An Inexpensive BDRF Model for Physically-Based Rendering

Christophe Schlick

LaBRI [1]
351, cours de la Libération
33405 Talence (FRANCE)
schlick@labri.greco-prog.fr

Abstract : A new BRDF model is presented which can be viewed as an kind of intermediary model between empirism and theory. Main results of physics are observed (energy conservation, reciprocity rule, microfacet theory) and numerous phenomena involved in light reflection are accounted for, in a physically plausible way (incoherent and coherent reflection, spectrum modifications, anisotropy, self-shadowing, multiple reflection, surface and subsurface reflection, differences between homogeneous and heterogeneous materials). The model has been especially intended for computer graphics applications and therefore includes two main features : *simplicity* (a small number of intuitively understandable parameters controls the model) and *efficiency* (the formulation insures adequation to Monte-Carlo rendering techniques and/or hardware implementations).

Keywords : Physically-Based Rendering, Bidirectional Reflectance Distribution Function, Optimization

1 Introduction

Computation of a reflectance model is the heart of every rendering method because it provides the illumination of objects in the scene, and therefore the color of pixels in the image. Reflectance models currently in use can be divided in two main families : empirical models and theorical ones. Empirical models [12, 3, 6] are computationally efficient but are lacking of physical validity (energy conservation law, for instance) and thus do not provide plausible values of energy or intensity. In fact, they are generally only used to create bright spots on surfaces in order to add some tridimensional information which helps to understand the image. Therefore they are limited to applications where good-looking pictures are sufficient (computer generated imagery for movies or commercials). On the other side, theoretical models [5, 9, 20] involve higher computational costs but provide quantitative values that have shown to be in good adequation with experimental data. Therefore they are well adapted to applications for which physically-based rendering is essential (simulation for lighting industry or architecture).

This paper proposes a kind of intermediary model between empirical and theoretical models. In Section 2, simple notations are presented and used to reformulate several existing reflectance models. Section 3 focuses on some unsatisfactory points that can be find in these models. In Section 4, a general purpose optimization technique is detailled and several low-cost alternatives to expensive terms involved in existing models are proposed. Finally, Section 5 presents the new reflectance model, which uses that optimization technique to combine several interesting features of previously disjointed work into an inexpensive formulation well-suited to computer graphics.

2 Background

The interaction of light with a surface is described by relating incoming and outcoming radiances at a given point P on the surface. This expression usually involves a function $R_\lambda(P, V, V')$ called *bidirectional reflectance distribution function* (BRDF, for short) :

$$L_\lambda(P, V) = \int_{V' \in \mathcal{V}} R_\lambda(P, V, V')\ L_\lambda(P, -V')\ (N \cdot V')\ dV' \tag{1}$$

[1] Laboratoire Bordelais de Recherche en Informatique (*Université Bordeaux I* and *Centre National de la Recherche Scientifique*). The present work is also granted by the *Conseil Régional d'Aquitaine*.

where

- $L_\lambda(P,V)$ is the reflected radiance leaving point P in direction V
- $L_\lambda(P,-V')$ is the incident radiance reaching point P from direction $-V'$
- $R_\lambda(P,V,V')$ is the BRDF of the surface at point P between directions V and V'
- $\mathcal{V}$ is the set of possible directions for incident light (*ie* hemisphere above the surface)
- dV' is a differential solid angle surrounding direction V'
- V, V' and N are unit vectors so $(N \cdot V')$ is the cosine of the angle between N and V'

The reflected radiance is the integral, for all possible directions, of incident radiances scaled by the BRDF and the projected solid angle. It should be noted that such a formulation is well adapted to rendering algorithms (radiosity, path tracing, two pass methods) that perform effectively a true integration and for which the solid angle information is available. Other rendering algorithms (direct illumination methods, ray tracing) consider only a discrete sum of light contributions, and therefore are unable to provide close simulation of real phenomena as required by physically-based rendering.

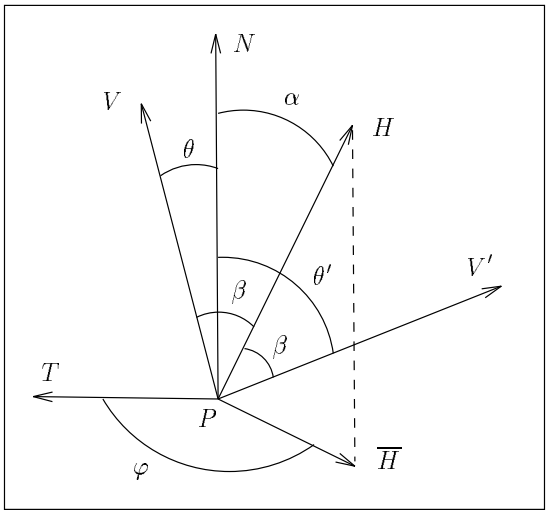
Equation 1 is a monochromatic equation expressed for a given wavelength λ. In the present paper, we use the following notation convention : every term that is function of the wavelength will be subscripted by λ. Such a term has to be defined and/or computed, theoretically for every wavelength of the visible spectrum, and practically for a given number of samples (three in trichromatic models, up to twenty in spectral models [6]).

The BRDF has got two important properties that result directly from physics of light [2]. First, due to the *Helmholtz Reciprocity Rule*, R_λ is symmetric relative to V and V' :

$$\forall\ V \in \mathcal{V} \quad \forall\ V' \in \mathcal{V} \quad R_\lambda(P,V,V') = R_\lambda(P,V',V) \tag{2}$$

Second, due to the *Energy Conservation Law*, R_λ has to fulfill the normalization condition :

$$\forall\ V \in \mathcal{V} \quad \int_{V' \in \mathcal{V}} R_\lambda(P,V,V')\ (N \cdot V')\ dV' \leq 1 \tag{3}$$



V	=	Outcoming direction of light
V'	=	Incoming direction of light
N	=	Surface normal vector
T	=	Surface tangent vector
H	=	Bisector vector of V and V'
$\overline{H}$	=	Projection of $H - N$

α	$= < H, N >$	t	$= \cos\alpha$
β	$= < V, H >$	u	$= \cos\beta$
θ	$= < V, N >$	v	$= \cos\theta$
θ'	$= < V', N >$	v'	$= \cos\theta'$
φ	$= < \overline{H}, N >$	w	$= \cos\varphi$

Figure 1 : Angles and vectors for definition of BDRF

According to the shape of the BRDF, two kinds of surfaces are traditionally distinguished :

Diffuse surfaces : The light is reflected in every direction. The limit case — *perfectly diffuse surfaces* or *Lambert surfaces* — is obtained when the BRDF becomes a constant function (*ie* the light is equally reflected in every direction).

