

Constant-Time Energy-Normalization for the Phong Specular BRDFs

Ian Mallett and Cem Yuksel

Abstract The Phong and Modified Phong specular BRDFs, although of limited physical basis, are nevertheless some of the simplest BRDFs exhibiting glossy and specular qualities to understand and to implement, making them useful for validation and teaching. Unfortunately, although it is well-known how to make these BRDFs *conserve* energy (that is, never gain energy), making them *energy-normalized* (that is, never lose nor gain energy) is far more difficult. Lesser-known algorithms exist, but require the specular exponent n to be integer-valued, and have $O(n)$ runtime cost. We express these algorithms as mathematical formulae and generalize to the real-valued specular exponent case. We then simplify and optimize to finally attain an algorithm that is $O(1)$. Energy normalization makes the Phong BRDFs more physically plausible and therefore both more practically and theoretically useful—and our improvements allow for this energy normalization to be done efficiently and without arbitrary limitations.

1 Introduction

The importance of a reflectance model, i.e. the bidirectional reflectance distribution function (BRDF), to be *energy-conserving* is well-understood in computer graphics. An energy-conserving BRDF ensures that the material does not reflect more light than it receives—that is, the material is permitted to absorb some of the light it receives, but it cannot add light.

In the absence of any absorption, real materials still exhibit some darkening in their appearance based on the view angle. When light hits a surface, the Fresnel equations define the portion that is reflected or transmitted. Collectively, the Fresnel coefficient, multiplied by a *distribution function* that describes the distribution of reflectance, gives the BRDF. For a material to be physically realistic, the distribution function must integrate to one—that is, any darkening in the BRDF should be due to the Fresnel term, not due to artifacts of the distribution; when there is no absorption, the distribution function must scatter *all* light it receives. We refer to this criterion as *energy normalization*. Note that energy normalization is a stronger criterion than energy conservation, and is arguably part of the definition of the BRDF itself [12].

Energy normalization has received less attention in computer graphics than energy conservation, as its solution is usually more difficult. In production (such as AAA video games, special effects, and feature animation), microfacet BRDFs [2, 15] are common. Energy normalization for a microfacet BRDF can be physically interpreted, mostly, as accounting for multiple scattering off of secondary microfacets, and convenient solutions have only become available recently [5, 7, 9, 16].

In this paper, we concern ourselves with calculating the correct energy normalization terms for the specular component BRDFs of the Phong [14] and Modified Phong [10] reflectance models. These BRDFs are still widely implemented and used and, even if microfacet models are more popular, Phong specular BRDFs remain valuable for their simplicity and intuition. These BRDFs are not physically based, but energy normalization makes them behave in a physically sensible way, which may also make them more appealing in a contest with more-modern BRDF models. Nonetheless, we view the contribution of our work as more-theoretical

Ian Mallett
University of Utah School of Computing
imallett@cs.utah.edu

Cem Yuksel
University of Utah School of Computing
cem@cemyuksel.com

in nature, rather than necessarily having practical importance for realistic graphics productions.

Arvo [1] introduced a powerful framework which can be applied for expressing BRDFs, including (approximately) modern BRDFs [8]. However, it provides an *exact* solution for the Phong specular BRDFs. Unfortunately, this algorithm only works for integer values of the specular exponent n and it has linear computational complexity $O(n)$. Considering that the normalization factor must be computed every time the BRDF is evaluated and n can be arbitrarily (even infinitely) large, this algorithm quickly becomes impractical for shiny materials.

We introduce an algorithm for exact energy normalization of the Phong specular BRDFs, which can instead be computed in constant time and space. Our formulation also supports non-integral specular exponent values. To achieve this, we simply convert Arvo's method (as it is expressed in [4]) into a mathematical language, apply continuation and subsequently various simplifications to the mathematical form, and then interpret the result back as an algorithm. The factors introduced in this paper can be considered a practical "fix" for the Phong specular BRDFs.

2 Background

When light hits a surface point \mathbf{x} , a part of it is transmitted (absorbed, for "opaque" materials) and the rest is reflected. Let ϵ be the ratio of reflectance (i.e. $1 - \epsilon$ accounts for transmission) and L_i be the incoming radiance from a particular direction ω_i . Thus, ϵ portion of the light received by the surface $(\mathbf{N} \cdot \omega_i) L_i$ is reflected toward outgoing directions ω_o , where \mathbf{N} is the surface normal. The incoming light multiplied by ϵ must be equal to the integral over the whole hemisphere for outgoing illumination, such that

$$(\mathbf{N} \cdot \omega_i) L_i \epsilon = \int_{\Omega} (\mathbf{N} \cdot \omega_i) L_i f_r(\omega_i \rightarrow \mathbf{x} \rightarrow \omega_o) (\mathbf{N} \cdot \omega_o) d\omega_o.$$

By canceling out the terms $(\mathbf{N} \cdot \omega_i) L_i$ on either side, we can write:

$$\epsilon = \int_{\Omega} f_r(\omega_i \rightarrow \mathbf{x} \rightarrow \omega_o) (\mathbf{N} \cdot \omega_o) d\omega_o. \quad (1)$$

A realistic material cannot reflect more light than it receives, meaning it conserves energy, such that $\epsilon \leq 1$. If a material has no transmission/absorption, we must get exactly $\epsilon = 1$.

