A Spherical Gaussian Framework for Bayesian Monte Carlo Rendering of Glossy Surfaces

R. Marques,[1] C. Bouville,[2] M. Ribardière,[2] L.P. Santos,[3] K. Bouatouch[2]

[1]INRIA, Rennes, France,

[2]IRISA, Rennes, France,

[3]University of Minho, Braga, Portugal

Abstract—The Monte Carlo method has proved to be very powerful to cope with global illumination problems but it remains costly in terms of sampling operations. In various applications, previous work has shown that Bayesian Monte Carlo can significantly outperform importance sampling Monte Carlo thanks to a more effective use of the prior knowledge and of the information brought by the samples set. These good results have been confirmed in the context of global illumination but strictly limited to the perfect diffuse case. Our main goal in this paper is to propose a more general Bayesian Monte Carlo solution that allows dealing with non-diffuse BRDFs thanks to a spherical Gaussian-based framework. We also propose a fast hyperparameters determination method which avoids learning the hyperparameters for each BRDF. These contributions represent two major steps towards generalizing Bayesian Monte Carlo for global illumination rendering. We show that we achieve substantial quality improvements over importance sampling at comparable computational cost.

Index Terms—Bayesian Monte Carlo, Gaussian Process, Spherical Gaussian.

✦

1 INTRODUCTION

In global illumination rendering, the main issue lies in the computation of multi-dimensional integrals involving intensive ray-traced sampling. Although the Monte Carlo method has proved to be very powerful when coping with this problem, it remains costly in terms of sampling operations. As the computational cost of sampling is very high compared to the sheer quadrature cost, we can ask ourselves if Monte Carlo methods make an efficient use of information brought by the samples. This question has been raised by O'Hagan which led him to propose the "Bayes-Hermite quadrature" [1], a new form of quadrature which is referred to as "Bayesian Monte Carlo" (BMC) by other authors. While keeping the fundamental property of data dimension independence of Monte Carlo methods, it considerably broadens the set of theoretical tools that can be used to exploit the information produced by sampling. In particular, BMC uses the information regarding the samples location, which is ignored in the classic Monte Carlo method (CMC). Moreover, BMC offers much more flexibility in the exploitation of the prior knowledge compared to CMC which mainly relies on sampling strategies. However, all these advantages are obtained at the expense of the quadrature complexity and additional preprocessing. Brouillat et al. [2] have proposed efficient solutions that make the overhead of computing the BMC quadrature negligible compared to CMC. Moreover, their results show that BMC can significantly outperform CMC methods, even when including the preprocessing step. Nevertheless, their work only considers diffuse reflections and their strategy for efficiently computing the quadrature coefficients does not apply to non-diffuse BRDFs. A direct application of the method to view dependent BRDFs would require massive precomputations of multidimensional tables so as to allow an acceptable rendering time. The prior model construction strategy presented in Brouillat et al. [2] would also be inappropriate to the sharpness features of glossy BRDFs. It would require performing a learning phase for each BRDF present in the scene, which would represent a high computational cost. Furthermore, the proposed samples set optimization method is not suited to the highly non-uniform samples distributions needed to efficiently compute glossy reflections. Applied to glossy BRDFs, their method converges towards local minima which are very far from the optimal solution.

In this paper we propose a new theoretical framework that includes a novel method of quadrature computation based on spherical Gaussian functions that can be generalized to a broad class of BRDFs (any BRDF which can be approximated by a sum of one or more spherical Gaussian functions) and potentially to other rendering applications. We account for the BRDF sharpness by using a new computation method for the prior mean function and by introducing a new factor in the method of Brouillat et al. [2] for constructing optimized samples set. Lastly, we propose a fast hyperparameters evaluation method that avoids the learning step.

In the following, after a presentation of related work, we introduce the theoretical aspects of BMC. Then we present the application of BMC to the illumination integral and develop our theoretical framework. It is followed by a description of the full rendering algorithm and a presentation of comparative results. Then, we suggest various

(a) Reference   (b) LDIS (RMSE = 0.039, time = 2m18s)   (c) BMC (RMSE = 0.026, time = 2m20s)
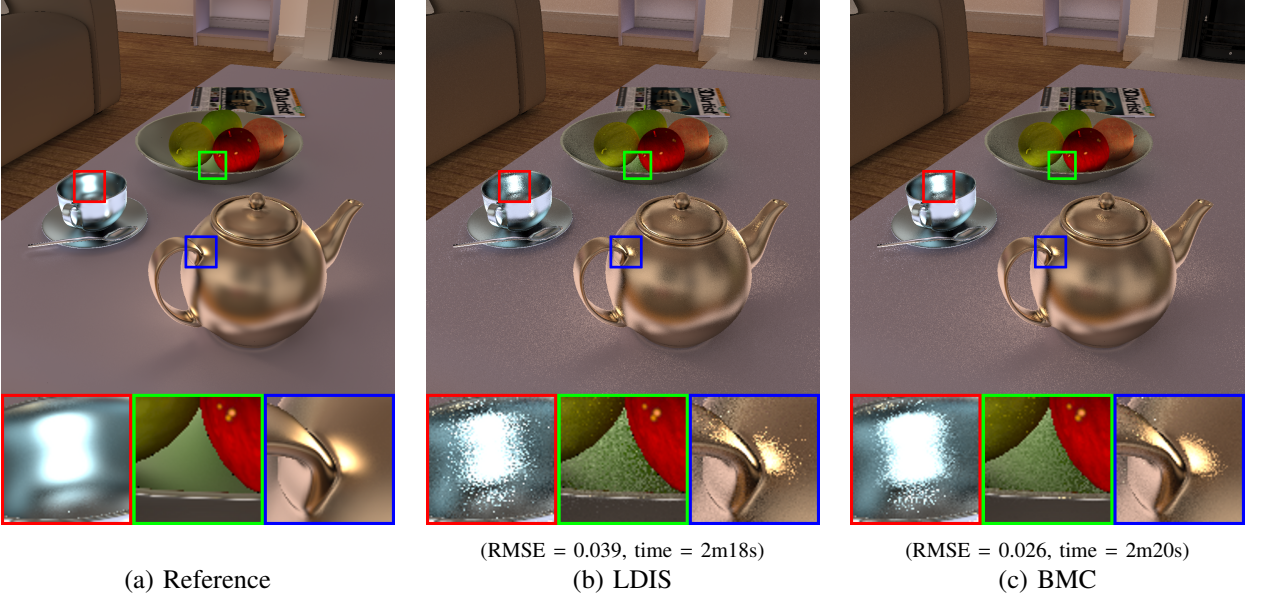
Fig. 1. Indirect radiance component for the Room scene rendered with low discrepancy Monte Carlo Importance Sampling (LDIS, (b)) and BMC (c). The shown images have been multiplied by a factor of 4. The RMSE and the time are computed by considering only the glossy component. 16 ray samples per visible point were used for the materials with a glossy BRDF (teapot, tea cup and fruit-dish), while 64 samples were used for the diffuse BRDFs.

research directions in a discussion section. We conclude by summarizing what was learned from our implementation and mention new perspectives of application of BMC.

2 RELATED WORK

The fundamental difference between BMC and CMC lies in the use of a prior stochastic model of the function $f(\mathbf{x})$ to be integrated. Note that in the Bayesian framework the argument $\mathbf{x}$ is considered deterministic, as opposed to CMC methods. Therefore, the randomness is not introduced by random sampling as in CMC but by the uncertainty (in a Bayesian sense) we put on the function to be integrated before any samples are drawn. Several approaches have been proposed in the literature regarding how to model this uncertainty. Our approach is based on the Bayes-Hermite quadrature of O'Hagan [1] which uses a Gaussian process (GP) for the prior model. We have found this method more appropriate for the computation of global illumination integrals because it easily leads to closed-form solutions. Rasmussen and Ghahramani [3] have shown that GP-based BMC can significantly outperform Monte Carlo importance sampling (MCIS). Several applications of the Bayes-Hermite quadrature have then been proposed in the literature and among them, the work of Pfingsten et al. [4] is particularly interesting in the way they derive a closed-form solution. However, it is not suited to the particularities of spherical functions we are faced with in global illumination problems. Brouillat et al. [2] have proposed solutions in the context of final gathering for photon mapping and have clearly shown the benefit of BMC over MCIS. However, this method can only deal with diffuse reflections. The main goal of this paper is thus to propose a BMC framework which can efficiently deal with the problems raised by glossy reflections.

$\mathbf{x}$	Point in $\mathbb{R}^D$
$p(\mathbf{x})$	Analytically known function
$f(\mathbf{x})$	Unknown function interpreted as a random quantity and modeled through a Gaussian Process (GP)
$\bar{f}(\mathbf{x})$	Mean function $\mathrm{E}[f(\mathbf{x})]$
$\bar{\mathbf{F}}$	Vector of mean function values
$k(\mathbf{x}, \mathbf{x}')$	General non-stationary covariance function
$k(\mathbf{x} - \mathbf{x}')$	Stationary covariance function
$k'(\mathbf{x} - \mathbf{x})$	Stationary correlation function
$\mathbf{k}(\mathbf{x})$	Vector of covariance values between a position $\mathbf{x}$ and a set of sample locations $\mathbf{x}_1, \ldots, \mathbf{x}_n$
K	Noise-free covariance matrix
Q	Covariance matrix
Y_i	Noisy observation of f at location $\mathbf{x}_i$ ($f(\mathbf{x}_i) + \epsilon_i$)
$\mathbf{Y}$	Vector of observations $(Y_1, \ldots, Y_n)$
ϵ_i	Sample of an i.i.d. Gaussian noise ($\mathrm{E}[\epsilon_i] = 0$)
σ_n^2	Variance of ϵ_i
σ_f^2	Variance of the GP
σ_y^2	Variance of the observations ($\sigma_n^2 + \sigma_f^2$)
$L_i(\omega)$	Incident radiance from direction ω
l	Lengthscale of the GP which models L_i
$G()$	Spherical Gaussian function
$\rho()$	BRDF
m	Shininess coefficient of the BRDF
w	Gaussian BRDF lobe width $w = 1/\sqrt{m}$

TABLE 1
Notations description.