Specular surfaces : The light is reflected only in a small area around the mirror direction. The limit case — *perfectly specular surfaces* or *Fresnel surfaces* — is obtained when the BRDF becomes a Dirac function (*ie* the light is reflected in a single direction).

Figure 1 presents the notations that will be used to formulate BRDF models throughout the paper. A complete review of all the models that have been proposed in the litterature is out of scope here (the interested reader may refer to [17] where such a survey has been done), we will focus especially on the three theoretical models that have been used to define our new model.

2.1 The Cook-Torrance Model

The first theoretical reflectance model has been introduced in the computer graphics field by Cook & Torrance [5] using work previously done in physics by Beckmann & Spizzichino [2] and Torrance & Sparrow [19] about the reflection of electromagnetic waves on rough surfaces. In that model, a surface is supposed to be composed of so-called *microfacets* which are small smooth planar elements. Only microfacets whose normal vector is in the direction H (see Figure 1) contribute to the reflection between V and V'. The BDRF depends on five different angles and is expressed as a linear combination of a diffuse reflector and a specular one :

$$R_\lambda(\alpha,\beta,\theta,\theta',\varphi) = \frac{d}{\pi} C_\lambda + \frac{s}{4\pi v v'} F_\lambda(\beta)\ G(\theta,\theta')\ D(\alpha,\varphi) \quad \text{with} \quad d+s=1 \tag{4}$$

where

- d (resp. s) $\in [0,1]$ is the ratio of the surface behaving as a diffuse (resp. specular) reflector
- $C_\lambda \in [0,1]$ is the ratio of light at wavelength λ, reflected by the diffuse reflector (a complete set of C_λ, one for each wavelength sample, enables to define the color of the diffuse reflector)
- $D(\alpha,\varphi) \in [0,\infty)$ is the microfacets slope distribution function which defines the fraction of the facets that are oriented in direction H
- $F_\lambda(\beta) \in [0,1]$ is the Fresnel factor which defines the ratio of light at wavelength λ, reflected by each microfacet (a complete set of F_λ gives the color of the specular reflector)
- $G(\theta,\theta') \in [0,1]$ is the geometrical attenuation coefficient which expresses the ratio of light that is not self-obstructed by the surface

One important point to notice is that Equation 4 is only valid when it fulfills the normalization condition (Equation 3). Therefore, it implies a condition on the slope distribution function [2] :

$$\int_0^{\pi/2} \int_0^{2\pi} D(\alpha,\varphi)\ \cos\alpha\ \sin\alpha\ d\alpha\ d\varphi = \pi \tag{5}$$

If we assume an isotropic behaviour of the surface (*ie* the BDRF is invariant by rotation around the normal vector) then Equation 5 becomes :

$$\int_0^{\pi/2} D(\alpha)\ 2\ \cos\alpha\ \sin\alpha\ d\alpha = 1 \tag{6}$$

Moreover, if we express the angular dependence of the three factors (D, F and G) in terms of their cosines, Equation 6 may be rewritten as :

$$\int_0^1 2t\ D(t)\ dt = 1 \tag{7}$$

and finally Equation 5 becomes :

$$R_\lambda(t,u,v,v') = \frac{d}{\pi} D_\lambda + \frac{s}{4\pi v v'} F_\lambda(u)\ G(v,v')\ D(t) \tag{8}$$

2.2 The He-Torrance-Sillion-Greenberg Model

A more complete model for BRDF has been proposed by He *et al.* [9] which accounts comprehensively for every physical phenomena (polarization, diffraction, interference, conductivity, effective and apparent roughness...) involved in light reflection on rough surfaces. When restricted to unpolarized light (the usual case in computer graphics), it can be rewritten in a form that mimics Equation 8 :

$$R_\lambda(t,u,v,v') = \frac{d}{\pi} C_\lambda + \frac{s}{4\pi v v'} F_\lambda(u)\ G(v,v')\ D(t) + \frac{s}{v'\ dV'} F_\lambda(v)\ G(v,v')\ A(v,v')\ \Delta \tag{9}$$

Such a formulation hides the complexity of the model (for instance, the expression of the slope distribution function $D(t)$ has discouraged many potential users to implement it) but lets clearly appear one of the main difference with the Cook-Torrance model : there is an additional term in the linear combination which corresponds to coherent reflection on the mean plane of the surface (*ie* not the microfacets). Coherent reflection has been used for years in computer graphics (it is the fundamental principle of recursive ray tracing) but only for perfectly specular surfaces; He *et al.* have shown that this term exists also for non-smooth surfaces, though it decreases rapidly when the roughness increases, due to the presence of the roughness attenuation coefficient $A(v,v')$.

2.3 The Ward Model

Relatively few reflectance models have been proposed for anisotropic reflection (*ie* the BRDF is function of angle φ and thus of its cosine w) and usually they involve high computational costs [10, 4, 13]. Ward has presented a simple model [20] in which the rotational symmetry of isotropic BRDF is replaced by an elliptical asymmetry of varying excentricity :

$$R_\lambda(t,v,v',w) = \frac{d}{\pi} C_\lambda + \frac{s}{4\pi\sqrt{vv'}} C'_\lambda\ D(t,w) \quad \text{with} \quad D(t,w) = \frac{1}{mn}\ e^{\frac{t^2-1}{t^2}\left(\frac{w^2}{m^2} + \frac{1-w^2}{n^2}\right)} \tag{10}$$

Intuitively, the model considers scratches on the surface (oriented along the tangent vector of the local frame) leading to different roughnesses (defined by m and n) when considering directions parallel or perpendicular to the scratches : the more m and n are different, the more anisotropy is created.