2.1 The Phong Specular Component BRDFs

The Phong reflectance model [14] proposes a reflection model that is a combination of a diffuse and a specular

term, controlled by albedo parameters ρ_d and ρ_s , respectively. Later, the rendering equation [6] formalized the geometry term $\mathbf{N} \cdot \omega_i$. We recognize Phong's diffuse term as a Lambertian diffuse BRDF times the geometry term from the rendering equation, but his specular term lacks the geometry term, and thereby would implicitly have a divided term in the BRDF. Dropping this detail gives the Modified Phong BRDF [10]. In modern conventions, the Phong and Modified Phong BRDFs therefore respectively take the forms:

$$f_{r,P}(\omega_i \rightarrow \mathbf{x} \rightarrow \omega_o) = \frac{\rho_d}{\pi} + \frac{\rho_s}{I_P} \cdot \frac{\max(0, \mathbf{R}(\omega_i) \cdot \omega_o)^n}{(\mathbf{N} \cdot \omega_i)}$$

$$f_{r,M}(\omega_i \rightarrow \mathbf{x} \rightarrow \omega_o) = \frac{\rho_d}{\pi} + \frac{\rho_s}{I_M} \cdot \max(0, \mathbf{R}(\omega_i) \cdot \omega_o)^n,$$

where ω_i and ω_o are the incoming and outgoing light directions, \mathbf{N} is the surface normal, I_P and I_M are the energy normalization terms for the two BRDFs, and $\mathbf{R}(\mathbf{u}) := -\mathbf{u} + 2(\mathbf{N} \cdot \mathbf{u})\mathbf{N}$ is the perfect reflection direction. For these BRDFs to be sensible, the diffuse and specular albedos must sum to no more than 1, i.e. $\rho_d + \rho_s \leq 1$. The specular exponent n is an extended real number whose sensible values range in $[0, \infty]$. In particular, infinite values are allowed, and correspond to perfect specular (i.e. mirror) reflection.

The Phong BRDFs bundle diffuse and specular into what we would now call a multi-layer model¹. The diffuse term is already normalized as a simple Lambertian BRDF; however, the specular component is not. Without loss of generality, we need normalize only the Phong and Modified Phong *specular component* BRDFs, and the distinction will be hereafter dropped:

$$f_{r,P}^{spec}(\omega_i \rightarrow \mathbf{x} \rightarrow \omega_o) = \frac{1}{I_P} \cdot \frac{\max(0, \mathbf{R}(\omega_i) \cdot \omega_o)^n}{(\mathbf{N} \cdot \omega_i)} \quad (2)$$

$$f_{r,M}^{spec}(\omega_i \rightarrow \mathbf{x} \rightarrow \omega_o) = \frac{1}{I_M} \cdot \max(0, \mathbf{R}(\omega_i) \cdot \omega_o)^n \quad (3)$$

Substituting these BRDFs into Equation 1 with $\epsilon = 1$ while applying reciprocity tells us that the values of I_P and I_M must be:

$$I_P = \int_{\Omega} \max(0, \mathbf{R}(\omega_o) \cdot \omega_i)^n d\omega_o \quad (4)$$

$$I_M = \int_{\Omega} \max(0, \mathbf{R}(\omega_o) \cdot \omega_i)^n (\mathbf{N} \cdot \omega_o) d\omega_o \quad (5)$$

It has been widely observed that when $\omega_i = \mathbf{N}$, the integrals are maximal. The equations then simplify easily (see Appendix B), giving energy conservation (but not normalization) terms:

$$I_P = \frac{2\pi}{n+1}, \quad I_M = \frac{2\pi}{n+2} \quad (6)$$

Unfortunately, being only conservation terms, these

¹ As in other multi-layer composite-BRDF models, one would typically weight the specular term by the Fresnel reflection and the diffuse term by the Fresnel transmission.

Algorithm 1 Variant of Arvo’s method, used for the Phong BRDF.

```

procedure ARVOINTEGRATEPHONG( $\mathbf{N}, \omega_i, n$ )
   $c \leftarrow (\mathbf{N} \cdot \omega_i)$ 
   $s \leftarrow \sqrt{1 - c^2}$ 
  if  $n$  even then
     $A \leftarrow \pi - \arccos(c)$ 
     $k \leftarrow 1$ 
     $T_k \leftarrow s$ 
  else
     $A \leftarrow \pi/2$ 
     $k \leftarrow 0$ 
     $T_k \leftarrow \pi/2$ 
  end if
   $t \leftarrow 0$ 
  while  $k \leq n - 1$  do
     $t \leftarrow t + T_k$ 
     $k \leftarrow k + 2$ 
     $T_k \leftarrow s^2 ((k - 1)/k) T_k$ 
  end while
   $I_P \leftarrow 2 (A + c t) / (n + 1)$ 
  return  $I_P$ 
end procedure

```

only normalize the Phong BRDFs at normal incidence: even with $\rho_s = 1$, energy is lost for any $\omega_i \neq \mathbf{N}$, particularly near grazing angles. We must turn to more-advanced techniques to truly normalize Phong for all directions ω_i and ω_o .

