{L})(L(P_i, \eta_{2,i}) - \bar{L}),$$

where P_i is a visible point. For each value of d , we trace only one pair of rays $(\eta_{1,i}, \eta_{2,i})$ from a visible point P_i . After having fit the obtained $\bar{k}(d)$ values to Equation (20), we get \bar{l} and $\bar{\sigma}^2$. For a given view of the Cornell Box scene, we obtain $\bar{l} = 0.45$ and $\bar{\sigma}^2 = 0.22$. Figure 6 shows the quality improvement (BMC RMSE over MC RMSE) obtained for this scene with different values of \bar{l} and $\bar{\sigma}^2$ (RMSE has been computed with respect to a reference rendering using 16k final gathering rays per visible point). The values obtained with the fitting process are close to the optimal ones that we can deduce from Figure 6, say $\bar{l}_{fig} = 0.5$ and $\bar{\sigma}_{fig}^2 = 0.3$. While these global values may be locally sub-optimal for a particular visible point, they prove to be quite good estimates. We used a total of 262k rays to estimate $\bar{k}(d)$ with $d \in [0, \pi/4]$: the cost of evaluating our hyperparameters is equivalent to the cost of one additional gathering ray per visible point.

5.7.3. Making BMC rendering practical

During rendering, we want to compute the BMC estimate given by Equation (24) for each visible point of the scene. As shown in Section 5.7.1, evaluating \bar{E} and \bar{f} is straightforward. We still have to compute the values of \mathbf{z} and Q^{-1} . Naively evaluating the integral for \mathbf{z} and inverting Q for each visible point is not practical. In Section 5 we showed that the values of \mathbf{z} and Q depends only on the relative position of the sampling directions η_i within a given set of directions, \mathcal{D} . Thus, we propose to precompute the whole vector of quadrature coefficients $\mathbf{c} = Q^{-1}\mathbf{z}$ for a fixed number of sets \mathcal{D}_j , then use only these sets at rendering time. Since a set of sampling directions can be rotated around the normal without changing the quadrature coefficients (cf. 5.2), we have found that a small number N_d of different sets \mathcal{D}_j is sufficient to

reduce the artefacts due to repeated sampling patterns. In addition, N_d can be reduced as the number of samples n per set increases.

We have shown in Section 5.2 that the values of $f_z(\eta_i)$ depend only on θ_i and l . In addition, θ_i always ranges between 0 and $\pi/2$, and the typical values of \bar{l} are between 0 and 1. We chose to precompute $f_z(\theta_i, l)$ for a set of values of θ_i and l and store it into a table. At runtime, $f_z(\eta_i)$ is interpolated from the closest entry values in the table. Moreover, due to the properties of the covariance function $k(\mathbf{x}, \mathbf{x}')$, the matrix Q is symmetric and positive-definite. We can use Cholesky factorization to speed-up the matrix inversion process. The resulting algorithm is given by Algorithm 1. Note that the

Algorithm 1: Final gathering algorithm using Bayesian Monte Carlo.

```

// Determination of the
  hyperparameters
Compute  $\bar{k}(d)$  by casting pairs of rays;
Compute  $\bar{l}$  and  $\bar{\sigma}^2$  by fitting;
// Get  $N_d$  sets of directions  $\mathcal{D}_j$  and
  the associated  $\mathbf{c}_j$ 
 $\{(\mathcal{D}_j), (\mathbf{c}_j)\} = \text{GetSets}(N_d, n, \bar{l}, \bar{\sigma}^2)$ ;
// at rendering time, F.G. pass
foreach visible point  $p$  do
  Pick a set of directions  $\mathcal{D}_j$  and associated  $\mathbf{c}_j$ ;
  Rotate  $\mathcal{D}_j$  along the normal at  $p$  by a random  $\phi$ ;
  Cast rays according to  $\mathcal{D}_j$  and compute  $\mathbf{Y}$ ;
  Compute  $\bar{f}$  and  $\bar{E}_{MC}$ ;
  // BMC estimate of the irradiance
  at  $p$ 
   $\bar{E}_{BMC} = \bar{E}_{MC} + \pi \mathbf{c}_j(\mathbf{Y} - \bar{f})$ ;
end