3 Unsatisfactory Points

By examinating existing reflectance models, one can find several points that appear somewhat unsatisfactory. For instance, the BRDF is formulated as a linear combination with constant weights between a diffuse part and a specular one. The justification usually given by the authors is that, for a large class of materials, diffuse and specular components come from different physical phenomena, and thus they may have different colors. One classical example is a plastic surface on which light can be reflected either by the uncolored substrat in a coherent way (*ie* surface reflection is specular) or by the colored pigments beneath the surface in an incoherent way (*ie* subsurface reflection is diffuse) [5].

But, as noticed by Shirley [15], such a linear combination with constant weights is incorrect because proportions of diffuse and specular components are usually not constant but function of the incident angle. Taking the example of a varnished wood floor (see Figure 2), one can see that, according to the Fresnel law, for large incident angles most light is reflected specularly by the varnish, whereas for small incident angles, most light penetrates the varnish before beeing reflected diffusely by the wood.

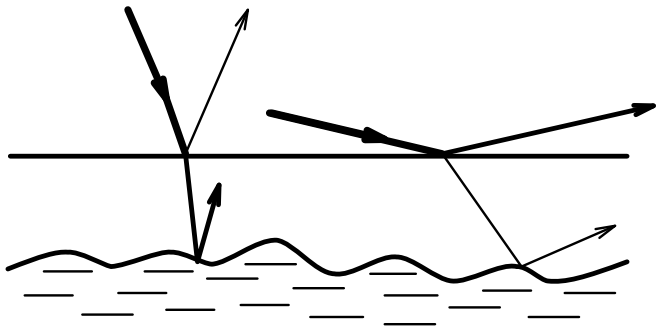


Figure 2 : Influence of the incident angle on surface and subsurface reflection

Beside these *heterogeneous* materials, there are a lot of *homogeneous* ones, for which the diffuse/specular distinction is unnecessary. For such materials (metals, for instance) there is rather a kind of continuum between perfect diffuse and perfect specular behaviours (see Figure 3) according to the roughness of the surface. Therefore a linear combination with constant weights is inadequate again.

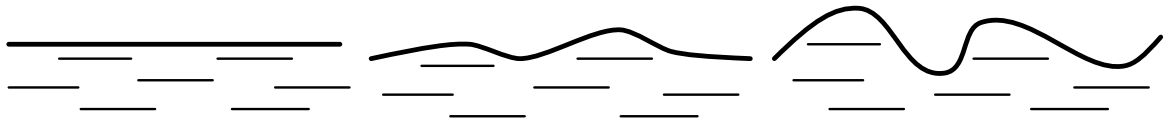


Figure 3 : Continuum between diffuse and specular for surface reflection

Another weak point in existing models appears when light reaches or leaves a rough surface where self-obstruction (from one microfacet to another) occurs. Usually, a geometrical attenuation coefficient (G in Equation 8) is used as a multiplicative factor to express the ratio of light that is not subject to that obstruction. But in real life, the remainder of the light (*ie* $1-G$) is reflected in other directions and not simply blocked. Currently, none of the existing reflectance models does correctly account for that reemission of self-obstructed light.

The last — and perhaps the most — unsatisfactory point is about the accuracy/cost ratio. As said, empirical reflectance models are inexpensive but their lack of physical validity prevents their use in any physically-based rendering system. On the other hand, theoretical models are physically accurate but involve complex mathematical expressions which are computationally expensive and preclude hardware implementations. Moreover, when including such a reflectance model in an image synthesis software, the error generated by other stages of the rendering pipeline (tessellation for geometrical modeling, spectral sampling for optical modeling, directional sampling for global illumination, interpolation at almost every steps) does usually totally cancel the benefit of greater accuracy. In other words, there is no need to compute BDRF at a precision of 0.1%, if directional sampling is only done at 5% and spectral sampling at 15%. A possible solution could be to replace expensive formulas in theoretical reflectance models by some well-chosen low-cost alternative functions. In Section 4, we present a new technique which enables to find such approximations.

4 Optimization by Rational Fraction Approximation

One classical optimization technique (which has been applied several times in computer graphics) to speed up an algorithm that involves the computation of a complex function is to store many sample

values of the function in a table, and compute missing values by interpolation (usually linear or cubic). Implementing a whole theoretical reflectance model which such a technique would require numerous tables (in order to account for various surface properties and illuminating conditions) which means high memory costs and difficulties to switch to hardware implementations.

A well-known technique in mathematics is to replace such a function by its Taylor expansion, giving a polynomial that can be computed with a handful of multiplications and additions by Horner's rule. Unfortunately a Taylor expansion is only valid nearby the origin ; therefore the approximation is usually only accurate in such a neighbourhood. To overcome this limitation, a possible solution is to use piece-wise Taylor approximants. But creating large ranges of values where the approximation is accurate implies to use many pieces, for which continuity in their derivatives cannot always be insured.

Another classical technique (which exists since the beginning of the century and has been applied to numerous scientific fields) is to use Padé approximants [1] in which a rational fraction is generated according to the Taylor expansion of the function. Compared to pure polynomial approximations, Padé approximants have usually a much better accuracy when leaving the neighbourhood of the origin. Piece-wise Padé approximants have also been proposed but rarely used in practice because insuring continuity of the derivatives becomes almost impossible.

The previous approximation methods, which deal all with Taylor expansions, have got two strong limitations. First, the Taylor expansion of the function has to be known, this is not always possible even with numerical techniques. Second, specific properties of the function are generally not preserved. For instance, if we want to approximate a statistical distribution function f on a range $[a, b]$ (which by definition has to fulfill $\int_a^b f(t)dt = 1$) by one of the previous approximation techniques, the approximant has virtually no chance to fulfill that condition too, giving something that is mathematically incorrect.

4.1 Principle

We propose here another method that we simply call *rational fraction approximation*. This method differs from the Padé approximation technique by the fact that we do not use Taylor expansions to find the coefficients of the numerator and denominator polynomials. The idea is to study the function that we want to approximate, in order to find what we call *kernel conditions*.