2.2 Energy Normalization

Unfortunately, computing the correct energy normalization factors in Equations 4 and 5 is difficult. Intuitively, the specular lobe intersects the ecliptic plane in a complex way when the outgoing direction is not \mathbf{N} . Moreover, when $\omega_o \neq \mathbf{N}$, the geometry term no-longer aligns with the specular lobe, and so must be treated separately.

Arvo [1] developed algorithms that can be applied to this problem. Specifically, I_M can be computed using Algorithm 2 (slightly modified; see Arvo’s §6.4 and Integral 31a in [4]). Similarly, I_P can be computed using a result derived from the same paper (see [4] Integral 31b), which we rework into a similar form in Algorithm 1. These algorithms are not widely implemented or known, but were the first, and to our knowledge heretofore-only, way to get the correct normalization terms for the Phong BRDFs. Notice that apart from the direction of the branch, the loop bound, and the last line, the algorithms are identical.

Unfortunately, these algorithms have two major limitations. The first is that they have $O(n)$ time complexity. This is unacceptable considering that BRDFs are evaluated millions or billions of times in a typical render. The specular exponent n is unbounded, and is

Algorithm 2 Arvo’s method, used for the Modified Phong BRDF.

```

procedure ARVOINTEGRATEMODIFIEDPHONG( $\mathbf{N}, \omega_i, n$ )
   $c \leftarrow (\mathbf{N} \cdot \omega_i)$ 
   $s \leftarrow \sqrt{1 - c^2}$ 
  if  $n$  even then
     $A \leftarrow \pi/2$ 
     $k \leftarrow 0$ 
     $T_k \leftarrow \pi/2$ 
  else
     $A \leftarrow \pi - \arccos(c)$ 
     $k \leftarrow 1$ 
     $T_k \leftarrow s$ 
  end if
   $t \leftarrow 0$ 
  while  $k \leq n - 2$  do
     $t \leftarrow t + T_k$ 
     $k \leftarrow k + 2$ 
     $T_k \leftarrow s^2 ((k - 1)/k) T_k$ 
  end while
   $I_M \leftarrow 2 (T_k + A c + c^2 t) / (n + 2)$ 
  return  $I_M$ 
end procedure

```

in practice quite large (thousands or at-least hundreds would be expected for a shinier material).

The second limitation is that both algorithms are limited to values of n that are nonnegative integers. This adds implementation complexity, but is mainly objectionable because discretization artifacts are likely to be produced if n varies (such as when roughness is textured), especially at the lower values.

Precomputing a 1D table for a particular specular exponent n and a range of cosine values $\mathbf{N} \cdot \omega_i$ allows the BRDF to be evaluated fairly efficiently, albeit due to interpolation, only approximately. For acceptable results at glancing angles, the table must be quite large, incurring excessive precomputation and caching cost. A 2D table, which allows the specular exponent to vary, is even larger; the precomputation cost can be seconds or minutes.

3 Energy Normalization of the Phong BRDFs

Our goal is to convert Arvo’s algorithms into a constant-time method. In this section, we achieve this by first formulating the algorithms mathematically. Various sums and products of this can then be evaluated either in closed-form or in terms of various easy-to-work-with functions, allowing further simplifications and evaluation at non-integral specular exponents. Interpreting the result back as an algorithm yields a constant-time computation.

We begin by providing the following reformulation of both Algorithms 1 and 2 into mathematical terms:

$$c = (\mathbf{N} \cdot \omega_i), \quad s = \sqrt{1 - c^2}, \quad \text{specular exponent } n$$

$$A = \begin{cases} \pi - \arccos(c) & \text{if } n \text{ even (Phong) or} \\ & n \text{ odd (Modified Phong)} \\ \pi/2 & \text{if } n \text{ odd (Phong) or} \\ & n \text{ even (Modified Phong)} \end{cases}$$

$$T_k = \begin{cases} (\pi/2) s^k \prod_{j=1}^{k/2} \left(\frac{2j-1}{2j} \right) & \text{if } k \text{ even} \\ s^k \prod_{j=1}^{(k-1)/2} \left(\frac{2j}{2j+1} \right) & \text{if } k \text{ odd} \end{cases}$$

$$S_k = \begin{cases} \sum_{j=0}^{k/2} T_{2j} & \text{if } k \text{ even} \\ \sum_{j=1}^{(k+1)/2} T_{2j-1} & \text{if } k \text{ odd} \end{cases}$$

$$I_P = 2 (A + c S_{n-1}) / (n+1)$$

$$I_M = 2 (T_n + A c + c^2 S_{n-2}) / (n+2)$$

It is worth noting that when written out mathematically, the algorithm has quadratic complexity: computing T_k is $O(k)$ and so computing S_k is $O(k^2)$.

We now turn to writing these expressions in terms of functions without summations or products. In the following, let n be odd for the Phong BRDF and n be even for the Modified Phong BRDF so that we only need to compute T_k and S_k for k even (it turns out we will get the same result without this assumption; this is just to simplify the presentation).