```

function GetSets(N_d, n, l, σ^2)

```

// Precompute  $N_d$  sets  $\mathcal{D}_j$  of  $n$ 
  directions  $\eta_i$  and their associated
  quadrature coefficient  $\mathbf{c}_j$ 
if Optimal Sampling then
  Get optimal  $\{(\mathcal{D}_j), (\mathbf{c}_j)\}$  from a precomputed
  database of optimal sets;
else
  for  $j = 0$  to  $N_d-1$  do
     $\mathcal{D}_j =$  set of  $n$  random directions over the
    hemisphere;
    // compute  $\mathbf{z}_j$ 
    foreach direction  $\eta_i$  do
      Interpolate  $f_z(\eta_i)$  from the table;
    end
    Compute  $Q_j$ ;
    Invert  $Q_j$  using Cholesky factorization;
     $\mathbf{c}_j = Q_j^{-1}\mathbf{z}_j$ ;
  end
end

```

sets of directions \mathcal{D}_j and their associated quadrature coefficient vector \mathbf{c}_j can be computed on-the-fly before rendering, or read from a scene-independent database.

5.7.4. Optimal sampling

Basically, BMC integration schemes do not rely on random samples. In our application to final gathering, instead of generating \mathcal{D}_j using random directions, we could search for a set of directions which minimize Equation (7) as explained in Section 5.4. Although the optimization process is computationally expensive, the generated optimal sets \mathcal{D}_j depend only on the values of N , \bar{l} and $\bar{\sigma}^2$, but they do not depend on the scene. One could precompute a scene-independent database of optimal sets \mathcal{D}_j for some values of n , \bar{l} and $\bar{\sigma}^2$, storing directions and associated optimal \mathbf{c}_j . Then, during the rendering process, we only have to pick some of the optimal sets, trace corresponding rays and compute a dot product between \mathbf{c}_j and \mathbf{Y} .

6. Results

All results were gathered on a Intel Core Duo T2600 (2.16GHz) with 2GB RAM.

6.1. General observations

During the final gathering pass, our algorithm does not have to draw random numbers to generate directions on the hemisphere. Instead, we pick a previously computed set \mathcal{D}_j of directions and retrieve the associated quadrature coefficient vector \mathbf{c}_j . Then, the only additional computation with respect to MC integration is a dot product between \mathbf{c}_j and \mathbf{Y} . During the final gathering pass, the computation time of our algorithm is essentially the same as that of classical Monte Carlo (the computation time is dominated by ray casting and photon map queries). So, the overhead of our algorithm lies in the evaluation of the hyperparameters and the computation of the sets \mathcal{D}_j together with the associated quadrature coefficient vectors.

We found that we can get a good approximation of the actual covariance function of the incoming radiance using a total number of $N_c = 262k$ rays. In our test scenes, the determination of \bar{l} and $\bar{\sigma}^2$ takes 1.1 s for the Cornell Box scene, 5 s for the Sibenik Cathedral scene and 7s for the Sponza Lucy scene.

On-the-fly generation of sets of directions and the associated quadrature coefficients depends mainly on n , the number of rays shot from each visible point. The cost of this generation is dominated by the inversion of the $n \times n$ matrix Q . We found that we do not need a large number N_d of sets \mathcal{D}_j , to avoid artefacts. Moreover, the larger the number of samples n , the smaller N_d . Computing all the \mathbf{c}_j takes 273 ms when

$n = 16$ and $N_d = 1000$, 1.16 s when $n = 64$ and $N_d = 250$, and 9.2 s when $n = 256$ and $N_d = 60$.

6.2. Bayesian Monte Carlo integration using random sampling

As Bayesian Monte Carlo does not make any assumption on the sample directions η_j , we can make use of importance sampling and/or stratified sampling to generate sample direction sets \mathcal{D}_j . In this way, we will be able to compare BMC with MC for the same sampling directions \mathcal{D}_j generated either with uniform sampling or importance sampling.