A kernel condition can be any intrinsic characteristic of the function : value at a given point either of the function or of one of its derivative, integral or differential equation it obeys to... The detection and the choice of kernel properties can be done in several ways, by using its mathematical definition, by picking some of its remarkable values or even by plotting the function and examinating the graph. In fact, the only generic method for finding kernel conditions is to answer the question : "For me what are the characteristics of the function that every approximation should fulfill ?"

Once the kernel conditions have been found, coefficients of the rational fraction are simply obtained by identifying the function and its approximation for each kernel condition. This gives a system of n equations and n unknowns where n is the number of conditions.

For instance, let's suppose that we want to approximate the function $f(x) = \sin x$ on the range $[0, \pi/2]$. By plotting the corresponding curve or by specific knowledge on the function, we can find at least four characteristics that appear to be essential for every approximation function :

$$f(0) = 0 \quad f'(0) = 1 \quad f(\pi/2) = 1 \quad f'(\pi/2) = 0$$

Because there are four kernel conditions, we search for a rational fraction containing four independent parameters (a, b, c and d). For instance :

$$\forall x \in [0, \pi/2] \quad \overline{f}(x) = \frac{x^2 + ax + b}{cx + d} \tag{11}$$

When we express the kernel conditions with that function[2] , it leads to a system of four equations :

$$\begin{cases} b = 0 & \text{(condition 1)} \\ a = d & \text{(condition 2)} \\ a + \pi/2 = c + 2a/\pi & \text{(condition 3)} \\ c\pi^2 + 4\pi a + 4a^2 = 0 & \text{(condition 4)} \end{cases}$$

giving finally the following approximation of $f(x) = \sin x$:

$$\forall x \in [0, \pi/2] \quad \overline{f}(x) = x\frac{u+x}{u+vx} \quad \text{with} \quad u = -\pi^2/4 \quad \text{and} \quad v = \pi + u \tag{12}$$

Plotting the original function and the approximation (see Figure 4) enables to control visually the similarities between the curves. In order to test our approximations in a quantitative way, we have also developped a statistical test (evaluation of the function and its approximation for one million random values) which provides two measures : ε = relative error and γ = speed-up factor (see [16] for more complete testing results). For instance, with our sine example, we have obtained $\varepsilon = 1.4\%$ and $\gamma = 230\%$

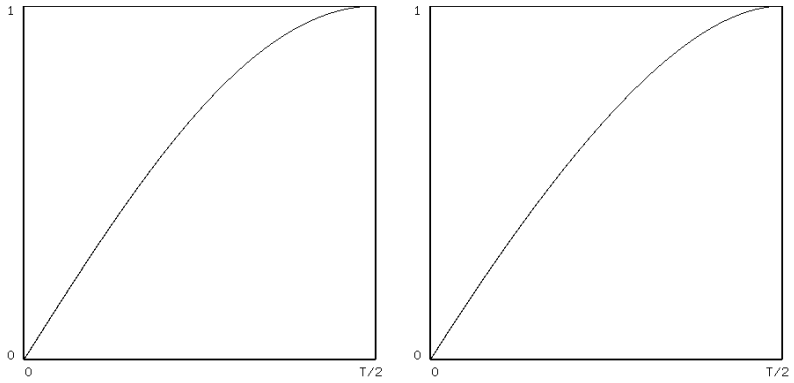


Figure 4 : $f(x) = \sin x$ where $x \in [0, \pi/2]$
LEFT : True function RIGHT : Rational approximation

4.2 Approximation of the Fresnel Factor

The Fresnel factor $F_\lambda(u)$ expresses the reflection of light on the well-oriented microfacets (*ie* the facets whose normal vector is H). For a non-polarized electromagnetic wave, its formulation is [15] :

$$\begin{aligned} F_\lambda(u) &= \frac{1}{2}\frac{(a-u)^2+b^2}{(a+u)^2+b^2}\left[\frac{(a+u-1/u)^2+b^2}{(a-u+1/u)^2+b^2}+1\right] \\ a^2 &= \frac{1}{2}\left(\sqrt{(n_\lambda^2-k_\lambda^2+u^2-1)^2+4n_\lambda^2k_\lambda^2}+n_\lambda^2-k_\lambda^2+u^2-1\right) \\ b^2 &= \frac{1}{2}\left(\sqrt{(n_\lambda^2-k_\lambda^2+u^2-1)^2+4n_\lambda^2k_\lambda^2}-n_\lambda^2+k_\lambda^2-u^2+1\right) \end{aligned} \tag{13}$$

where n_λ is the ratio of the refraction indices above and below the surface and k_λ is the extinction coefficient of the surface. An interesting characteristic of the Fresnel factor is that $F_\lambda = 1$ at a grazing incidence ($\beta = \pi/2$ so $u = 0$) whatever the wavelength λ.

[2]Notice that the fraction $(ax+b)/(cx+d)$ does not have four independent parameters, because one of them can be eliminated by simultaneous division of the numerator and the denominator.

One difficulty that precludes a general use of F_λ in every rendering environment comes from the fact that n_λ and k_λ are seldom known. Some experimental values exist [11] but usually one can only find a single value $\overline{n}$ and $\overline{k}$ for a wavelength in the middle of the visible spectrum. On the other hand, a data which has been measured for thousands of materials is the spectral distribution f_λ of the Fresnel factor at normal incidence ($\beta = 0$ so $u = 1$). When $\overline{n}$, $\overline{k}$ and f_λ are the only known data, Cook & Torrance have proposed the following approximation [5] :

- Compute $\overline{F}(u)$ and $\overline{f} = \overline{F}(1)$ with $\overline{n}$ and $\overline{k}$, using Equation 20
- For each wavelength λ, compute $F_\lambda(u)$, using Equation 21

$$F_\lambda(u) = f_\lambda + (1 - f_\lambda)\,\frac{\overline{F}(u) - \overline{f}}{1 - \overline{f}} \tag{14}$$

This approximation not only solves the lack of experimental data, but also speeds-up the calculation because the complete expression of $F_\lambda(u)$ is evaluated only once, for an average n_λ and k_λ. But even so, the computation of the Fresnel factor remains expensive and further optimization should be possible.