T_k (again, for k even) can be computed as ^{2 3}:

$$\begin{aligned} T_k &= \frac{\pi}{2} s^k \prod_{j=1}^{k/2} \frac{2j-1}{2j} = \frac{\pi}{2} s^k \frac{(k-1)!!}{2^{k/2} (k/2)!} \\ &= \frac{\pi}{2} s^k \frac{2^{k/2} \Gamma((k+1)/2) / \sqrt{\pi}}{2^{k/2} \Gamma(k/2 + 1)} \\ &= \frac{\sqrt{\pi}}{2} s^k \frac{\Gamma((k+1)/2)}{\Gamma(k/2 + 1)} \end{aligned}$$

This result is resistant to further simplification, but the ratio of gamma functions can be computed efficiently and robustly by taking the exponential of the difference of the logs of the numerator and denominator; the log-of-gamma function is widely available and computationally efficient in most programming environments, both on the CPU and GPU (e.g. `lgamma(...)` in C/C++, CUDA, and OpenCL).

To solve the series S_k , after substituting in the appropriate T term, we break the summation into two infinite series:

$$S_k = \sum_{j=0}^{k/2} T_{2j} = \frac{\sqrt{\pi}}{2} \sum_{j=0}^{k/2} s^{2j} \frac{\Gamma(j + \frac{1}{2})}{\Gamma(j+1)} = \frac{\sqrt{\pi}}{2} (S_{k,1} - S_{k,2})$$

² The double-factorial function $n!! = n(n-2)(n-4) \cdots (1)$ must not be confused with the factorial function $n! = n(n-1)(n-2) \cdots (1)$ applied twice.

³ The gamma function $\Gamma(\cdots)$ generalizes the factorial function from the non-negative integers to most complex numbers. For any positive integer n , we have $\Gamma(n) = (n-1)!$, but in particular $\Gamma(\cdots)$ also works for floating-point values.

where $S_{k,1}$ and $S_{k,2}$ are defined as:

$$S_{k,1} := \sum_{j=0}^{\infty} s^{2j} \frac{\Gamma(j + \frac{1}{2})}{\Gamma(j+1)} \quad \text{and} \quad S_{k,2} := \sum_{j=\frac{k}{2}+1}^{\infty} s^{2j} \frac{\Gamma(j + \frac{1}{2})}{\Gamma(j+1)}.$$

The first infinite subseries $S_{k,1}$ can be rearranged by applying the double-factorial identity:

$$\begin{aligned} S_{k,1} &= \Gamma(\frac{1}{2}) \sum_{j=0}^{\infty} s^{2j} \frac{2^j \Gamma(j + \frac{1}{2}) / \Gamma(\frac{1}{2})}{2^j j!} \\ &= \sqrt{\pi} \sum_{j=0}^{\infty} s^{2j} \frac{(2j-1)!!}{(2j)!!} \end{aligned}$$

This is just a binomial series:

$$\begin{aligned} S_{k,1} &= \sqrt{\pi} \sum_{j=0}^{\infty} (-1)^j \binom{-\frac{1}{2}}{j} (s^2)^j \\ &= \sqrt{\frac{\pi}{1-s^2}} = \frac{\sqrt{\pi}}{c} \end{aligned}$$

The second infinite subseries $S_{k,2}$ can be solved by first shifting it back to $j=0$. Letting $a := k/2 + 1$:

$$S_{k,2} = \sum_{j=a}^{\infty} s^{2j} \frac{\Gamma(j + \frac{1}{2})}{\Gamma(j+1)} = \sum_{j=0}^{\infty} (s^2)^{j+a} \frac{\Gamma(j+a + \frac{1}{2})}{\Gamma(j+a+1)}$$

and then rearranging in terms of the Pochhammer symbol $(x)_n := \Gamma(x+n) / \Gamma(x)$:

$$\begin{aligned} S_{k,2} &= (s^2)^a \sum_{j=0}^{\infty} (s^2)^j \frac{\Gamma(j+a + \frac{1}{2})}{\Gamma(j+a+1)} \\ &= (s^2)^a \frac{\Gamma(a + \frac{1}{2})}{\Gamma(a+1)} \sum_{j=0}^{\infty} \frac{(s^2)^j \Gamma(1+j) \Gamma(a + \frac{1}{2} + j) \Gamma(a+1)}{\Gamma(1+j) \Gamma(1) \Gamma(a + \frac{1}{2}) \Gamma(a+1+j)} \\ &= (s^2)^a \frac{\Gamma(a + \frac{1}{2})}{\Gamma(a+1)} \sum_{j=0}^{\infty} \frac{(1)_j (a + \frac{1}{2})_j (s^2)^j}{(a+1)_j j!} \end{aligned}$$

we get the form of a hypergeometric series. Applying the Gaussian hypergeometric function ${}_2F_1(\cdots)$:

$$S_{k,2} = (s^2)^a \frac{\Gamma(a + \frac{1}{2})}{\Gamma(a+1)} {}_2F_1(1, a + \frac{1}{2}; a+1; s^2)$$

Then, because the hypergeometric function takes a special form (the first argument is 1), on to the incomplete beta function $B(x; a, b)$:

$$\begin{aligned} S_{k,2} &= (s^2)^a \frac{\Gamma(a + \frac{1}{2})}{\Gamma(a+1)} a (s^2)^{-a} (1-s^2)^{-1/2} B(s^2; a, \frac{1}{2}) \\ &= \frac{\Gamma(a + \frac{1}{2})}{c \Gamma(a)} B(s^2; a, \frac{1}{2}) \end{aligned}$$