Figure 7 compares the variation of RMSE with respect to a reference rendering, as the number of samples increases. We can see that for the exact same \mathcal{D}_j , BMC greatly reduces the noise in the rendered image (Cornell Box, Figure 1 and 7(a)). We can notice that BMC (with importance sampling) outperforms MC (with importance sampling) (Sibenik Cathedral, Figure 7(b)). Table 1 gives some results on timing and image quality obtained with BMC and MC from the same sets of sample directions \mathcal{D}_j .

Figure 8 shows the rendering of a scene containing glossy objects. Our current implementation of BMC computes final gathering only on diffuse surfaces. The glossy reflections were computed using Monte Carlo with importance sampling. When comparing RMSE values of different methods, we only consider the pixels computed with BMC. Figure 8(e) and (f) show the difference images. We can see from these difference images that the error is generally smaller with BMC especially for bright surfaces but this is hardly visible on the rendered picture. This is due to the fact that, thanks to the Gaussian Process prior of BMC, a few samples are necessary to evaluate the contribution of a bright surface seen from a relatively narrow angle whereas much more samples are necessary with importance sampling.

6.3. BMC estimator using optimal sets of directions

BMC has the advantage (over MC) of assigning small weights to nearby samples and larger ones to relatively distant samples. As for SI-MC (Monte Carlo with stratified importance sampling), nearby samples are rare, only relatively distant samples are generated. This is why SI-BMC (BMC with stratified importance sampling) does not perform better than SI-MC, but provides results that are similar to those of SI-MC in the worst case. However, with O-BMC (BMC with optimal sample sets \mathcal{D}_j), the obtained results are better as the method significantly reduces noise in the rendered image. Note that if we want SI-MC to achieve the same RMSE as the ones of O-BMC, more than 33% additional sample rays are needed (we save 25% rays with respect to SI-MC, Table 1).

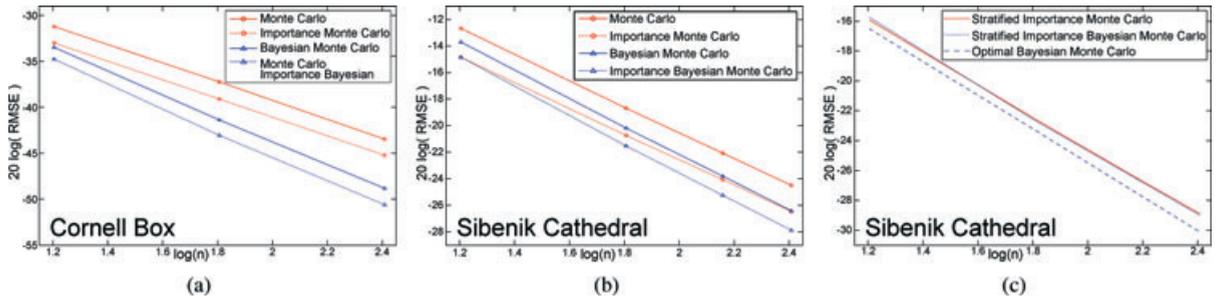


Figure 7: RMSE variation as a function of the number of rays. (a) Cornell Box scene. (b) and (c) Sibenik Cathedral scene. Compared to SI-MC, SI-BMC gives similar results. However, Optimal BMC significantly improves the rendering quality.

Table 1: Quality comparison between BMC renderings and the associated MC renderings (using the exact same sets \mathcal{D}_j). Total time includes determination of hyperparameters (few seconds, depending on the scene) and on-the-fly computation of \mathcal{D}_i and associated \mathbf{c}_j (9.2 sec when $n = 256$). The overhead of our algorithm is negligible with respect to the time spent on ray casting. Same quality MC #rays is the number of rays needed to compute a rendering yielding the same RMSE value as the corresponding BMC rendering.