3 BACKGROUND

In this section, we give a brief introduction to the theoretical basis of BMC, and make a short description of the previous application of BMC to the diffuse BRDF case. Detailed information can be found in [1], [2], [3]. The notations used throughout the paper are described in Tab. 1.

3.1 The Gaussian process prior

As in any Bayesian method, we need to state our knowledge prior to performing observations. This is usually defined as a probability distribution often simply called the prior. In BMC, the prior is modeled by a Gaussian process as explained in the following.

Let us consider the computation of the integral:

$$I = \int f(\mathbf{x})p(\mathbf{x})d\mathbf{x} , \qquad \text{with } \mathbf{x} \in \mathbb{R}^D \qquad (1)$$

where $p(\mathbf{x})$ is analytically known and $f(\mathbf{x})$ can only be determined through physical observations or, as in our case, through numerical evaluations. Given the high cost of these evaluations, it is infeasible to have a detailed knowledge about $f(\mathbf{x})$ in all its domain. Our knowledge about $f(\mathbf{x})$ is restricted to a limited set of samples while anywhere else in the domain, we are not sure about the value of $f(\mathbf{x})$. This uncertainty about $f(\mathbf{x})$ leads us to interpret it as a random quantity. The Bayesian reasoning states that all forms of uncertainty can be modeled by probability. Consequently, $f(\mathbf{x})$ can be considered as random because its value is unknown and thus uncertain. To model our uncertainty about $f(\mathbf{x})$, we will use a stochastic model called Gaussian process (GP). For a more practical knowledge about GP, the reader may refer to [5]. Formally, a GP is a collection of random variables, any finite number of which has 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 must be positive definite [6]:

$$\bar{f}(\mathbf{x}) = \mathrm{E}[f(\mathbf{x})] \qquad (2)$$
$$k(\mathbf{x}, \mathbf{x}') = \mathrm{E}[(f(\mathbf{x}) - \bar{f}(\mathbf{x}))(f(\mathbf{x}') - \bar{f}(\mathbf{x}'))] \qquad (3)$$

and will be denoted as:

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

We use the GP formalized by Eqs. (2) and (3) as our prior model. $\bar{f}(\mathbf{x})$ is our expectation of $f(\mathbf{x})$ before any observation is made. The covariance function characterizes the belief we have on the smoothness of $f(\mathbf{x})$, i.e., how correlated are nearby samples. A strong correlation between the samples implies that $f(\mathbf{x})$ is very smooth. The covariance function of the prior GP is often assumed stationary. A covariance function k is stationary when $k(\mathbf{x}, \mathbf{x}') := k(\mathbf{x} - \mathbf{x}')$, $\forall(\mathbf{x}, \mathbf{x}')$, in which case the variance of $f(\mathbf{x})$, i.e., $k(\mathbf{x}, \mathbf{x}) = k(\mathbf{0})$, is constant. The prior covariance function k is parametrized by a set of *hyperparameters* which need to be determined in a preprocessing step. This problem will be addressed in Sections 4.2.1 and 5.2. For now, we assume that the hyperparameters are known.

3.2 The posterior Gaussian process

Once the prior is defined, we can collect observations of $f(\mathbf{x})$ so as to refine our model by leveraging the prior GP. The resulting process is also a GP and is called posterior GP. In this section, we describe how to compute the posterior GP given the prior and a set of observations.

Let us suppose that we are provided with a set $\mathcal{D}$ of noisy samples of $f(\mathbf{x})$:

$$\mathcal{D} = \{(\mathbf{x}_i, Y_i) \,|\, i = 1, \ldots, n\} \quad \text{with} \quad Y_i = f(\mathbf{x}_i) + \varepsilon_i$$

the ε_i being samples of an independent, identically distributed Gaussian noise with zero mean and variance σ_n^2, $\mathbf{x}_i$ being a sample location (the input) and Y_i its corresponding sample value (the output). Given this additive noise assumption, the covariance of the observations becomes:

$$\mathrm{cov}(Y_p, Y_q) = k(\mathbf{x}_p, \mathbf{x}_q) + \sigma_n^2 \delta_{pq} \qquad (4)$$

where δ_{pq} is the Kronecker symbol. σ_n^2 is thus a hyperparameter of the prior model and represents the variance which is unexplained by the GP. It is important to understand that this noise component is not only due to measurement errors and computation inaccuracies. In a very broad sense, the noise term allows to accommodate highly discontinuous data that do not fit the smoothness assumption implied by the prior GP. Indeed, illumination functions are generally discontinuous and cannot be modeled accurately with smooth basis functions. The noise term provides the necessary flexibility so that when building the posterior process, the most plausible smooth model can be fitted to discontinuous data. This data fitting has the same effect as the pre-filtering method proposed by Křivánek and Colbert [7] which aims at low-pass filtering the observed samples to reduce aliasing. However GP model fitting does not assume any band-limited functions hypothesis. Moreover, smoothness has to be understood in terms of order of continuity of the integrand, not in terms of spectral bandwidth. In conclusion, BMC is applicable whatever the smoothness or the bandwidth of the integrand but of course, it performs better when the noise level σ_n^2 is low, that is when the prior fits well the data.

The posterior process results from applying the Bayes' rule to incorporate the information brought by the samples. It can be shown that the obtained posterior process is also a GP with mean and covariance functions given by [6]:

$$\mathrm{E}[f(\mathbf{x})|\mathcal{D}] = \bar{f}(\mathbf{x}) + \mathbf{k}(\mathbf{x})^t Q^{-1}(\mathbf{Y} - \bar{\mathbf{F}}) \quad (5)$$
$$\mathrm{cov}[f(\mathbf{x}), f(\mathbf{x}')|\mathcal{D}] = k(\mathbf{x}, \mathbf{x}') - \mathbf{k}(\mathbf{x})^t Q^{-1}\mathbf{k}(\mathbf{x}')$$

with:

$$\begin{aligned}
\mathbf{k}(\mathbf{x}) &= (k(\mathbf{x}_1, \mathbf{x}), \ldots, k(\mathbf{x}_n, \mathbf{x}))^t \\
K_{i,j} &= k(\mathbf{x}_i, \mathbf{x}_j) \qquad \text{with } (i, j) \in [1, n]^2 \\
Q &= (K + \sigma_n^2 I_n) \qquad\qquad (6) \\
\mathbf{Y} &= (Y_1, \ldots, Y_n)^t \\
\bar{\mathbf{F}} &= \left(\bar{f}(\mathbf{x}_1), \ldots, \bar{f}(\mathbf{x}_n)\right)^t
\end{aligned}$$

and I_n being the $n \times n$ identity matrix. $\mathbf{Y}$ is the vector of observed samples value while $\bar{\mathbf{F}}$ is the vector of prior GP mean values at the sampling locations. Q contains the covariance between the samples and is called covariance matrix. Eq. (5) gives an estimate of $f(\mathbf{x})$ for an unobserved input $\mathbf{x}$ given the observed data set $\mathcal{D}$. This particular form of regression is called *Bayesian regression*.

3.3 Bayesian quadrature

Eq. (5) gives a posterior estimate of $f(\mathbf{x})$ given the observed samples. Then, from Eq. (1), a posterior estimate of I, that is $\hat{I} = \mathrm{E}(I|\mathcal{D})$, is obtained by integrating both terms of Eq. (5):

$$\hat{I} = \bar{I} + \mathbf{z}^t Q^{-1}(\mathbf{Y} - \bar{\mathbf{F}}) \tag{7}$$

with:

$$\bar{I} = \int \bar{f}(\mathbf{x})p(\mathbf{x})d\mathbf{x} \tag{8}$$

$$\mathbf{z} = \int \mathbf{k}(\mathbf{x})p(\mathbf{x})d\mathbf{x} \tag{9}$$

It can be seen that the posterior estimate $\hat{I}$ results from adding to the prior expectation $\bar{I}$ a corrective term that represents the effect of the observed samples $\mathcal{D}$. This term includes three factors: $(\mathbf{Y} - \bar{\mathbf{F}})$, which measures how wrong was our prior expectation for each observed sample value, Q^{-1}, the inverse covariance matrix which accounts for the relative positions of the observed samples, and the $\mathbf{z}$ vector, which captures the influence of each sample on the deterministic part of the integrand given the covariance function.

3.4 Variance analysis

3.4.1 Variance of the integral estimate and optimal sampling pattern

The posterior estimate of I given by Eq. (7) has a Gaussian distribution of mean $\hat{I}$ and a variance given by:

$$\mathrm{Var}(I|\mathcal{D}) = \bar{V} - \mathbf{z}^t Q^{-1}\mathbf{z} \tag{10}$$

with:

$$\bar{V} = \iint k(\mathbf{x}, \mathbf{x}')p(\mathbf{x})p(\mathbf{x}') \, d\mathbf{x}d\mathbf{x}' \tag{11}$$

Note that the estimate of the posterior variance in Eq. (10) does not depend on the observed values $\mathbf{Y}$. It only depends on the location $\mathbf{x}$ of the samples. This might seem not plausible at first sight but actually, the covariance function of our prior model (Eq. (4)) already comprises the statistical information necessary to estimate $\mathrm{Var}(I|\mathcal{D})$. This variance is due to the implicit variability of our observations given our prior. It can be interpreted as a measure of the confidence we may attribute to our integral estimate. Note also that BMC does not require drawing the samples randomly according to a PDF. However, Eq. (10) shows that the $\mathrm{Var}(I|\mathcal{D})$ strongly depends on the choice of the samples set position $\{\mathbf{x}_i\}$. An optimal choice consists in selecting the $\{\mathbf{x}_i\}$ that minimizes $\mathrm{Var}(I|\mathcal{D})$. This optimal set will reflect both the influence of the deterministic function $p(\mathbf{x})$ (through the $\mathbf{z}$ vector) and that of the prior knowledge on $f(\mathbf{x})$ (through the covariance function). The resulting effect can be compared to product sampling or bidirectional sampling techniques that try to obtain information on $f(\mathbf{x})$ in order to improve the efficiency of importance sampling.