Substituting the $S_{k,1}$ and $S_{k,2}$ subseries back in, we find S_k is:

$$\begin{aligned} S_k &= \frac{\sqrt{\pi}}{2} \left(\frac{\sqrt{\pi}}{c} - \frac{\Gamma(a + \frac{1}{2})}{c \Gamma(a)} B(s^2; a, \frac{1}{2}) \right) \\ &= \frac{\sqrt{\pi}}{2c} \left(\sqrt{\pi} - \frac{\Gamma((k+3)/2)}{\Gamma(k/2+1)} B(s^2; \frac{k}{2} + 1, \frac{1}{2}) \right) \end{aligned}$$

Evaluating A and S_{n-1} for Phong, we find that the normalization factor I_P is:

$$I_P = \frac{2}{n+1} \left(\frac{\pi}{2} + c \frac{\sqrt{\pi}}{2c} \left(\sqrt{\pi} - \frac{\Gamma(n/2+1)}{\Gamma((n+1)/2)} B(s^2; \frac{n+1}{2}, \frac{1}{2}) \right) \right)$$

Similarly, evaluating T_n , A , and S_{n-2} for Modified Phong, we find that the normalization factor I_M is:

$$I_M = \frac{2}{n+2} \left(\frac{\sqrt{\pi}}{2} s^n \frac{\Gamma((n+1)/2)}{\Gamma(n/2+1)} + \frac{\pi}{2} c + c^2 \frac{\sqrt{\pi}}{2c} \left(\sqrt{\pi} - \frac{\Gamma((n+1)/2)}{\Gamma(n/2)} B(s^2; \frac{n}{2}, \frac{1}{2}) \right) \right)$$

Simplification yields our final two normalization terms:

$$I_P = \frac{1}{n+1} \left(2\pi - \sqrt{\pi} \frac{\Gamma(n/2+1)}{\Gamma((n+1)/2)} B(s^2; \frac{n+1}{2}, \frac{1}{2}) \right) \quad (7)$$

$$I_M = \frac{1}{n+2} \left(2\pi c + \sqrt{\pi} \frac{\Gamma((n+1)/2)}{\Gamma(n/2+1)} \left[s^n - \frac{n}{2} c B(s^2; \frac{n}{2}, \frac{1}{2}) \right] \right) \quad (8)$$

For Modified Phong, there is a complicating detail: when $n = 0$, the incomplete beta function becomes indeterminate. In this case, we must replace that term by its limit. To compute the limit, we use the integral definition of the incomplete beta function and evaluate the term:

$$\begin{aligned} \lim_{n \rightarrow 0} -\frac{n}{2} c B(s^2; \frac{n}{2}, \frac{1}{2}) &= -c \lim_{n \rightarrow 0} \frac{n}{2} \int_0^{s^2} t^{n/2-1} (1-t)^{1/2-1} dt \\ &\quad u := \sqrt{1-t} \\ &= -c \lim_{n \rightarrow 0} n \int_c^1 (1-u^2)^{n/2-1} du \end{aligned}$$

The integrand does not satisfy the requirements for bringing the limit into the integral, so we instead proceed by series expansion. We factor the integrand and take the Taylor series at $x = 1$, which converges everywhere and will simplify nicely:

$$\begin{aligned} (1-x^2)^p &= (1-x)^p (1+x)^p \\ &= (1-x)^p \sum_{j=0}^{\infty} (1+x)^{p-j} \binom{p}{j} (x-1)^j \\ &= 2^p \sum_{j=0}^{\infty} \binom{p}{j} \left(\frac{-1}{2} \right)^j (1-x)^{p+j} \end{aligned}$$

We substitute the series into the integral:

$$\begin{aligned} &-c \lim_{n \rightarrow 0} n \int_c^1 (1-u^2)^{n/2-1} du \\ &= -c \lim_{n \rightarrow 0} n \int_c^1 \left[2^{n/2-1} \sum_{j=0}^{\infty} \binom{n/2-1}{j} \left(\frac{-1}{2} \right)^j (1-u)^{n/2-1+j} \right] du \end{aligned}$$

From here, we simplify and rearrange:

$$= -c \lim_{n \rightarrow 0} 2^{n/2} \sum_{j=0}^{\infty} \frac{n}{2} \binom{n/2-1}{j} \left(\frac{-1}{2} \right)^j \int_c^1 (1-u)^{n/2-1+j} du$$

$$= -c \lim_{n \rightarrow 0} 2^{n/2} \sum_{j=0}^{\infty} \frac{n}{2} \binom{\frac{n}{2}-1}{j} \left(\frac{-1}{2} \right)^j \left(\frac{(1-c)^{n/2+j}}{\frac{n}{2}+j} \right)$$

When $j = 0$, the leading $n/2$ cancels with the $n/2 + j$ in the denominator, leaving the term nonzero. This cancellation does not happen in subsequent terms, so the leading $n/2$ causes them all to disappear when we take the limit $n \rightarrow 0$. Therefore, we have but to pull the **first term** out of the summation to attain the limiting value:

$$\begin{aligned} &= -c \lim_{n \rightarrow 0} 2^{n/2} \frac{n}{2} \binom{\frac{n}{2}-1}{0} \left(\frac{-1}{2} \right)^0 \left(\frac{(1-c)^{n/2+0}}{\frac{n}{2}+0} \right) \\ &= -c \lim_{n \rightarrow 0} 2^{n/2} (1-c)^{n/2} \\ &= -c \end{aligned}$$

Therefore, if the specular exponent $n = 0$, the incomplete beta term $-\frac{n}{2} c B(s^2; \frac{n}{2}, \frac{1}{2})$ for computing I_M in **Equation 8** should be replaced by $-c$.