Scene	Sampling method	# FG Rays	RMSE (10^{-3})		RMSE gain	Same quality MC #Rays	% Rays saved	Total time (in seconds)
			MC	BMC				
Cornell Box	Uniform sampling	256	6.71	3.63	5.34 dB	900	72%	211
	Importance sampling	256	5.49	2.95	5.40 dB	896	71%	
Cathedral	Uniform sampling	256	59.6	47.8	1.92 dB	402	36%	1310
	Importance sampling	256	47.5	40.4	1.41 dB	362	29%	
	Stratified + importance	256	35.9	35.5	0.1 dB	266	4%	1310
	Optimized sampling	/	/	31.2	1.22 dB	342	25%	1301
Sponza Lucy	Importance sampling	256	9.1	8.03	0.99 dB	445	49%	1648
	Stratified + importance	256	7.67	7.61	0.07 dB	271	6%	1648
	Optimized sampling	/	/	7.35	0.37 dB	332	23%	1639

In Figure 9, the image quality is the same for SI-MC and O-BMC in some parts of the image (the global \bar{l} and $\bar{\sigma}^2$ are far from optimal) and better for O-BMC in some other parts (the global \bar{l} and $\bar{\sigma}^2$ are closer to optimal). This means that finding a fast method of computing local estimates of l and σ^2 would further improve the performances of O-BMC.

6.4. Discussion

The results that we have obtained are surprisingly good given the strong smoothness assumption that we have put in the Gaussian Process prior. First, we have used a very smooth covariance function as explained in Section 4.2, second we have taken a constant mean function that does not allow to reduce radiance function discontinuities as discussed in Section 4.3.1 and third, the model hyperparameters are selected at a global level, which leads to quite long lengthscale values, (i.e. narrow bandwidth assumption for the radiance function). These assumptions may seem incompatible with

the sharp discontinuities met in the test scenes. However, if we consider the results presented in Figure 3, the integral estimate error remains below 7% with the longest length-scale ($l = 0.53$). Furthermore, Figure 4 and 6 show that the variance of BMC estimates increases slowly with l when l is greater than the optimal value. These observations explain why optimizing hyperparameters at the scene level leads to rather long lengthscales since the error on the integrals estimates will remain low despite significant deviations from the optimal values.

Figure 10 shows the strong smoothing effect brought by the Bayesian prediction with the globally optimal hyperparameters, which strengthens the conclusions drawn from the simulations of Section 5.5. Hyperparameters must be selected so as to obtain the best integral estimates and not the best data fittings. In this regard, our hyperparameters determination method leads to values that are close to optimum as explained in Section 5.7.2 However, the simulation results presented in Section 5.5 suggest that better estimates can be expected if the hyperparameters could be locally adapted.