By examinating Figure 5, one can see that the shape of the curves does not vary very much according to the kind of material. The main difference is the value f_λ where the curve arrives at $u = 1$. Therefore, a step further in the approximation process could be to make $F_\lambda(u)$ only dependent on f_λ. By choosing the following kernel conditions :

$$F_\lambda(0) = 1 \qquad F_\lambda(1) = f_\lambda \qquad F'_\lambda(1) = 0 \qquad F''_\lambda(1) = 0$$

we have found

$$F_\lambda(u) = f_\lambda + (1 - f_\lambda)\,(1 - u)^5 \tag{15}$$

which costs only 4 multiplications and 2 additions in an optimized implementation. Our statistical testing process shows that the approximation can be computed almost 32 times faster with less than 1% error : $\varepsilon = 0.6\%$ and $\gamma = 3180\%$

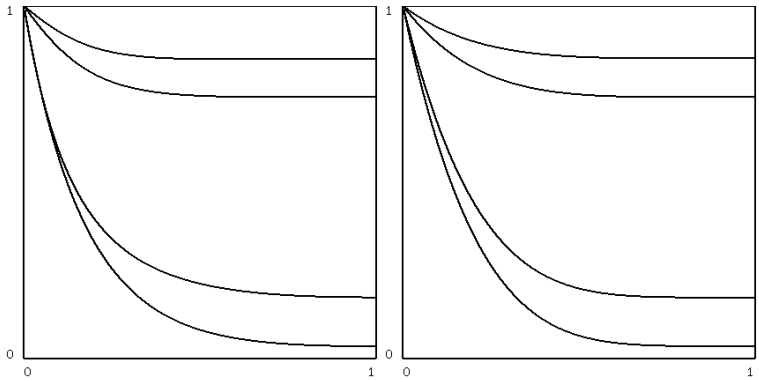


Figure 5 : $F_\lambda(s)$ by Fresnel for diamant, glass, copper and gold
LEFT : True function RIGHT : Rational approximation

4.3 Approximation of Geometrical Attenuation Coefficient

Several expressions for the geometrical attenuation coefficient $G(v, v')$ have been proposed in physics [19, 18, 14]. In their original paper, Cook & Torrance used the formulation derived in [19] :

$$G(t, u, v, v') = \min\left[1, 2\frac{tv}{u}, 2\frac{tv'}{u}\right] \tag{16}$$

This expression results from coarse approximations about the surface geometry and therefore does not meet experimental results : its first derivative is not continuous, it is not invariant by rotation around the normal vector and it is independent of the surface roughness. The formulation proposed by Smith [18] (introduced in the computer graphics field by [9]) is not subjected to these restrictions and has been experimentally validated. Moreover, it depends only on v and v', and is separable in v and v' :

$$G(v, v') = G(v)\ G(v') \tag{17}$$

After several equivalences, the original expression of $G(v)$ can be written more compactly :

$$G(v) = \frac{g}{g+1} \quad \text{with} \quad g = \sqrt{h\pi}\ (2 - \text{erfc}\ \sqrt{h}) \quad \text{and} \quad h = \frac{v^2}{2m^2\ (1 - v^2)} \tag{18}$$

where m is the root mean square (RMS) slope of the microfacets (theoretically $m \in [0, \infty)$ but in fact, it almost never exceeds 0.5 for real surfaces). Despite its complicated form, the shape of the function (see Figure 6) is quite simple. To characterize it, we choose the following kernel conditions :

$$G(0) = 0 \qquad G(1) = 1 \qquad G'(0) = \sqrt{\frac{\pi}{2m^2}}$$

that leads to a very simple expression :

$$G(v) = \frac{v}{v - kv + k} \quad \text{with} \quad k = \sqrt{\frac{2m^2}{\pi}} \tag{19}$$

By precomputing k and $1-k$, $G(v)$ needs only 1 division, 1 multiplication and 1 addition. Except for a small neighbourhood of $v = 1$, the two curves on Figure 6 are very similar. That visual feeling is confirmed by the testing process : $\varepsilon = 1.8\%$ and $\gamma = 2870\%$

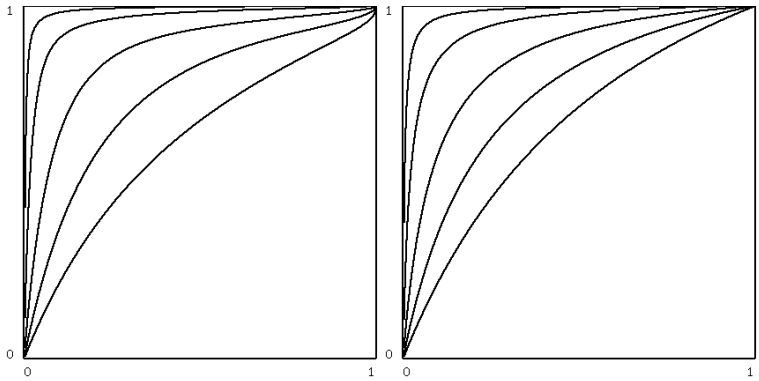


Figure 6 : $G(v)$ by Smith for $m = 0.01, 0.05, 0.1, 0.25, 0.5$
LEFT : True function RIGHT : Rational approximation

4.4 Approximation of Slope Distribution Function

Among the different formulations of the slope distribution function that exist in the litterature [2, 19, 3], only the one proposed by Beckmann fulfills Equation 7. Moreover, compared to others, this formulation depends only on the RMS slope m of the microfacets and does not introduce any arbitrary constant :

$$D(t) = \frac{1}{m^2 t^4}\ e^{\frac{t^2-1}{m^2 t^2}} \tag{20}$$

When the surface is rough (large values for m), orientations of microfacets are very dispersed. When the surface is smoother (small values for m) microfacet normals H come closer to the average normal N. And for perfectly smooth surfaces (m is null) $D(t)$ becomes a Dirac function (see Figure 7).