Compare Equations 7 and 8 to the energy conservation terms in **Equation 6**. We can see that when $c = 1$ (i.e. at normal incidence), the former reduces to the latter (since $c = 1$ means $s = 0$, and $B(0; \dots, \dots) = 0$). However, these normalization formulae are correct for other incidence also. No series or products remain, and in practice computing the (ratio of) gamma function(s) and the incomplete beta function are constant-time operations (see next section), so our approach has $O(1)$ complexity.

Although we derived this result using n even, the same result is obtained when n is odd, and so this formula works for all integer-valued $n \geq 0$. In-fact, empirically, the algorithm produces correct results for *real-valued* $n \geq 0$.

4 Results

We evaluate these normalization methods with a GPU pathtracer based on OptiX [13] running on an NVIDIA GTX 1080, using standard importance sampling [4].

4.1 Numerical Validation

The top graph of **Figure 1** shows different values of I_M computed for the full range of incident light angles and a range of exponent values n . Arvo's method only works with integer values of n , so its results are drawn as curves (in red). Our results match Arvo's method (down to numerical precision), as expected. In addition, our method also smoothly interpolates between them for non-integer values.

The bottom graph of **Figure 1** shows the difference between our method and a numerically computed integral. Notice that our results closely match the

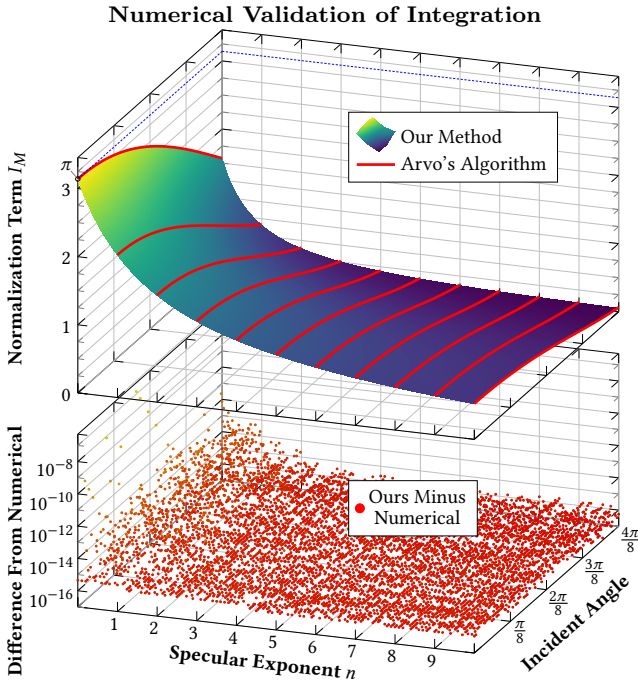


Fig. 1 Validating the Modified Phong BRDF (the plots for Phong are qualitatively similar). The interesting range of n is shown; at larger values the graph tapers smoothly as one expects. **Top:** Arvo’s algorithm can normalize at integer specular exponents. Our method works for any specular exponent. **Bottom:** Our algorithm differs from a numerical validation by no more than 2.54×10^{-7} , which is comparable to the worst-case convergence of the numerical procedure itself.

numerically computed results to 2.54×10^{-7} , which is roughly the numerical integrator’s worst-case accuracy. This validates that we can compute the correct normalization terms, including particularly for floating-point specular exponents where Arvo’s method cannot provide a ground truth.

To visualize the shape of the integral far from $n = 0$, we present a visualization of the (reciprocal of) the integrals of both BRDF variants in **Figure 2**. With increasing n , the reciprocals $1/I_P$ and $1/I_M$ (used in Equations 2 and 3) continue to grow, as expected: since the areas of the BRDFs become progressively smaller, their values approach the delta function of a perfect specular BRDF. While we did not encounter numerical problems in this range, they are inevitable for any specular BRDF that becomes sufficiently close to a delta but remains finite.