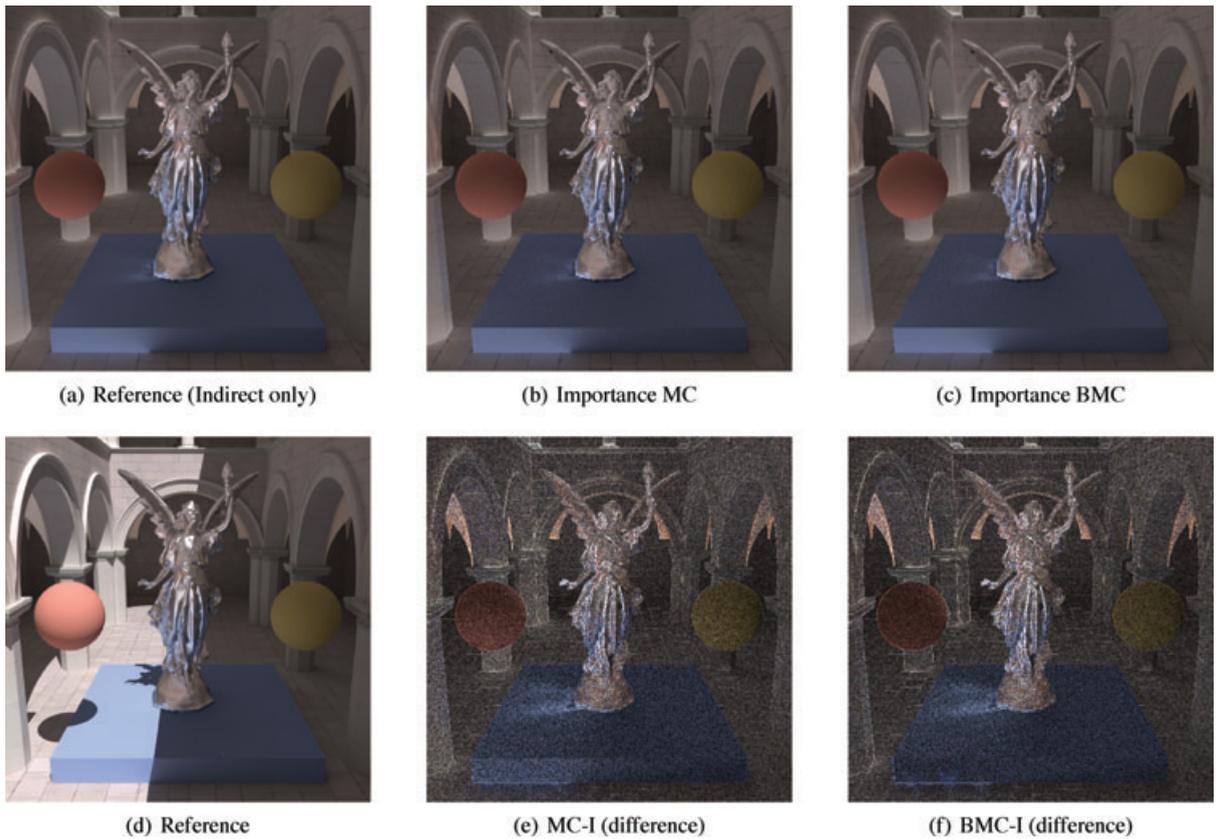


Figure 8: Rendering of Sponza Atrium scene, with a glossy Lucy. $n = 256$, $\bar{l} = 0.42$, $\bar{\sigma}^2 = 0.31$. Glossy reflections have been computed using importance sampling. Difference pictures have been multiplied by 10. To obtain the same quality, Monte Carlo integration would require $n = 445$.

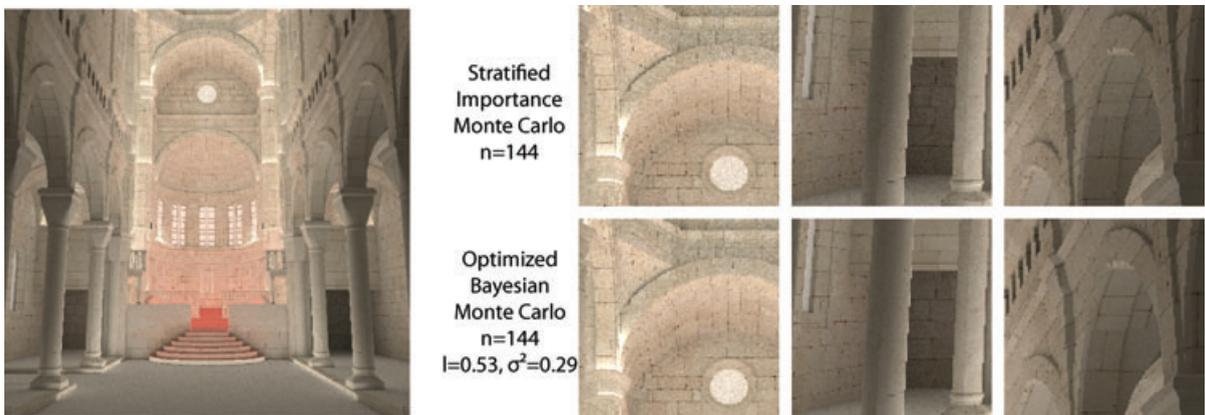


Figure 9: Optimal BMC rendering of the Sibenik Cathedral (Indirect only). $n = 144$, $\bar{l} = 0.53$, $\bar{\sigma}^2 = 0.29$. Locally, OBMC performance depends on how close are the global values of \bar{l} and $\bar{\sigma}^2$ to the local optimal values. Note that O-BMC never performs worse than SI-MC, hence the global quality of the rendering is improved. The overall reduction of the RMSE value is 11%. The reduction of the RMSE value for each of the detail view is respectively 8%, 22% and 16%.