The normalization condition (Equation 7) is an obvious kernel condition for $D(t)$. Another important characteristic (already noticed by Beckmann) is that $D(t)$ is almost null for $t < 1-m$. And finally, $D(1)$ gives a last kernel condition :

$$\forall t \in [0, 1-m] \quad D(t) = 0 \qquad D(1) = \frac{1}{m^2} \qquad \int_0^1 D(t)\, 2t\, dt = 1$$

Having an integral equation as a kernel condition complicates somewhat the process of finding a good approximation. A solution that works well for many cases is to search a rational fraction having a u'/u^2 form, for which an analytic integration can be done. Therefore, we propose :

$$\forall t \in [1-m, 1] \quad D(t) = \frac{m^3 x}{t\ (m x^2 - x^2 + m^2)^2} \quad \text{with} \quad x = t + m - 1 \tag{21}$$

An optimized implementation of $D(t)$ costs only 1 division, 4 multiplications and 2 additions. The curves drawn in polar coordinates are shown on Figure 7. The quantitative results are somewhat less good than for the two other approximations but still interesting : $\varepsilon = 2.7\%$ and $\gamma = 1650\%$

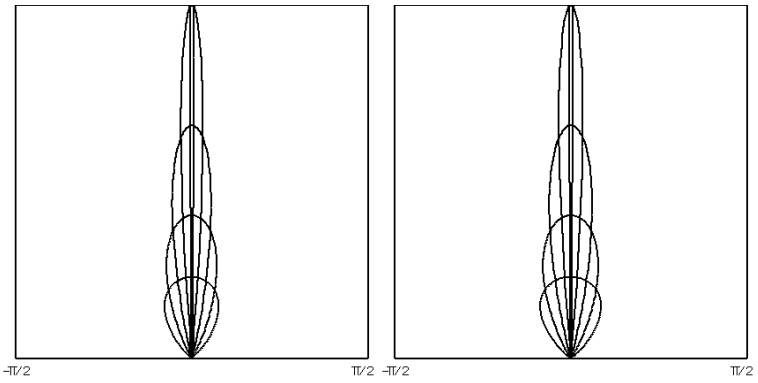


Figure 7 : $D(t)$ by Beckmann for $m = 0.01, 0.05, 0.1, 0.25, 0.5$
LEFT : True function RIGHT : Rational approximation

When collecting the results of Equations 15, 19 and 21, one gets a kind of approximated/optimized Cook-Torrance model (more than 20 times faster with less than 3% error[3] compared to an implementation using the true formulas). Because it uses only basic arithmetic operations $(+ - */)$ the approximated model is well-suited to hardware implementation and constitutes a possible answer to the accuracy/cost ratio inadequacy of current BRDF models (see Section 3).

5 A New BDRF Model

The rational fraction approximation scheme enables to speed-up the computation of reflectance models but does not provide a solution for the other unsatisfactory points discussed in Section 3. With regards to that discussion, an appealing BRDF model should include the following features :

- Main results of physics *(Energy Conservation Law, Helmholtz Reciprocity Rule, Fresnel Equation, Microfacet Theory)* should be fulfilled to enable physically-based rendering
- A continuum between lambertian and smooth surfaces should be provided

[3] In fact, it is much more precise than the original Cook-Torrance implementation in which Equation 16 is used for G, giving an error of $\varepsilon = 53\%$ with our statistical test.

- A distinction between homogeneous and heterogeneous materials should be made
- Both isotropic and anisotropic behaviours should be accounted for
- Only a small number of simple and meaningful parameters should control the model
- Only expressions with low computational costs should be used

A new model which includes all these features is presented here and can be viewed as an intermediary model between empirism and theory. Before giving the formulation of the BRDF, let's examine how we modelize optical properties of surfaces. In fact, two different kinds of materials SINGLE/DOUBLE are distinguished in opposition to the classical diffuse/specular separation[4] :

- SINGLE : Materials having homogeneous optical properties (metal, glass, rough paper, tissue)
- DOUBLE : Materials having heterogeneous optical properties (plastic, skin, stratified or varnished or painted surfaces) usually composed of a transparent layer over an opaque one, each of them being SINGLE materials.

We propose to characterize a SINGLE material by a set of parameters (C_λ, r, p) and a DOUBLE material by two sets (C_λ, r, p) and (C'_λ, r', p'), one for each layer :

- $C_\lambda \in [0,1]$: Reflection factor at wavelength λ
- $r \in [0,1]$: Roughness factor ($r = 0$: perfect specular, $r = 1$: perfect diffuse)
- $p \in [0,1]$: Isotropy factor ($p = 0$: perfect anisotropy, $p = 1$: perfect isotropy)

The choice of these parameters was motivated mainly by two of there characteristics. First, the role of every parameter can be understood intuitively and therefore easily defined by a non-physician user. Second, the parameters can also be assigned according to experimental data [11]. Indeed, C_λ can be viewed as the reflectivity at normal incidence f_λ (see Equation 14), r can be related to the RMS slope m of the surface (see Equation 20), and p is in fact the ratio of the RMS slopes m/n between the scratch ($\varphi = 0$) and the ortho-scratch ($\varphi = \pi/2$) directions for an anisotropic surface (see Equation 10).

5.1 Formulation

When only geometrical optics is involved (an hypothesis made by almost every rendering technique), the spectral and the directional behaviours of the BDRF can be separated (*ie* rays are reflected in the same direction, whatever their wavelength) in two multiplicative factors S_λ and D. According to the kind of material, we propose the following expression :

$$\begin{cases} \text{SINGLE} & : R_\lambda(t,u,v,v',w) = S_\lambda(u)\ D(t,v,v',w) \\ \text{DOUBLE} & : R_\lambda(t,u,v,v',w) = S_\lambda(u)\ D(t,v,v',w) + [1 - S_\lambda(u)]\ S'_\lambda(u)\ D'(t,v,v',w) \end{cases} \tag{22}$$

Spectral factor

The simplest expression for the spectral factor is to consider it as a constant function :

$$S_\lambda(u) = C_\lambda \tag{23}$$

But in fact, $S_\lambda(u)$ depends on the incident angle and should obey to the (approximated) Fresnel law :

$$S_\lambda(u) = C_\lambda + (1 - C_\lambda)\ (1-u)^5 \tag{24}$$

[4] This idea of a layered surface model has appeared several times in physics (but has never been completely invastigated) as well as in a very recent paper by Hanrahan & Krueger in computer graphics [8].