3.4.2 Bias considerations

As opposed to CMC, in BMC the observed data are considered as known and thus deterministic whereas the function $f(\mathbf{x})$ (Eq. (1)) is considered as uncertain. The BMC integral estimate is obtained by computing the integral of the most probable function among all possible realizations of the posterior GP. The estimator of Eq. (7) is thus unbiased (in the Bayesian sense) as its value coincides with the expected value of the posterior probability distribution. However this unbiasedness has nothing to do with the usual interpretation of this term in a classic Monte Carlo framework in which sampling is produced by a random process. That is why unbiasedness considerations can be misleading when comparing classic and Bayesian Monte Carlo, since the randomness in each of the methods is of a very different nature.

Specifically, in a frequentist approach such as CMC, the estimate can be refined by averaging estimates with respect to multiple sample sets whereas a Bayesian method uses averaging with respect to the posterior distribution instead. In [8] (Chapter 3, Section 2), Bishop applies a frequentist bias-variance analysis to regression and he shows on a parametric regression example how the regularization parameter (which, in BMC, corresponds to the noise ratio introduced in the next section) can be used to control the frequentist bias. However, Bishop points out that this is of limited practical interest since averaging w.r.t. multiple sample sets would affect the efficiency of a Bayesian approach. Indeed, as noted by Bishop, to fully benefit from the Bayesian approach, a direct application of BMC would combine multiple sample sets into one large set rather than considering each set individually. Therefore, in our BMC method, we will consider only one single samples set and apply a full-fledged Bayesian method. Furthermore, the Bayesian framework offers much more efficient methods to implement progressive refinement than averaging with respect to multiple sample sets.

3.5 The case of stationary covariance functions and the noise ratio hyperparameter

In the absence of accurate information on $f(\mathbf{x})$ variations, it is sensible to assume a stationary covariance function for the prior GP. Note however that this does not mean that the prior GP is second-order stationary. For this to be true, the mean function $\bar{f}(\mathbf{x})$ must be constant, which is not the case in general.

As mentioned in Section 3.1, the variance of the prior GP is constant in the case of stationary covariance functions and we have:

$$k(\mathbf{x} - \mathbf{x}') = \sigma_f^2 k'(\mathbf{x} - \mathbf{x}') \tag{12}$$

where $\sigma_f^2 = k(0)$ is the prior GP variance and $k'(\mathbf{x} - \mathbf{x}')$ is its correlation function. σ_f is thus an hyperparameter of our GP model. It represents the strength of the correlation induced by the GP. As the additive noise ε is assumed independent from the prior GP, the variance of the observed

samples value σ_y^2 can be expressed as follows:

$$\sigma_y^2 = \sigma_f^2 + \sigma_n^2$$

The above equations reveal the σ_f and σ_n hyperparameters as totally independent variables. However, in global illumination problems, many integrals of the same type but with different data have to be computed to render a full image. The observed variance σ_y^2 can strongly vary from one integral to the other, but in practice, we can expect that the values of both hyperparameters will roughly have the same order of magnitude as σ_y^2. Therefore, the noise contribution is better characterized by the ratio σ_n/σ_f rather than σ_n alone. Henceforth, we will thus use the noise ratio hyperparameter $\sigma_n' = \sigma_n/\sigma_f$ instead of σ_n to represent the noise contribution.

In the following of this section, we show that once σ_n' has been determined, the value of σ_f is not required to compute the integral estimate $\hat{I}$.

Given Eq. (12), the Q matrix defined in Eq. (6) can be expressed as follows:

$$
\begin{aligned}
Q &= \sigma_f^2 Q' \\
\text{with} \quad Q' &= K' + \sigma_n'^2 I_n
\end{aligned}
\tag{13}
$$

where K' is obtained by replacing $k(\mathbf{x} - \mathbf{x}')$ by $k'(\mathbf{x} - \mathbf{x}')$:

$$K_{i,j}' = k'(\mathbf{x}_i - \mathbf{x}_j) \qquad \text{with } (i, j) \in [1, n]^2$$

We can now rewrite Eq. (7) with terms that depend on $\sigma_n'^2$ instead of σ_f^2 and σ_n^2:

$$\hat{I} = \bar{I} + \mathbf{z}'^t Q'^{-1}(\mathbf{Y} - \bar{\mathbf{F}}) \tag{14}$$

where the $\mathbf{z}'$ vector is obtained by replacing $k(\mathbf{x} - \mathbf{x}')$ by $k'(\mathbf{x} - \mathbf{x}')$ in Eq. (9). Each element of this vector can then be expressed by the following convolution integral:

$$z_i' = \int k'(\mathbf{x} - \mathbf{x}_i)p(\mathbf{x})d\mathbf{x} \tag{15}$$

where $\mathbf{x}_i$ is a sample location. Assuming that the σ_n' hyperparameter is known, the integral estimate given by Eq. (14) does not require determining σ_f^2 and σ_n^2. The noise ratio σ_n' can be interpreted as the level of confidence we have in the GP model fitting. The higher is σ_n', the lower is the proportion of the observed samples variance that can be explained by the GP.

3.6 BMC algorithm overview

To summarize, the BMC method consists of the following steps:

1) Build the GP prior by choosing a covariance function and a mean function
2) Learn the hyperparameters associated with the covariance function
3) Select the set of sampling positions $\{\mathbf{x}_i\}$ which minimize the variance of the BMC estimate and compute the inverted covariance matrix Q'^{-1}
4) Compute the $\mathbf{z}'$ vector and the vector of quadrature coefficients $\mathbf{c} = \mathbf{z}' Q'^{-1}$

5) Collect the observed samples value $\mathbf{Y}_i$ for each sampling position $\{\mathbf{x}_i\}$
6) Compute the prior mean value vector $\bar{\mathbf{F}}$
7) Compute the posterior estimate with Eq. (14)

As mentioned above, global illumination problems imply the computation of many integrals of the same type with different data and of course, all these steps need not be repeated for each integral evaluation. As a matter of fact, in implementing a BMC algorithm, we will try to transfer the most computer intensive steps to a preprocessing stage. By using the same set of hyperparameters for all integral evaluations, Brouillat et al. [2] have shown that steps 1 to 4 can be preprocessed in the case of diffuse reflection, which makes the computing load of BMC comparable to CMC methods. However, their approach is not applicable to the case of glossy BRDF because of the dependency on the viewing direction. The deterministic function $p(\mathbf{x})$ that essentially contains the BRDF will then change at each integral evaluation and steps 2 and 3 can no longer be preprocessed as we will see in the following. In the next sections, we will propose a more general approach based on a spherical Gaussian framework that can be applied to a large class of problems.

4 A SPHERICAL GAUSSIAN-BASED BMC FRAMEWORK

We will now consider the computation of the illumination integral at a given shading point:

$$L_o(\omega_o) = \int_{\Omega_{2\pi}} L_i(\omega_i) \, \rho(\omega_i, \omega_o) \, (\omega_i \cdot \mathbf{n}) \, d\Omega(\omega_i) \tag{16}$$

In this integral, the analytical part, which corresponds to $p(\mathbf{x})$ in Eq. (1) is naturally the BRDF factor $\rho(\omega_i, \omega_o)(\omega_i \cdot \mathbf{n})$. The unknown function $f(\mathbf{x})$ in Eq. (1), modeled with a Gaussian Process, is the incident radiance $L_i(\omega_i)$ at the shading point in Eq. (16). Our goal is to compute an estimate of $L_o(\omega_o)$ from a set of samples value $\{L_i(\omega_j), j \in [1, n]\}$, where ω_j are the samples location (called $\mathbf{x}_i$ in Section 3.2).

Throughout this section, we describe our approach to the different problems involved in computing the Bayesian quadrature. We start by describing a spherical Gaussian-based framework for BMC which greatly simplifies the computation of the integrals for the $\mathbf{z}'$ vector. Then we show how the prior GP covariance and mean functions are built, and how optimal sampling patterns are built. Finally, we elaborate on our rendering algorithm.

4.1 Our theoretical approach

In this section, we essentially address step 4 of the BMC algorithm presented in Section 3.6 and more particularly the problem of making the computation of the $\mathbf{z}'$ vector tractable knowing that it is not possible to precompute the z_j' coefficients given their dependency on ω_o. Our approach is to model both the BRDF and the correlation function with spherical Gaussian functions (SGF), a choice which allows us to reduce the $\mathbf{z}'$ computation to a simple query to a scene-independent 2-entry look-up table.

An interesting property of SGFs is that the product of two SGFs yields a SGF. This property will be very useful for the computation of the Eq. (15) integral. As far as $p(\mathbf{x})$ can be modeled by a SGFs product or mixture, the method described in this section can be applied to compute the quadrature of Eq. (14) given its linearity properties.