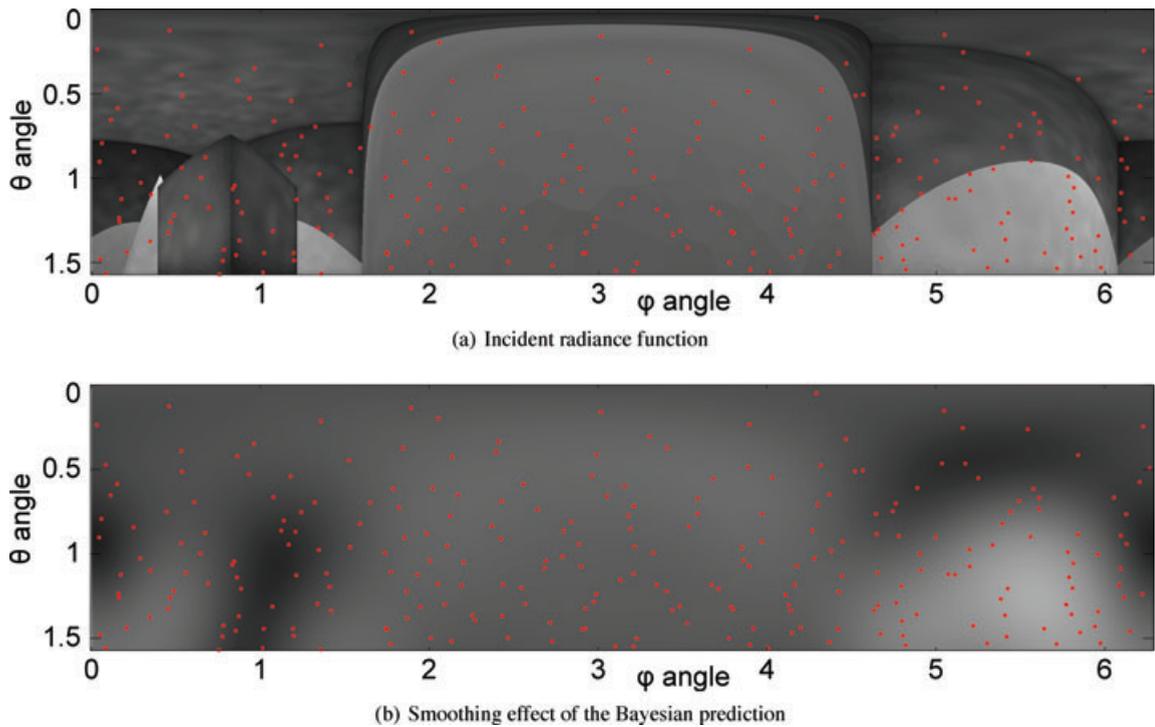


Figure 10: Bayesian prediction of incident radiance seen from a given point in the Sibenik Cathedral scene. The red dots are the sample points.

7. Conclusion

This work clearly shows the benefits of BMC over traditional MC methods. Similarly to deterministic quadrature methods, BMC takes into account sample positions but remains a probabilistic method in that it assumes a Gaussian prior on the function to be integrated. Both these features contribute to the strength of this method. In the BMC implementation described in this paper, we have chosen to leave unchanged the computing load of the rendering loop, i.e. estimates are computed by simple weighted sums just as in standard MC. We have shown that this method can lead to significant savings in terms of number of rays. However, as already mentioned earlier, this implementation does not fully exploit the advantages of the BMC method. Ideally, the prior model should be adapted to the data for each integral computation, which is too costly with the usual likelihood minimization methods. Our future research will thus be directed toward new methods allowing local adaptation of the hyperparameters values as well as the mean function and the number of sample points. The results we have obtained with this first implementation lead us to think that there is a great margin for improvement in this direction. We are also planning to extend our BMC approach to the rendering of glossy surfaces. Similarly to the diffuse reflection case, sets of sampling directions are generated and associated weight coefficients are computed.