Directional factor

Following a similar derivation as in [2], one can imagine a straigthforward formulation for the directional factor (extended to anisotropy), in which the dependence on the zenith angle α and on the azimuth angle φ can be separated and expressed by two factors $Z(t)$ and $A(w)$:

$$D(t, v, v', w) = \frac{1}{4\pi v v'} \; Z(t) \; A(w) \tag{25}$$

$D(t, v, v', w)$ in only valid when the product $Z(t)A(w)$ obeys to Equation 5. If we suppose an identical anisotropic behaviour in all four quadrants around the normal vector, Equation 5 can be rewritten as :

$$\int_0^{\pi/2} Z(\alpha) \; \cos\alpha \; \sin\alpha \; d\alpha = \frac{1}{2} \qquad \text{and} \qquad \int_0^{2\pi} A(\varphi) \; d\varphi = 4 \int_0^{\pi/2} A(\varphi) \; d\varphi = 2\pi \tag{26}$$

and thus

$$\int_0^1 2t \; Z(t) \; dt = 1 \qquad \text{and} \qquad \int_0^1 \frac{1}{\sqrt{1-w^2}} \; A(w) \; dw = \frac{\pi}{2} \tag{27}$$

Using Equation 27 as kernel conditions, as well as other characteristics, we have found simple expressions for Z and A, which are in fact polar equations of ellipses (the former with the pole on the boundary, the latter with the pole in the middle) :

$$Z(t) = \frac{r}{(1 + rt^2 - t^2)^2} \qquad \text{and} \qquad A(w) = \sqrt{\frac{p}{p^2 - p^2 w^2 + w^2}} \tag{28}$$

When looking at the resulting curves (see Figure 8), one can notice that when $r = 1$, $Z(t)$ is a constant function (perfect diffuse) and when $r = 0$, $Z(t)$ becomes a Dirac function (perfect specular). The same remark can be made for $A(w)$ which varies continuously between a constant function when $p = 1$ (perfect isotropy) and a Dirac function when $p = 0$ (perfect anisotropy).

Adequation to Monte-Carlo Techniques

One main feature of the expressions for $Z(t)$ and $A(w)$ is that they are well-suited to Monte-Carlo rendering methods [15, 20]. Indeed, a usual technique for a Monte-Carlo process is to generate a stochastic *importance sampling* to improve the convergence [7]. Such an importance sampling for α and φ (limited to the first octant but easily generalized by duplication) can be simply obtained from two uniform random variables $(a, b) \in [0, 1]^2$ with :

$$\alpha = \operatorname{acos} \sqrt{\frac{a}{r - ar + a}} \qquad \text{and} \qquad \varphi = \frac{\pi}{2} \sqrt{\frac{p^2 b^2}{1 - b^2 + b^2 p^2}} \tag{29}$$

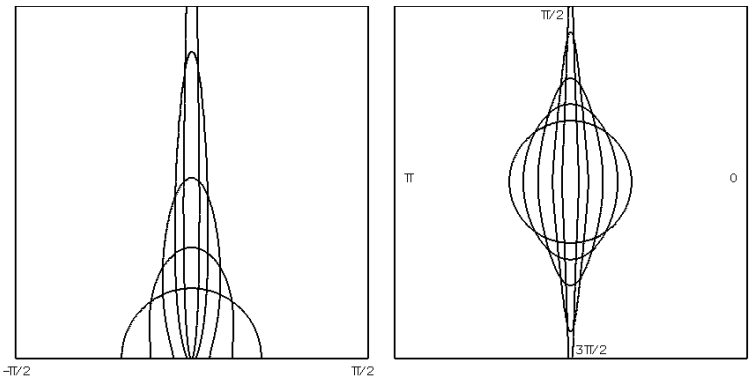


Figure 8 : Directional factor in logarithmic polar coordinates
LEFT : Zenith angle dependence $Z(t)$ for $r = 0.01, 0.05, 0.2, 0.5, 1.0$
RIGHT : Azimuth angle dependence $A(w)$ for $p = 0.01, 0.05, 0.2, 0.5, 1.0$

As in [18], self-obstruction can be included by a geometrical attenuation factor $G(v)G(v')$ where $G(v)$ (resp. $G(v')$) expresses the ratio of reflected (resp. incident) non-obstructed light. But, as discussed in Section 3, we want to account for reemission of self-obstructed light (*ie* 1–G). Because of the stochastic orientation of microfacets, the direction of light after several reflections is essentially random, therefore it seems logical to reemit the light with $Z(t) = A(w) = 1$:

$$D(t, v, v', w) = \frac{G(v)G(v')}{4\pi v v'} \; Z(t) \; A(w) \quad + \quad \frac{1 - G(v)G(v')}{4\pi v v'} \tag{30}$$

As with the Fresnel factor, the approximated formulation of the Smith factor is used :

$$G(v) = \frac{v}{r - rv + v} \qquad \text{and} \qquad G(v') = \frac{v'}{r - rv' + v'} \tag{31}$$

Diffuse-Specular Continuum

Due to the presence of v and v' on the denominator, Equation 25 does not provide a complete transition from perfect diffuse to perfect specular. A classical result in physics [2, 8] says that there exists no microfacets configuration (even a constant slope distribution function) that provides a Lambert reflector. Similarly, because coherent reflection is not considered, Equation 25 cannot provide a Fresnel reflector (even with a Dirac slope distribution function). Again, we propose a kind of empirical/theoretical solution, inspired by Equation 9, to create a continuum between these limit situations. The directional factor is defined as a sum of three different reflectors (Lambert, Beckmann, Fresnel), each of them having a specific weight (l, b, f) :

$$D(t, v, v', w) = \frac{l}{\pi} \; + \; \frac{b}{4\pi v v'} \; B(t, v, v', w) \; + \; \frac{f}{v' \; dV'} \; \Delta \quad \text{with} \quad l + b + f = 1 \tag{32}$$

where $B(t, v, v', w)$ is the anisotropic Beckmann-like factor given either by Equation 25 or 29 and Δ is a Dirac fonction (equal to 1 in dV' and 0 otherwise). The weights (l, b, f) could be specified directly by the user, but it would represent three additional parameters per material. We propose rather a (physically plausible) automatic generation scheme which provides a quadratic interpolation between the three fundamental behaviours of a surface, according to the roughness factor.