4.2 White Furnace Test

Figure 3 shows a standard white furnace test. A sphere is placed into a uniform environment, and the specular albedo is set to $\rho_s = 1$ (i.e. $\rho_d = 0$). Under these circumstances, every point on the object must have the

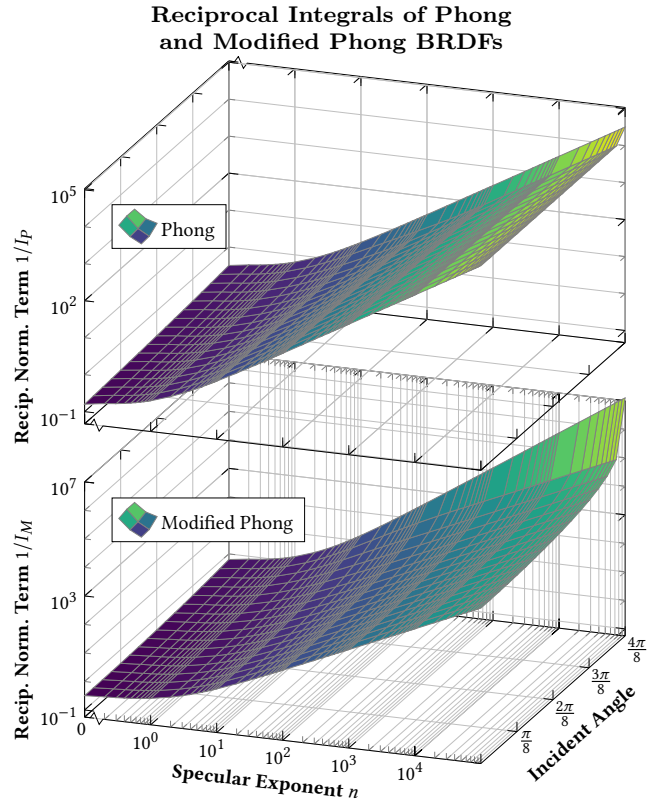


Fig. 2 Log-log plots of the reciprocal normalization terms $1/I_P$ and $1/I_M$ over a wide range of specular exponent n values. Notice that as n increases, I_P and I_M get closer to zero and their reciprocals continue to increase.

same radiance as the environment. The object should therefore “disappear” into the background. The specular exponent for both BRDFs is arbitrarily set to $n = 5$ (note that the qualitative behavior in all four images is not affected by the exponent).

On the left, we see the effect of the typical energy conservation terms used today (i.e., **Equation 6**). As the specular lobe tilts over at glancing angles, it intersects progressively more with the ecliptic plane. This energy is lost, causing erroneous darkening of the edges.

On the right, we see the effect of energy normalization (computed with our algorithm, although Arvo’s algorithms would give the same result for integer-valued specular exponents). The sphere cannot be readily distinguished from the background (though due to quantization and sampling issues, there is variation of about half a gray level, which may be visible).

4.3 Performance Comparison

Figure 4 shows the performance scaling of various methods, as measured for the white furnace test.

Only doing energy conservation (pale orange) requires negligible computation, and so should be viewed

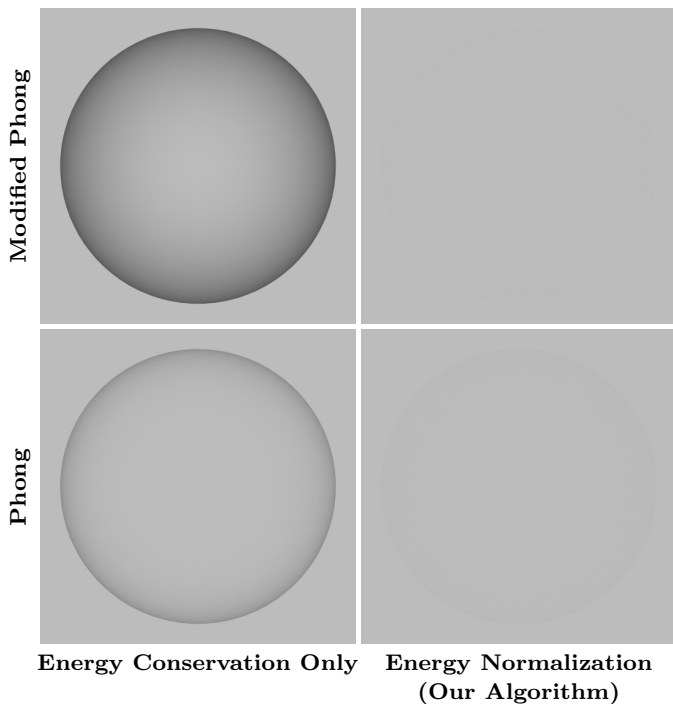


Fig. 3 In the white furnace test, a sphere is placed in an uniform environment. **Left:** energy conservation terms prevent energy from being gained, but energy is still lost at glancing angles. **Right:** energy normalization correctly ensures energy is neither lost nor gained.

as a lower-bound on BRDF latency for this rendering setup. Our constant-time approach (blue) adds a small amount of latency over energy conservation (for all data, at worst about 3.6%). We feel like this is a small price to pay for correctness—and in scenes with actually interesting geometry, it will be much smaller. Meanwhile, Arvo’s method (red) has non-constant time complexity. Although performance is good in the diffuse and weakly glossy regime, at higher specular exponents, the algorithm quickly becomes intractable.

While it is true that not all BRDFs are specular or highly glossy, Arvo’s method has an additional practical disadvantage in that it cannot handle floating-point specular exponents: on the left side of the plot, the red line can be seen is really merely a series of points. Our method has no such difficulty.