A SGF results from the restriction of a Gaussian RBF (Radial Basis Function) to the unit sphere S^2. Consequently, for $x, y \in S^2$, we have:

$$|x - y|^2 = 2(1 - x \cdot y) \qquad (17)$$

since $|x| = |y| = 1$. A SGF can thus be expressed as follows:

$$G(x - y; \mu, \lambda) := \mu \exp\left(\frac{x \cdot y - 1}{\lambda^2}\right)$$

Since to each point $x \in S^2$ corresponds a direction ω, we model the correlation function k' by a SGF $G_{k'}$ defined as follows:

$$k'(\omega - \omega') := G_{k'}(\omega - \omega'; 1, l) = \exp\left(\frac{\omega \cdot \omega' - 1}{l^2}\right) \quad (18)$$

where l is the lengthscale hyperparameter which controls the smoothness of the prior GP. We model the BRDF ρ by a SGF G_ρ defined as:

$$\rho(\omega_i, \omega_o) := G_\rho(\omega_r - \omega_i; k_s, w) = k_s \exp\left(\frac{\omega_r \cdot \omega_i - 1}{w^2}\right) \quad (19)$$

where k_s is the specular coefficient. G_ρ has an axially symmetric lobe whose axis is aligned with ω_r which is itself function of the outgoing direction ω_o. w characterizes the lobe sharpness, i.e., the "width" of the lobe. It is important to note that this choice does not restrict the generality of our analysis as several works have shown that most BRDFs can be approximated by a mixture of SGFs (e.g. in [9]). In particular, one single SGF is sufficient to model the glossy term of a Phong's BRDF. In this case, the lobe axis direction ω_r is the perfect mirror incident direction:

$$\omega_r = 2(\omega_o \cdot \mathbf{n})\mathbf{n} - \omega_o$$

and using the following approximation:

$$\cos^m \theta = e^{m \ln(\cos \theta)} \approx e^{m(\cos \theta - 1)} \qquad (20)$$

with $\cos \theta = (\omega_r \cdot \omega_i)$, we have $w = 1/\sqrt{m}$, m being the Phong shininess parameter. The resulting approximation RMSE is below 10^{-3} when $m > 10$. Note also that modeling BRDFs with SGFs (which are naturally isotropic) does not restrict our framework to the isotropic BRDF case since anisotropic BRDFs can be modeled using a weighted sum of (isotropic) SGFs as mentioned above. Moreover, we have found that the Blinn-Phong BRDF can be modeled with only 3 SGFs with a RMSE below 0.05. More details regarding the potential application of our approach to multiple-lobe BRDFs are given in Section 6.2.

Given Eqs. (15), (18) and (19), each coefficient of the $\mathbf{z}'$ vector can be expressed as follows:

$$z'_j = \int_{\Omega_{2\pi}} G_{k'}(\omega_j - \omega'; 1, l) \, G_\rho(\omega_r - \omega'; k_s, w) \, d\Omega(\omega')$$

where ω_j is the direction vector of sample j. As the product of two SGFs is also a SGF, we have:

$$z'_j = \int_{\Omega_{2\pi}} G(\omega_m - \omega'; c_m, l_m) \, d\Omega(\omega') \qquad (21)$$

with:

$$\begin{aligned}
\frac{1}{l_m^2} &= \left| \frac{\omega_j}{l^2} + \frac{\omega_r}{w^2} \right| \\
\omega_m &= l_m^2 \left(\frac{\omega_j}{l^2} + \frac{\omega_r}{w^2} \right) \\
c_m &= k_s \exp\left(\frac{1}{l_m^2} - \frac{1}{l^2} - \frac{1}{w^2} \right)
\end{aligned}$$

By developing Eq. (21), we have:

$$z'_j = c_m S_g(\omega_m, l_m)$$

where $S_g(\omega, l)$ is the spherical Gaussian integral (SGI):

$$S_g(\omega, l) = \int_{\Omega_{2\pi}} \exp\left(\frac{\omega \cdot \omega' - 1}{l^2} \right) d\Omega(\omega') \qquad (22)$$

The computation of the $\mathbf{z}'$ coefficients is then reduced to evaluations of the SGI of Eq. (22) which can easily be tabulated for quick evaluations. Only the elevation angle $\theta = (\omega, \mathbf{n})$ is necessary to specify the input direction ω in Eq. (22) given the axial symmetry of the SGI about the normal $\mathbf{n}$. Therefore, only a single 2-entry table with (θ, l) as inputs is necessary for SGI evaluations. This table is independent of the scene and the used BRDFs. It is computed just once and used for any BMC integration within this framework.

4.2 Constructing the prior GP model

4.2.1 Hyperparameters selection

In the preceding sections, we have shown that only two hyperparameters are needed to compute the integral estimate from Eq. (14): the noise ratio σ'_n and the lengthscale l. Our strategy for determining appropriate hyperparameters value consists in using the same set of hyperparameters for all illumination integrals that involve the same BRDF. We shall see in Section 5 that this choice is perfectly acceptable since the hyperparameters values mainly depend on the BRDF shininess, although they can vary depending on the scene lighting conditions. This simplification allows us to learn the hyperparameters for each BRDF at a scene level, and then precompute the inverse covariance matrix Q^{-1} and the optimal sampling pattern (i.e., steps 2 and 3 of the algorithm of Section 3.6).

To learn the hyperparameters at a scene level for a given BRDF, we use the same technique as that of Brouillat et al. [2], with the difference that we sample the incoming radiance with a distribution concentrated within the BRDF lobe. In this way, the fitting of the covariance function is adapted for the range of interest of the inter-samples distance, that is, a narrow range of distances for high shininess and a wide one for low shininess. But this approach has a limitation regarding scenes with several BRDFs since it implies learning the hyperparameters separately for

each shininess parameter used in the scene. Although this operation takes just a few seconds per shininess parameter (typically around 1 to 4 seconds depending on the scene complexity), it would make the BMC approach inefficient for scenes with different shininess parameters. To cope with this problem, a fast method to approximate the hyperparameters is proposed in Section 5.4.

4.2.2 Determining the mean function

To fully define the GP, a mean function $\bar{L}_i(\omega_i)$ corresponding to $\bar{f}(\mathbf{x})$ in Section 3.1 must be specified. Our approach to this question is to assume a locally constant mean function $\bar{L}_i$ that will be determined from the observed samples value $L_i(\omega_j)$. A simple average of the samples value would provide a strongly biased $\bar{L}_i$ estimate since the samples distribution is not at all uniform as explained in Section 4.3. Instead, we leverage the prior GP as described in Chapter 2, Section 7 of [6]. The method consists in inferring a mean function from the observed samples value through an explicit basis function model. In our case, we only use one basis function ($h(\omega) = 1$) and its associated weight is thus the desired constant mean value, which yields:

$$\bar{L}_i = \frac{HQ'^{-1}Y}{HQ'^{-1}H^t} \qquad (23)$$

where $H = [h(\omega_1), \ldots, h(\omega_n)] = [1, \ldots, 1]$ and $Y = [L_i(\omega_1), \ldots, L_i(\omega_n)]^t$ is the vector of observations. As shown in [6], this method slightly increases the estimates variance but this can be neglected in practice. Then from Eq. (8) and Eq. (19), we have:

$$\bar{I} = \bar{L}_i \, \mu(\omega_r) \qquad (24)$$

where $\mu(\omega_r)$ is the SGI:

$$\mu(\omega_r) = \int_{\Omega_{2\pi}} G_\rho(\omega_r - \omega; k_s, w) \, d\Omega(\omega)$$

To evaluate this integral, the same 2-entry table as the one required for evaluating the SGI of Eq. (22) can be used. Moreover, since $H = [1, \ldots, 1]$, Eq. (23) can be expressed as a weighted sum of the observed samples value:

$$\bar{L}_i = \frac{1}{\Gamma} \sum_j \gamma_j Y_j$$

where the weight γ_j is equal to the sum of the coefficients of the jth column of the Q'^{-1} matrix and $\Gamma = \sum_j \gamma_j$ is the sum of all coefficients of the Q'^{-1} matrix. All the γ_j/Γ weights can be precomputed once the sampling pattern and the hyperparameters have been determined.

4.3 Optimal sampling pattern

The choice of the sample directions $\{\omega_j\}$ is crucial to the quality of the rendering integral estimate, especially in the case of glossy BRDFs. Nevertheless, not only the BRDF but also the characteristics of the incident radiance function L_i should be taken into account for this choice. These characteristics are represented in a stochastic manner through the hyperparameters of the prior GP.

The variance estimate in Eq. (10) factors in the influence of the BRDF (through $\mathbf{z}'$), the prior GP hyperparameters, and the sampling pattern defined by the $\{\omega_j\}$ set. Moreover, since the *global hyperparameters* σ'_n and l are known at this stage (see Section 3.6), the variance only depends on the $\{\omega_j\}$ set. This is a strong feature of the BMC framework since we can optimize the samples direction in order to minimize $\text{Var}(I|\mathcal{D})$.

Optimizing the directions for a large set of samples can become cumbersome. A naive approach would be to consider that the variance is a function of n variables which are the samples directions. But minimizing such a function would become intractable even for medium size sample sets (40 samples or more). Our solution to this problem consists in modifying the spiral points algorithm [10] as shown in Alg. 1.

Algorithm 1 The modified spiral points algorithm.

1: $\Delta\phi \leftarrow \pi(3 - \sqrt{5})$ ▷ Compute the step on ϕ
2: $\phi \leftarrow 0$ ▷ Initialize ϕ for the first sample
3: $\Delta z \leftarrow 1/n$ ▷ Compute the step on z
4: $z \leftarrow 1 - \Delta z/2$ ▷ Initialize z for the first sample
5: **for all** $k \leftarrow [1:n]$ **do**
6:     $z_p \leftarrow \text{polyval}(coeff, z)$ ▷ Covariance effect
7:     $z_v \leftarrow \frac{1}{m} \ln[1 + z_p(e^m - 1)]$ ▷ BRDF effect
8:     $\theta_k \leftarrow \arccos(\min(z_v, 1))$ ▷ Compute θ angle
9:     $\phi_k \leftarrow \text{mod}(\phi, 2\pi)$ ▷ Compute ϕ angle
10:     $z \leftarrow z - \Delta z$ ▷ Give a step on z
11:     $\phi \leftarrow \phi + \Delta\phi$ ▷ Give a step on ϕ
12: **end for**

The original algorithm generates a discrete spiral on a sphere about the up axis z with a constant pitch (Δz). In our application, the z axis is aligned with the surface normal $\mathbf{n}$. The z coordinates of the original spiral points are thus uniformly distributed. We replace the uniform distribution by a polynomial distribution (whose coefficients are called *coeff* in Alg. 1, line 6) to account for the covariance effect (line 6). In addition, we change this algorithm to produce a discrete spiral on the BRDF lobe (line 7) rather than on the sphere. In Section 4.4 we show that this can be obtained by warping the z coordinates with the following function:

$$z_v = g(z) = \frac{1}{m} \ln[1 + z(e^m - 1)]$$

where m is the Phong shininess parameter.