$$\text{if} \;\; (r < 0.5) \;\; \text{then} \;\; \{b = 4r(1-r) \; ; \; l = 0 \; ; \; f = 1-b\} \;\; \text{else} \;\; \{b = 4r(1-r) \; ; \; f = 0 \; ; \; l = 1-b\} \tag{33}$$

5.2 Pictures

In order to show various illumination effects (incidence angles ranging from grazing to normal and varying either fast or slow) a simple test scene similar to [9] has been chosen. Every cylinder on the pictures has been rendered individually at a 256x512 resolution using Monte-Carlo ray-tracing. Picture 1 (resp. Picture 2) illustrates the continuum that is provided between diffuse (resp. isotropic) and specular (resp. anisotropic) reflection by taking four different values for r (resp. p) from 1 to 0.05. To achieve a better understanding of the behaviour of the new model, only direct illumination from a single light source put at the view point is shown. In order to exhibit anisotropy, the cylinder is made of brushed metal, having concentric circular scratches on its top and parallel horizontal scratches on its body. Picture 3 is similar to Picture 2, but indirect specular illumination is shown this time.

Picture 4 presents a set of disks on which circular scratches have created a characteristic butterfly pattern due to anisotropy. The influence of the light source position is shown : on the top row, the source is rotated around the disk and one can see that the pattern "follows" the light, one the bottom row, the source is translated from almost grazing to almost normal incidence and one can see that the pattern grows up and finally fits the whole disk. Such effects appear very familiar because brushed metal disks are common objects in our life (e.g. reverse side of a watch).

Finally, Picture 5 illustrates the importance of the Fresnel factor in S_λ. The cylindres are made here of a DOUBLE material, having a transparent specular layer on an opaque (black for the left cylinders, white for the right ones) diffuse layer. When using Equation 23 for S_λ (first and third cylinder), the whole cylinder exhibits the same amount diffuse and specular behaviours (linear combination with constant weights) which is both visually unnatural and physically incorrect. When using Equation 24 (second and fourth cylinders), different reflection behaviours are provided according to the incident angle as predicted by theory and measured experimentally [9].

Aknowledgments

Special thanks to G.Ward, K.Bouatouch and F.Sillion for their numerous comments during the development and the specification of the new model.

References

[1] G.A.Baker, P.G.Morris, *Padé approximants*, Encyclopedia of Mathematics, v13-14, Addison Wesley, 1981

[2] P.Beckmann, A.Spizzichino, *Scattering of Electromagnetic from Rough Surfaces*, Pergamon Press, 1963

[3] J.Blinn, *Models of Light Reflection for Computer Synthesized Pictures*, SIGGRAPH 77, Computer Graphics, v11, n4, p192-198, 1977

[4] B.Cabral, N.Max, R.Springmeyer, *Bidirectional Reflection Functions from Surface Bump Maps*, Proc. SIGGRAPH 87, Computer Graphics, v21, n4, p273-281, 1987

[5] R.Cook, K.Torrance, *A Reflectance Model for Computer Graphics*, Proc. SIGGRAPH 81, Computer Graphics, v15, n4, p307-316, 1981

[6] R.Hall, *Illumination and Color in Computer Generated Imagery*, Springer-Verlag, 1989

[7] J.M.Hammerley, D.C.Handscomb, *Monte-Carlo methods*, Wiley Editors, 1964

[8] P.Hanrahan, W.Krueger, *Reflection from Layered Surfaces due to Subsurface Scattering*, SIGGRAPH 93, Computer Graphics, v27, n3, p159-169, 1993

[9] X.He, K.Torrance, F.Sillion, D.Greenberg, *A Comprehensive Physical Model for Light Reflection*, SIGGRAPH 91, Computer Graphics, v25, n4, p187-196, 1991

[10] J.T.Kajiya, *Anisotropic Reflection Models*, SIGGRAPH 85, Computer Graphics, v19, n4, p15-22, 1985

[11] E.D.Palik, *Handbook of Optical Constants of Solids*, Academic Press, 1985

[12] B.T.Phong, *Illumination for Computer Generated Pictures*, Comm. of the ACM, v18, n6, p449-455, 1975

[13] P.Poulin, A.Fournier, *A Model for Anisotropic Reflection*, SIGGRAPH 90, Computer Graphics, v24, n4, p273-282, 1990

[14] M.Sancer, *Shadow-Corrected Electromagnetic Scattering from a Randomly Rough Surface*, IEEE Transactions on Antennas and Propagation, v17, n5, p577-585, 1969

[15] P.Shirley, *Physically Based Lighting Calculations for Computer Graphics*, PhD thesis, University of Illinois at Urbana-Champaign, November 1990

[16] C.Schlick, *Divers Eléments pour une Synthèse d'Images Réalistes*, PhD Thesis, Univ. Bordeaux I, 1992

[17] C.Schlick, *A Survey of Shading and Reflectance Models for Computer Graphics*, Technical Report 93/116, LaBRI, Université Bordeaux I (also submitted to Computer Graphics Forum), 1993

[18] B.G.Smith, *Geometrical Shadowing of a Random Rough Surface*, IEEE Transactions on Antennas and Propagation, v15, n5, p668-671, 1967

[19] K.E.Torrance, E.M.Sparrow, *Theory for Off-Specular Reflection from Roughened Surfaces*, Journal of the Optical Society of America, v57, n9, p1105-1114, 1967

[20] G.J.Ward, *Measuring and Modeling Anisotropic Reflection*, SIGGRAPH 92, Computer Graphics, v26, n2, p265-272, 1992

Picture 1 : Continuum between diffuse and specular reflection (direct illumination)

Picture 2 : Continuum between isotropic and anisotropic reflection (direct illumination)

Picture 3 : Continuum between isotropic and anisotropic reflection (indirect illumination)

Picture 4 : Influence of the light position on anisotropic reflection

Picture 5 : Influence of the Fresnel factor on DOUBLE materials