The difference with diffuse reflection lies in the fact that these directions are the half-angle directions. Given a viewing direction and the half-angle sampling directions together with their associated weight coefficients, we determine the incident directions and the quadrature coefficients.

References

- [Bir62] BIRNBAUM A.: On the foundations of statistical inference. *Journal of the American Statistical Association* 57, 298 (1962), 269–236.
- [CAE08] CLINE D., ADAMS D., EGBERT P. K.: Table-driven adaptive importance sampling. *Computer Graphics Forum* 27, 4 (2008), 1115–1123.
- [DLAW01] DROR R., LEUNG T., ADELSON E., WILLSKY A.: Statistics of real-world illumination. In *CVPR'01: Proc. IEEE Conference on Computer Vision and Pattern Recognition* (2001), vol. 2, pp. 164–171.
- [FCH*06] FAN S., CHENNEY S., HU B., TSUI K.-W., CHI LAI Y.: Optimizing control variate estimators for rendering. *Computer Graphics Forum* 25, 3 (2006), 351–357.

- [Her04] HERTZMANN A.: Introduction to Bayesian learning, 2004. <http://www.dgp.toronto.edu/hertzman/ibl2004/notes.pdf>.
- [HLO04] HICKERNELL F. J., LEMIEUX C., OWEN A. B.: Control variates for quasi-Monte Carlo. *Statistical Science* 20 (2004).
- [KNKS08] KUMAR A., NAIR P. B., KEANE A. J., SHAHPAR S.: Robust design using Bayesian Monte Carlo. *International Journal for Numerical Methods in Engineering* 73, 11 (2008), 1497–1517.
- [LW94] LAFORTUNE E., WILLEMS Y. D.: The ambient term as a variance reducing technique for Monte Carlo ray tracing. In *Proceedings of the Fifth Eurographics Workshop on Rendering* (1994), Springer-Verlag, Berlin, pp. 163–171.
- [LW95] LAFORTUNE E., WILLEMS Y.: A 5D tree to reduce the variance of Monte Carlo ray tracing. In *Rendering Techniques'95 (Proceedings of the Sixth Eurographics Workshop on Rendering)* (1995), Springer-Verlag, Berlin, pp. 11–20.
- [Min00] MINKA T. P.: *Deriving quadrature rules from Gaussian processes*. Tech. Rep., Statistics Department, Carnegie Mellon, 2000.
- [Nie92] NIEDERREITER H.: *Random Number Generation and Quasi-Monte Carlo Methods*. SIAM, 1992.
- [O'H87] O'HAGAN A.: Monte-Carlo is fundamentally unsound. *The Statistician* 36, 2–3 (1987), 247–249.
- [O'H91] O'HAGAN A.: Bayes-Hermite quadrature. *Journal of the Statistical Planning and Inference* 29, 3 (1991), 245–260.
- [OZ00] OWEN A., ZHOU Y.: Safe and effective importance sampling. *Journal of the American Statistical Association* 95, 449 (March 2000), 135–143.
- [PHR06] PFINGSTEN T., HERRMANN D. J. L., RASMUSSEN C. E.: Model-based design analysis and yield optimization. *IEEE Transactions on Semiconductor Manufacturing* 19, 4 (2006), 475–486.
- [RG02] RASMUSSEN C. E., GHAHRAMANI Z.: Bayesian Monte Carlo. In *Neural Information Processing Systems* (2002), MIT Press, pp. 489–496.
- [RW06] RASMUSSEN C. E., WILLIAMS C. K. I.: *Gaussian Process for Machine Learning*. MIT Press, Cambridge, MA, USA, 2006.
- [SK97] SAFF E., KUIJLAARS A.: Distributing many points on a sphere. *Mathematical Intelligencer* 19, 1 (1997), 5–11.
- [Vea97] VEACH E.: *Robust Monte Carlo Methods for Light Transport Simulation*. PhD thesis, Stanford University, 1997.