The optimizer will then take the polynomial coefficients *coeff* as input arguments to minimize $\text{Var}(I|\mathcal{D})$ instead of taking as input all the sampling directions. The number of parameters to optimize is thus reduced to the number of coefficients of the polynomial (a polynomial of degree 3 is sufficient). Actually, only the term $-\mathbf{z}^t Q^{-1} \mathbf{z}$ of Eq. (10) needs to be considered as $\bar{V}$ does not depend on the sampling pattern. The convergence is very fast (a few seconds) with the BFGS quasi-Newton method provided by Matlab and the result is not very sensitive to the initial values even if the initial polynomial is of degree 0.

The variance of the BMC estimate strongly depends on the correlation function which is, in its turn, parametrized by the lengthscale hyperparameter. We can thus expect the optimal sample set to depend on the lengthscale value. A very smooth L_i function implies a large lengthscale value. Consequently, a high concentration of samples around the BRDF apex would be redundant as the samples are highly correlated. This explains the sparsity of the samples on the lobe (Fig. 2(b)). Conversely, if L_i has a low lengthscale value the samples will tend to be concentrated around the BRDF lobe apex (Fig. 2(a)). In a preprocessing step, we compute optimal sampling patterns for a discrete set of pairs (l, m), l and m being the lengthscale and the Phong shininess parameter respectively. These optimal sampling patterns are independent of the scene to be processed. When rendering, we select the appropriate sampling patterns from the precomputed ones.

Note that the use of an optimized sampling pattern is beneficial to fully exploit the BMC method. It allows accounting for the prior information about the unknown function L_i on the samples distribution. This aspect is detailed in Section 5.
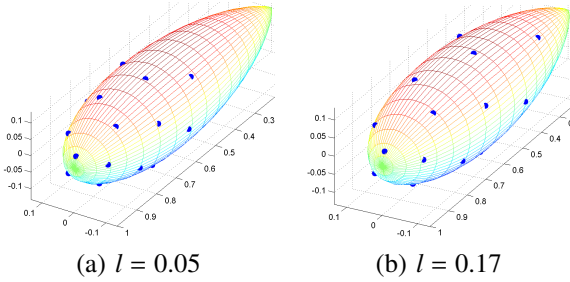

(a) $l = 0.05$ (b) $l = 0.17$

Fig. 2. Optimal sampling patterns of size 20 computed for two different lengthscale values and a purely glossy Phong BRDF with a shininess $m = 50$. Each blue point on the lobe corresponds to a sample direction.

4.4 The z warping function for Gaussian lobe

Our approach to this problem is to exploit the property of every function f on S^2 that:

$$\lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} f(\omega_{j,n}) = \frac{1}{4\pi} \int_{S^2} f(\omega) d\Omega(\omega)$$

for configurations of asymptotically uniformly distributed sampling directions $\{\omega_{1,n}, \ldots, \omega_{n,n}\}$ [10]. Therefore, at the beginning of the optimization algorithm, if the initial sampling pattern is such a configuration, all samples can contribute equally to the integral estimate, which can be considered as a sensible choice for a starting point. Our goal in this section is to show how such configurations can be built with the spiral points algorithm. In Section 4.3, we have assumed that the lobe central axis is aligned with the surface normal ($\omega_r = \mathbf{n}$) and consequently, using the spherical Gaussian BRDF expression given by Eq. (19), the illumination integral of Eq. (16) becomes:

$$L_o(\omega_o) = k_s \int_{\Omega_{2\pi}} L_i(\omega_i) \exp[m(\mathbf{n} \cdot \omega_i - 1)] d\Omega(\omega_i)$$

This equation can be further developed as follows:

$$L_o(\omega_o) = k_s \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\frac{\pi}{2}} L(\theta, \phi) e^{m(\cos \theta - 1)} \sin \theta \, d\theta d\phi$$

where (θ, ϕ) are the polar coordinates of the ω_i. If we make the substitution $z = \cos \theta$, we obtain:

$$L_o(\omega_o) = k_s \int_{\phi=0}^{2\pi} \int_{z=0}^{1} L(z, \phi) e^{m(z-1)} \, dz d\phi$$

If we directly sample the function $L(z, \phi) e^{m(z-1)}$, the distribution over the hemisphere $\Omega_{2\pi}$ must be uniform as stated above, which is precisely what the original spiral points algorithm does. But if we sample the function $L(z, \phi)$ only, another substitution is necessary to include the BRDF factor into new integration variables. This is possible by making the substitution:

$$z' = \frac{e^{mz} - 1}{e^m - 1} \tag{25}$$

and then:

$$dz' = \frac{m}{1 - e^{-m}} e^{m(z-1)} dz$$

which leads to:

$$L_o(\omega_o) = \frac{k_s(1 - e^{-m})}{m} \int_{\phi=0}^{2\pi} \int_{z'=0}^{1} L(z', \phi) \, dz' d\phi$$

Note that the integration domain remains the hemisphere $\Omega_{2\pi}$. This domain is then uniformly sampled as for the original integral but we still need to revert to the original z coordinate to obtain the polar coordinates of the sampling directions. By inverting Eq. (25), we have:

$$z = \frac{1}{m} \ln[1 + z'(e^m - 1)]$$

which gives the warping function we need for our spiral point algorithm. Its effect can be interpreted as a warping of the distance measure so that inter-samples distances are measured between points on the Gaussian lobe rather that on the unit sphere.

4.5 The rendering algorithm

The BMC integrator described in Alg. 2 is applied to every shading point. The BMC function (line 1) has two parameters: the outgoing direction ω_o and the surface normal $\mathbf{n}$. It returns the BMC estimate of $L_o(\omega_o)$ as defined in Eq. (16). All global data for the whole scene including the hyperparameters (l, σ'_n), the inverse covariance matrix Q'^{-1}, the optimal sampling pattern SS and the BRDF parameters (w, k_s) are assumed to belong to the program global scope and are thus not explicitly declared.

The optimal sampling pattern, initially centered around the surface normal $\mathbf{n}$, is rotated around $\mathbf{n}$ (line 2) by a random ϕ value. The idea behind this random rotation is to introduce a scrambling effect that allows masking the visibility of the regular sampling pattern, hence avoiding the introduction of artifacts in the final image. Note that since the covariance function is a SGF dependent only on the inter-samples distance, it is invariant to rotations of the whole sampling pattern. This property allows generating

---
Algorithm 2 The Bayesian Monte Carlo integrator

1: **function** BMC(ω_o, **n**)
2:     $SS_{rotated} \leftarrow$ rotateRandom$\phi(SS)$
3:     $\omega_r \leftarrow$ reflect$(\omega_o, \mathbf{n})$
4:     $SS_{rotated} \leftarrow$ centerOnR$(SS_{rotated}, \omega_r)$
5:     **for all** $\omega_j \in SS_{rotated}$ **do**
6:         $\mathbf{Y}[j] \leftarrow L_i(\omega_j)$
7:     **end for**
8:     $\bar{L}_i \leftarrow$ computeWeightedMean$(\mathbf{Y})$
9:     $\bar{I} \leftarrow \bar{L}_i \times \mu(\omega_r)$
10:     $\bar{\mathbf{F}} \leftarrow \bar{L}_i \times \mathbf{H}$
11:     $\mathbf{z}' \leftarrow$ computeZvector$(\omega_r, w, SS_{rotated})$
12:     $L_o \leftarrow \bar{I} + \mathbf{z}'^t \times Q'^{-1} \times (\mathbf{Y} - \bar{\mathbf{F}})$
13:     **return** L_o
14: **end function**

---

different optimal sampling patterns by randomly rotating the original optimal sampling pattern while keeping the same matrix Q'^{-1}. The mirror incident direction ω_r is computed at line 3. An additional rotation aligns the central axis of the sampling pattern with ω_r (line 4). The loop beginning at line 5 numerically evaluates the samples value for each sampling direction ω_j. $\bar{L}_i$ is computed at line 8 according to Eq. (23) and $\bar{I}$ is computed at line 9 according to Eq. (24). At line 11, $\mathbf{z}'$ is computed as described in Section 4.1 and Eq. (21) through a lookup in a scene-independent precomputed table of SGI values. Finally, the Bayesian quadrature (Eq. (14)) is applied to estimate L_o, which amounts to a vector-matrix and a vector-vector product.

As the sampling pattern SS is centered around ω_r, a part of the sampled BRDF lobe may be located under the plane tangent to the object surface at the shading point; in other words, some of the directions of the sampling pattern may be occluded by the object surface. A straightforward solution would be to simply discard these samples. But this would imply recomputing and inverting the covariance matrix whenever one or more samples of the samples set lie outside the domain of integration. To avoid this computational overhead, we use a fixed number of samples, which raises the problem of assigning a value to the occluded samples.

The values of the occluded samples could be naively set to zero, but such a method could potentially introduce wrong information with an important effect on the estimated integral value. The reason for this is that although the considered sampling direction ω_j is outside the integration domain, i.e., $\omega_j \notin \Omega_{2\pi}$, it may still contribute to the reconstruction of the function $L_i(\omega_i)$ within the integration domain. Setting the value of such samples to zero would introduce wrong information and force the reconstructed signal to tend towards zero when approaching the surface tangent plane. Indeed, there is no reason why an artificial discontinuity of the incident radiance should be introduced at the tangent plane since occlusion due to the surface is already accounted for through the choice of the integration domain. We have experimentally found that an efficient

solution to this problem is to assign to such samples the value of the closest sample that lies in the unoccluded part of the lobe. In particular, this solution performs better than assigning the occluded samples the mean value of the samples within the integration domain.

5 RESULTS

5.1 Experimental setup

In order to assess the effectiveness of the BMC method, its results are compared with those of BRDF-based low-discrepancy Monte Carlo importance sampling using a Halton sequence (LDIS), and pure random Monte Carlo importance sampling (MCIS). To evaluate the importance of the optimized sample sets in the BMC integration we also tested BMC using LD-based sample sets (LDBMC). The efficiency of each method is computed in terms of root mean square error (RMSE) with respect to reference images computed using LDIS with 10000 samples per pixel. We have chosen to test our framework using Phong shininess coefficients m varying from 10 to 200, a range of values which we consider illustrative of the common usage of glossy BRDFs. Six different scenes have been used, the details of which are presented in Tab. 2. In

Scene	triangles	m	l_l	σ'_{nl}
Dragon	100K	20	0.29	0.96
Buddha	170K	50	0.17	0.53
Horse	110K	80	0.13	0.67
VW	440K	20, 50, 80	nd	nd
Room	540K	20, 50, 80	nd	nd
Plates	20	10, 50, 80, 200	nd	nd

TABLE 2

Test profiles. Columns list the scene characteristics, the Phong shininess parameter m and the learned hyperparameters l_l, σ'_{nl}. Values of l_l and σ'_{nl} tagged as 'nd' indicate that the approximate method of hyperparameters is used, as shown in Section 5.4.

some scenes (i.e., Dragon, Buddha and Horse), for the sake of computation speed, incident illumination is simulated by an environment map, a simplification that does not limit the significance of our results and conclusions since our goal is to show that we can obtain better estimates with less samples and a negligible computing overhead whatever the used global illumination method. In other scenes (i.e., VW and Room), the results are generated using final gathering for photon mapping. As for the Plates scene, only direct incident illumination was computed. The results were generated at a resolution of 1024×1024 pixels for all the scenes except the Plates scene (1152×768 pixels), using a 64 bit machine equipped with a 2GHz Intel Core i7 processor and 8GB of RAM. The Mitsuba raytracer [11] was used to implement the methods of MCIS, LDIS, LDBMC and BMC.

5.2 Hyperparameters learning

The validity of our prior model depends on the good choice of the hyperparameters. Tab. 2 lists the hyperparameter

values obtained by learning for the scenes Dragon, Buddha and Horse. The learning method provides us with the hyperparameters that allow the best global adaptation of the GP model. But it is not guaranteed that these hyperparameters yield the best integral estimate. To validate our hyperparameters learning method, we have measured the RMSE for different values of the hyperparameters (see Fig. 3).
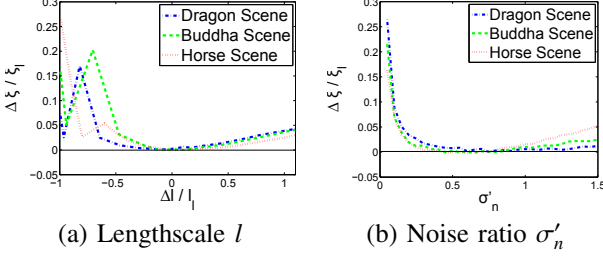

Fig. 3. Relative RMSE difference when varying one of the hyperparameters. $\Delta\xi = \xi - \xi_l$, ξ being the RMSE of an image computed when varying one hyperparameter, and ξ_l being the RMSE of an image computed with the learned hyperparameters (listed in Tab. 2). In (a) σ'_n is fixed to its learned value, while in (b) the lengthscale l is assigned its learned value. The abscissa axis of the lengthscale plot (a) is the relative difference between the tested lengthscale and the learned value lengthscale l_l.

Let l_l and σ'_{nl} be the learned hyperparameters listed in Tab. 2. To generate the RMSE plots of Fig. 3(a), we have fixed the value of σ'_n to σ'_{nl} and computed the RMSE for different values of l. We have repeated this process for the plot of Fig. 3(b) by varying the value of σ'_n while keeping $l = l_l$. Fig. 3(a) shows that the RMSE is minimum around $\Delta l = l_l - l = 0$ for the three scenes, which means that the learned lengthscale value l_l is appropriate for integration. Furthermore, Fig. 3(b) shows that we can draw the same conclusion for σ'_{nl}. Note that the RMSE is more sensitive to negative than to positive deviations from the optimal value. A more detailed analysis of this interesting feature will be presented in Section 5.4.

5.3 Comparison with importance sampling

Fig. 4 illustrates the results obtained with LDIS and BMC for the three scenes with a varying number of samples. The rendering times of both methods are similar as BMC with a single BRDF entails a negligible overhead compared to LDIS. The close-up views clearly demonstrate that BMC achieves a reduction of high frequency noise as compared to LDIS for the same number of samples. The RMSE is also reduced.

Fig. 5 shows the RMSE plots as a function of the number of samples for MCIS, LDIS, LDBMC and BMC. We can observe that the RMSE of BMC is consistently lower than that of the other methods. This result confirms the advantage of BMC compared to LDIS observed in Fig. 4. Moreover, it is interesting to note that the slope of the straight line fit of the RMSE plot as a function of the

number of samples n is slightly but consistently steeper for the BMC method ($n^{-0.72}$) than for LDIS ($n^{-0.68}$). Recall that the theoretical optimal rate of convergence for Quasi-Monte Carlo (QMC) integration over the unit sphere in $\mathbb{R}^3$ is of order $n^{-0.75}$ [12] under minimal integrand smoothness assumption. LDIS is below the $n^{-0.75}$ rate because this optimal rate assumes a hypothesis on the smoothness of the integrand which is generally not fulfilled by the incident radiance function. In BMC, the embedded Bayesian regression smooths out the discontinuities before integration, which explains its better performances with respect to the convergence rate. Numerical simulations based on the variance expression given by Eq. (10) show that this rate can even be better than $n^{-0.75}$ for lower noise ratio, that is when the GP prior fits well the incident radiance function.

Note also in Fig. 5 that the results obtained with LDBMC, when low discrepancy sequences are used instead of optimized point sets, do not show significant improvements compared to standard importance sampling except for small point set sizes. This result shows the importance of optimized point sets for BMC. It is only through an appropriate samples distribution that the information brought by the prior covariance function can be efficiently exploited.

5.4 Skipping the learning step

A detailed analysis of the sensitivity of the method suggests that BMC can still achieve good performances using approximated values for l and σ'_n, which would allow skipping the learning step. Indeed, Fig. 3(b) shows that for our test cases, we can use any value of σ'_n in the interval $[0.4, 1.1]$ without incurring a significant RMSE increase. However, we have observed that when increasing σ'_n, the error tends to increase faster in the areas of high variance of the incident radiance and is thus more conspicuous. We have found experimentally that $\sigma'_n = 0.5$ is a good trade-off.

Fig. 3(a) shows that the estimation of the integral value is not very sensitive to positive variations of l. Moreover, Fig. 6 demonstrates that the learned values of lengthscale exhibit a dependence on the shininess parameter m, such that $l_l \approx \alpha/\sqrt{m} = \alpha \times w$ where $m = 1/w^2$ as defined in Section 4.1.

A linear regression on the learned lengthscale values yields $\alpha \approx 1.25$ (see Fig. 6). To test this fast hyperparameter derivation method, we have compared the performances of BMC without hyperparameters learning and LDIS for three different scenes (VW, Room and Plates). The results are shown in Figs. 1, 7 and 8 and demonstrate that BMC still clearly outperforms LDIS both in terms of RMSE and visual quality. We have found that the slope of the line fit on the resulting RMSE plot is steeper for BMC, similarly to what we have obtained with learned hyperparameters in Fig. 5.

Fig. 8 shows images computed with the LDIS (Fig. 8(a)) and BMC (Fig. 8(b)) methods. The scene is made up of four plates, each one having a different shininess coefficient. It contains seven light sources of variable size and variable radiance producing an incident radiance along the plates

| Ref | LDIS | BMC | Ref | LDIS | BMC | Ref | LDIS | BMC |

(RMSE$_{LDIS}$ = 0.015 and RMSE$_{BMC}$ = 0.011) (RMSE$_{LDIS}$ = 0.011 and RMSE$_{BMC}$ = 0.008) (RMSE$_{LDIS}$ = 0.012 and RMSE$_{BMC}$ = 0.009)
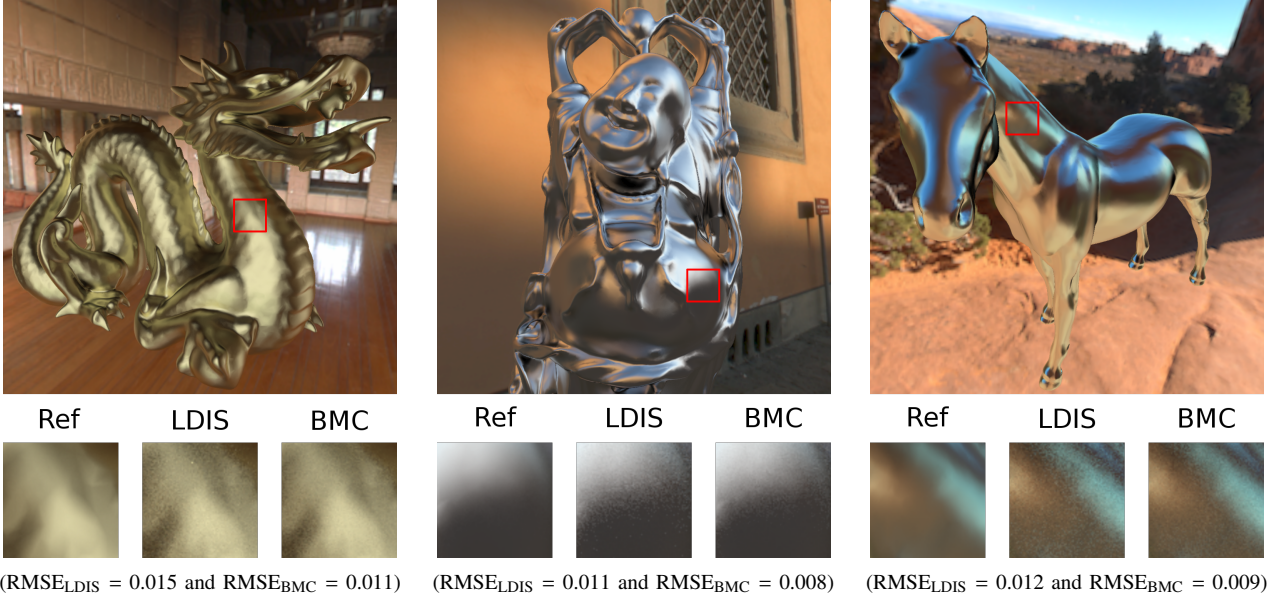
Fig. 4. Close-up views for three scenes rendered with Bayesian Monte Carlo (BMC) and low-discrepancy importance sampling (LDIS). The number of ray samples per visible point is 40 for the Dragon and the Buddha, and 20 for the Horse. Each of these objects has a purely glossy Phong BRDF with a shininess parameter of 20, 50 and 80 respectively from left to right.
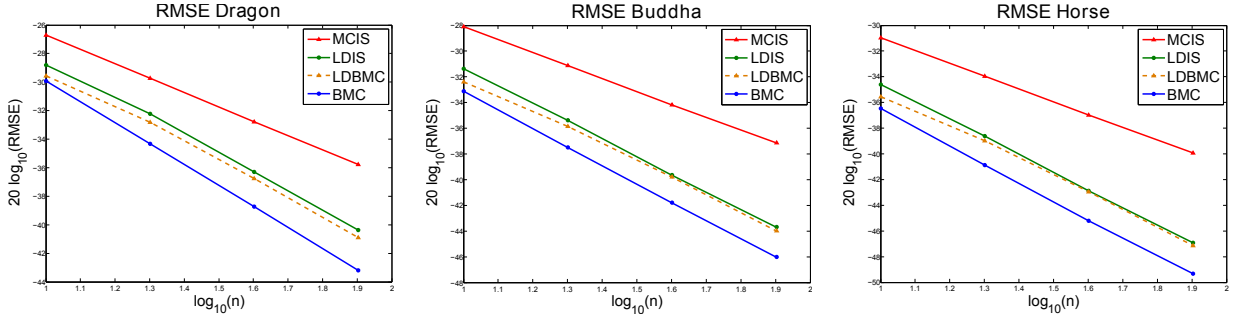
Fig. 5. RMSE plots as a function of the number of samples n. Note that the slope of the line fit for BMC ($n^{-0.72}$) is steeper than for LDIS ($n^{-0.68}$), LDBMC ($n^{-0.67}$), and for MCIS ($n^{-0.50}$).
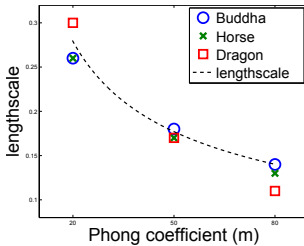

Fig. 6. Fitting of the learned lengthscales as a function of the shininess parameter m.

of variable frequency. Note that this frequency increases as we approach the center of the plates. The objective is to show, through this scene, how our BMC method behaves for different frequencies of the incident radiance and different shininess coefficients. Both the BMC and LDIS methods exhibit the same general features: for the same illumination conditions the noise increases as m decreases, and for the same shininess coefficient m the noise is higher for sharp

variations of the incident radiance. We can notice that the BMC method provides results with a lower noise when compared to the LDIS method. This can be explained by the fact that in BMC the GP acts as a low pass filter. The color transitions also appear smoother in the BMC images than in the LDIS images.

6 DISCUSSION

6.1 Possible Improvements

The results we have obtained demonstrate the soundness of our approach. However, the implementation of BMC presented in this paper is far from exploiting the full potential of the Bayesian approach. Indeed the main advantage of BMC over other methods lies in its ability to incorporate the prior knowledge, which can be obtained in two complementary ways. First by specifying a more accurate mean function for the prior GP with, for example, a rough approximation of incoming radiance using spherical harmonics or a SGF mixture (to this regard, there are

(a) Reference

(b) LDIS
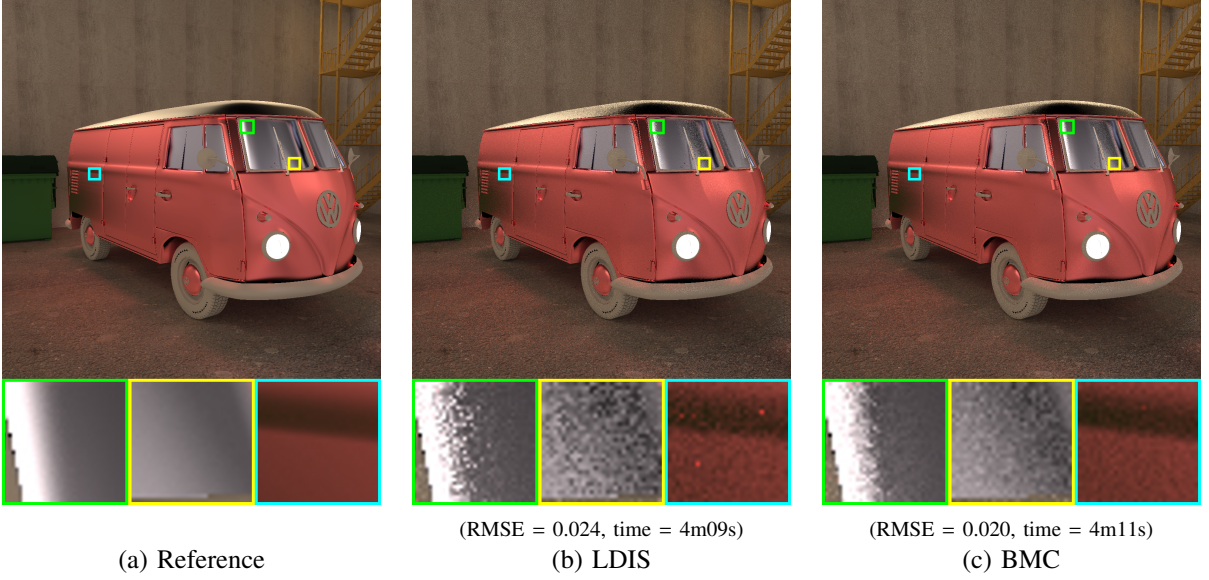(RMSE = 0.024, time = 4m09s)

(c) BMC
(RMSE = 0.020, time = 4m11s)

Fig. 7. Indirect radiance component for the VW scene rendered with LDIS (b) and BMC (c). The shown images have been multiplied by a factor of 4. The RMSE and the time are computed by considering only the glossy component. 16 ray samples per visible point were used for the materials with a glossy BRDF (VW glass, VW bodywork and VW roof), while 64 samples were used for the diffuse BRDFs.

(a) LDIS (RMSE = 0.096)
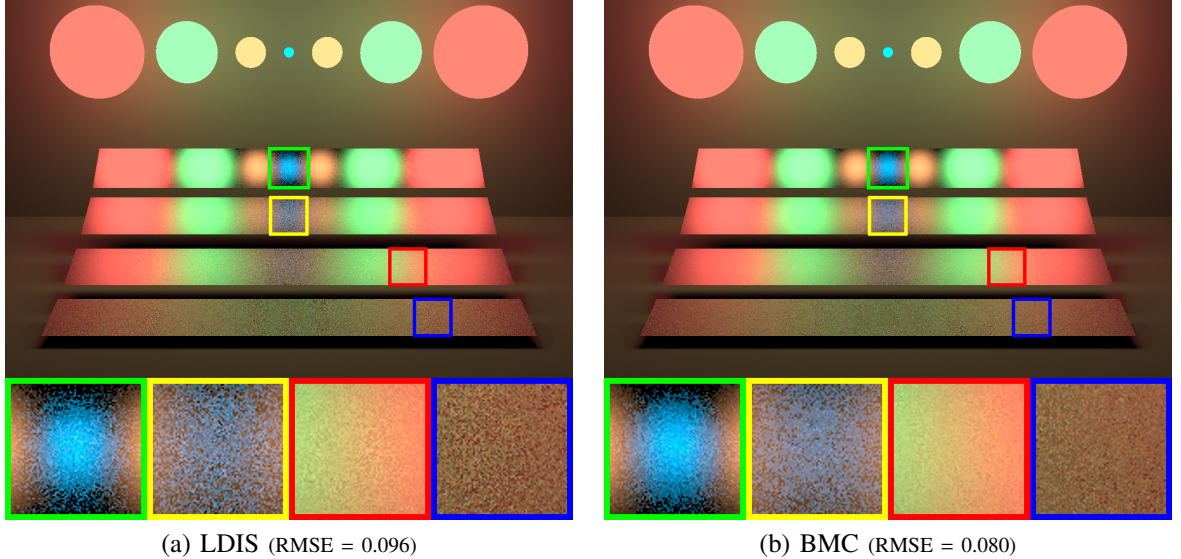
(b) BMC (RMSE = 0.080)

Fig. 8. Direct radiance component for the Plates scene rendered with LDIS (a) and BMC (b). The RMSE has been computed considering only objects with a glossy BRDF (i.e., the four plates), for which 32 ray samples per visible point were used. The Phong shininess coefficient m of each of the plates is 10, 50, 80 and 200 from bottom up respectively.

some similarities with the control covariate method [13]). Second by improving the covariance function adaptation. In particular, the hyperparameters can be adapted locally to better fit the local radiance function characteristics. For example, different classes of pixels can be distinguished based on the samples variance. In this way, different hyperparameters can be chosen for each class. Indeed, the pixels for which the samples variance is high have generally more impact on the visual quality and should be assigned proper hyperparameters value whereas the hyperparameters choice is much less critical for low variance pixels. As these latter pixels represent the great majority, this explains the low

RMSE sensitivity to hyperparameters variation observed in Fig. 3. Furthermore, noise ratios are expected to decrease with local hyperparameter adaptation and consequently higher convergence rate can be obtained as explained in Section 5.3.

6.2 BMC with many complex BRDFs

In this paper, we have focused our attention to single spherical Gaussian BRDF since it is the basic component in our approach to illumination integral computation. In Section 4.1 we have given some indications on how to generalize our approach to any type of BRDF, either analytical

or measured, but further research work is necessary in order to develop an efficient implementation. In the following, we try to be more specific on the different possible approaches. If a BRDF is approximated by a weighted sum of n SGFs, it is easy to see that the illumination integral can be broken down into a weighted sum of n integrals of the same type as Eq. (16). Our method can then be applied separately to each term of the sum and each integral will be assigned a samples set of size proportional to its weight.

However, this straightforward method cannot be optimal in terms of sampling efficiency since each integral estimate will not benefit from the information brought by the samples used in the other integrals. An alternative to this method is then to use a single samples set for all the BRDF lobes. In this case, only one global covariance matrix is necessary for computing the integral estimate with Eq. (14). In Eq. (15), since the function $p(\mathbf{x})$ that represents the BRDF is expressed as a weighted sum of SGFs $\sum_j \alpha_j G_j()$, the $\mathbf{z}'$ vector in Eq. (14) will also be expressed as a weighted sum of vectors $\sum_j \alpha_j \mathbf{z}'_j$ and each $\mathbf{z}'_j$ can be computed with the method described in Section 4.1. As for samples set optimization, the morphing function can be built from an approximation of the BRDF leading to a simpler analytical expression.

As mentioned above, one of the main problems to solve in BMC lies in the computation of the $\mathbf{z}'$ vector and the complexity of this computation depends on the choice of the deterministic function $p(\mathbf{x})$ in Eq. (1). So far, we have considered in a rigid manner that this function corresponds exactly to the BRDF but actually this choice can be made more flexible since $p(\mathbf{x})$ does not need to be an accurate approximation of the BRDF. If a function $g(\mathbf{x})$ (e.g. a weighted sum of SGFs) is a quite raw approximation of $p(\mathbf{x})$, it is possible to consider that the unknown function is $f(\mathbf{x})p(\mathbf{x})/g(\mathbf{x})$ and substitute $g(\mathbf{x})$ to $p(\mathbf{x})$ in Eq. (15). This will not change very much the behavior of the unknown function as the factor $p(\mathbf{x})/g(\mathbf{x})$ is generally a very smooth function but it could greatly simplify the computation of the $\mathbf{z}'$ vector for complex BRDFs.

As regards the application of BMC for scenes with a large number of different BRDFs, the single problem that could arise is the need for precomputing an optimized samples set for each shininess coefficient. This operation is much lighter than learning the hyperparameters and is scene independent. But still, if one wants to avoid performing this operation for each different shininess coefficient, the same strategy as the one applied for the hyperparameters could be used: optimize the sample sets for a subset of shininess values (which results in a polynomial of degree 3 as described in section 4.3) and interpolate the values of these polynomials for the intermediate values.

6.3 Limitations

As explained in this paper, the main difficulties that we are faced with in implementing BMC lie in the computation of the inverted covariance matrix and the $\mathbf{z}'$ vector. Our proposal in this paper is to solve this problem by precomputing optimized points sets and using spherical Gaussian

mixtures. As far as this solution is applicable, there is no reason why BMC should not perform better that CMC methods.

Directly applying our framework to the case of a very large samples set would increase the memory footprint and make costly the preprocessing step because of the inversion of the covariance matrix and the samples set optimization. One possible solution is to resort to a progressive approach which would refine the integral estimate by using successive small sample sets. At each iteration, the prior GP for the current estimate (for the current samples set) would use the posterior GP of the previous estimate (for the previous samples set). In this way the estimate is progressively refined.

The extension of BMC to larger dimensionality, such as illumination integral calculations requiring image plane sampling, lens sampling and multiple reflections, is not obvious. Indeed, finding an appropriate prior model (definition of the known and uncertain functions, determination of the hyperparameters and the mean function) while keeping a good trade-off between efficiency and performance is a difficult task.

7 Conclusion

In this paper, important steps are given towards the goal of reaching a general Bayesian Monte Carlo approach for solving global illumination problems, which was until now limited to the diffuse BRDF case. Modeling the BRDF and the covariance function with the same family of spherical Gaussian functions is a key point, since it allows making the computation of the Bayesian quadrature coefficients tractable for glossy BRDFs. The results confirm that Bayesian Monte Carlo outperforms the classical Monte Carlo methods. In a context where rendering times are dominated by the cost of sampling, Bayesian Monte Carlo has a clear advantage over low discrepancy Monte Carlo importance sampling since it exhibits less noise and lower RMSE for the same number of samples, while having a negligible overhead. Finally, we demonstrated that Bayesian Monte Carlo outperforms low discrepancy Monte Carlo importance sampling even with a suboptimal hyperparameters setting that avoids the learning step. Future work will aim at improving the efficiency of Bayesian Monte Carlo as explained above and also at extending its use to other problems of larger dimensionality, e.g., measured BRDFs, multiple reflections, depth blur, multi-view rendering, among others.

REFERENCES

[1] A. O'Hagan, "Bayes-hermite quadrature," *J. Statist. Plann. Inference*, vol. 29, no. 3, pp. 245–260, 1991.

[2] J. Brouillat, C. Bouville, B. Loos, C. Hansen, and K. Bouatouch, "A bayesian monte carlo approach to global illumination," *Computer Graphics Forum*, vol. 28, no. 8, pp. 2315–2329, 2009. [Online]. Available: http://dx.doi.org/10.1111/j.1467-8659.2009.01537.x

[3] C. E. Rasmussen and Z. Ghahramani, "Bayesian monte carlo," in *Neural Information Processing Systems*. MIT Press, 2002, pp. 489–496.

[4] T. Pfingsten, D. J. L. Herrmann, and C. E. Rasmussen, "Model-based design analysis and yield optimization," *Semiconductor Manufacturing, IEEE Trans. on*, vol. 19, no. 4, pp. 475–486, 2006.

[5] A. Geiger, "Gaussian process for machine learning," 2007, http://www.rainsoft.de/projects/gausspro.html. [Online]. Available: http://www.rainsoft.de/projects/gausspro.html

[6] C. E. Rasmussen and C. K. I. Williams, *Gaussian Process for Machine Learning*. MIT Press, 2006.

[7] J. Křivánek and M. Colbert, "Real-time Shading with Filtered Importance Sampling," *Computer Graphics Forum*, vol. 27, no. 4, pp. 1147–1154, 2008.

[8] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

[9] J. Wang, P. Ren, M. Gong, J. Snyder, and B. Guo, "All-frequency rendering of dynamic, spatially-varying reflectance," *ACM Trans. Graph.*, vol. 28, pp. 133:1–133:10, December 2009. [Online]. Available: http://doi.acm.org/10.1145/1618452.1618479

[10] E. Saff and A. Kuijlaars, "Distributing many points on a sphere," *Math. Intelligencer*, vol. 19, no. 1, pp. 5–11, 1997.

[11] W. Jakob, "Mitsuba renderer," 2010, http://www.mitsuba-renderer.org.

[12] J. S. Brauchart and J. Dick, "Quasi-monte carlo rules for numerical integration over the unit sphere $\mathbb{S}^2$," *ArXiv e-prints*, Jan. 2011, http://arxiv.org/abs/1101.5450.

[13] E. Lafortune and Y. Willems, "A 5D tree to reduce the variance of monte carlo ray tracing," in *Rendering Techniques '95 (Proceedings of the Sixth Eurographics Workshop on Rendering)*. Springer-Verlag, 1995, pp. 11–20.

Mickaël Ribardière received his Master's degree in computer science in 2007 from the University of Limoges, and a PhD degree in 2010 from the University of Rennes 1. He is currently Research Scientist in the FRVSense team at IRISA (Institut de Recherche en Informatique et Systémes Aléatoires). His research interests are global illumination, lighting simulation for complex environment and computer vision.

Luís Paulo Santos is an Assistant Professor at the Universidade do Minho, Portugal. He published several papers on both computer graphics and parallel processing on international conferences and journals. He manages two nationally funded graphics R&D projects and participates in multiple European projects with both academia and industry. His main research interests lie on Interactive Global Illumination and Parallel Processing. He is a member of the Direction Board of the Portuguese chapter of Eurographics since 2008, was Deputy Director of the University of Minho's Dept. of Informatics from 2010 to 2012 and Director of the Doctoral Program on Informatics over the same period.

Ricardo Marques received his MSc degree in Computer Graphics and Distributed Parallel Computation from Universidade do Minho (fall 2009), after which he worked as a researcher in the same university. He joined INRIA (Institut National de Recherche en Informatique et Automatique) and the FRVSense team as a PhD student in the fall 2010 under the supervision of Kadi Bouatouch. His research work focuses on spherical integration methods applied to light transport simulation.

Kadi Bouatouch was an electronics and automatic systems engineer (ENSEM 1974). He was awarded a PhD in 1977 and a higher doctorate in computer science in the field of computer graphics in 1989. He is working on global illumination, lighting simulation for complex environments, GPU based rendering and computer vision. He is currently Professor at the University of Rennes 1 (France) and researcher at IRISA (Institut de Recherche en Informatique et Systémes Aléatoires). He is member of Eurographics.

Christian Bouville is presently invited researcher in the FRVSense team at IRISA in Rennes (France). He has been team leader, project leader and Emeritus expert at Orange Labs until 2006 and has been involved in many European and national projects. His main fields of research are now global illumination models and image-based rendering with a special interest in machine learning approaches.