It is worth noting, again, that energy conservation *does not give* the correct result, whereas Arvo’s method and our method do. Therefore, energy conservation is simply a lower-bound on latency, not a performance that must (or even can) be matched. Nevertheless, for completeness, we report on the added latency of the individual pieces in a CPU isolation test⁴; we find a la-

⁴ Profiling small sections of code is theoretically problematic (pipelining, caching, context-switching, reordering etc. issues), but we report on our best effort on an Intel i7-6850K.

tency factor of $68\times$ when considering the normalization factor alone and $4.4\times$ when considering evaluation of the whole BRDF (without textures)—compared to the above $1.036\times$ when a realistic rendering context is considered. This last statistic is the most salient: the low textural and geometric complexity of the figure serves as a worst-case for bounding rendering by BRDF evaluation, and we would expect this (or better) performance for any actual usage of our algorithm.

4.4 Qualitative Behavior

We compare the Phong and Modified Phong BRDFs with energy conservation versus with normalization terms. **Figure 5** shows an example object with Lambertian diffuse and Phong specular variants, with specular exponent values keyed to roughness, under environment illumination. Although the appearance in all cases is similar, without energy normalization, energy loss can be observed whenever a significant portion of the specular lobe intersects the ecliptic plane, such as at glancing angles, or on surfaces with lower specular exponent where the lobe is wider overall.

5 Conclusion

We have presented a method for normalizing the Phong and Modified Phong specular component BRDFs. While previous work was able to, in linear- or quadratic-time, normalize for nonnegative integer-valued specular exponents, our result is constant-time and works for real-valued nonnegative (i.e., all reasonable) specular exponents.

Conflict of Interest: The authors declare that they have no conflict of interest.

References

1. Arvo, J.: Applications of irradiance tensors to the simulation of non-lambertian phenomena. In: Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH ’95, pp. 335–342. ACM, New York, NY, USA (1995). DOI 10.1145/218380.218467. URL <http://doi.acm.org/10.1145/218380.218467>
2. Beckmann, P., Spizzichino, A.: The scattering of electromagnetic waves from rough surfaces. Norwood, MA, Artech House, Inc., 1987, 511 p. (1987)
3. Boost: Boost C++ Libraries. <http://www.boost.org/> (1999)
4. Dutré, P.: Global illumination compendium. Tech. rep., Katholieke Universiteit Leuven (2003). <https://people.cs.kuleuven.be/~philip.dutre/GI/>

Latency of Normalization Approaches

(Lower is Better)

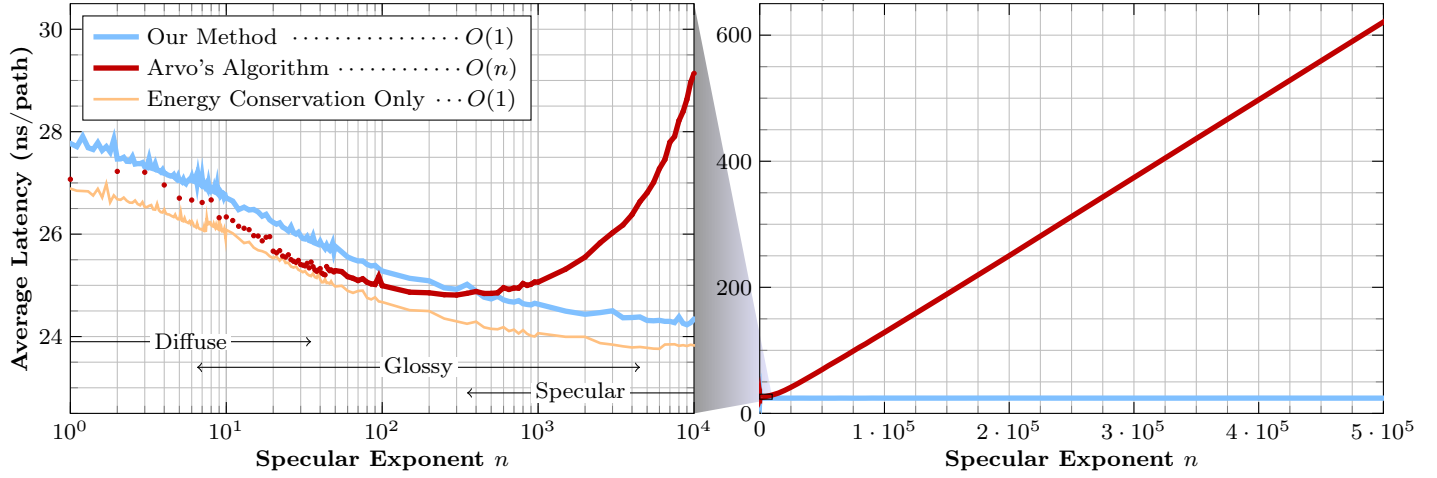


Fig. 4 Cost when using our approach vs. Arvo’s method to normalize Phong (the plot for Modified Phong is similar) in the scene for [Figure 3](#). Both methods are efficient for low specular exponents (compare to enforcing energy conservation only, which requires negligible computation and should be seen as a lower-bound). As specular exponents increase into the glossy and specular range, the linear runtime of Arvo’s method quickly makes it intractable, but our constant-time method is unaffected. Note that all methods receive a slight benefit as exponents increase due to increased locality of secondary rays.

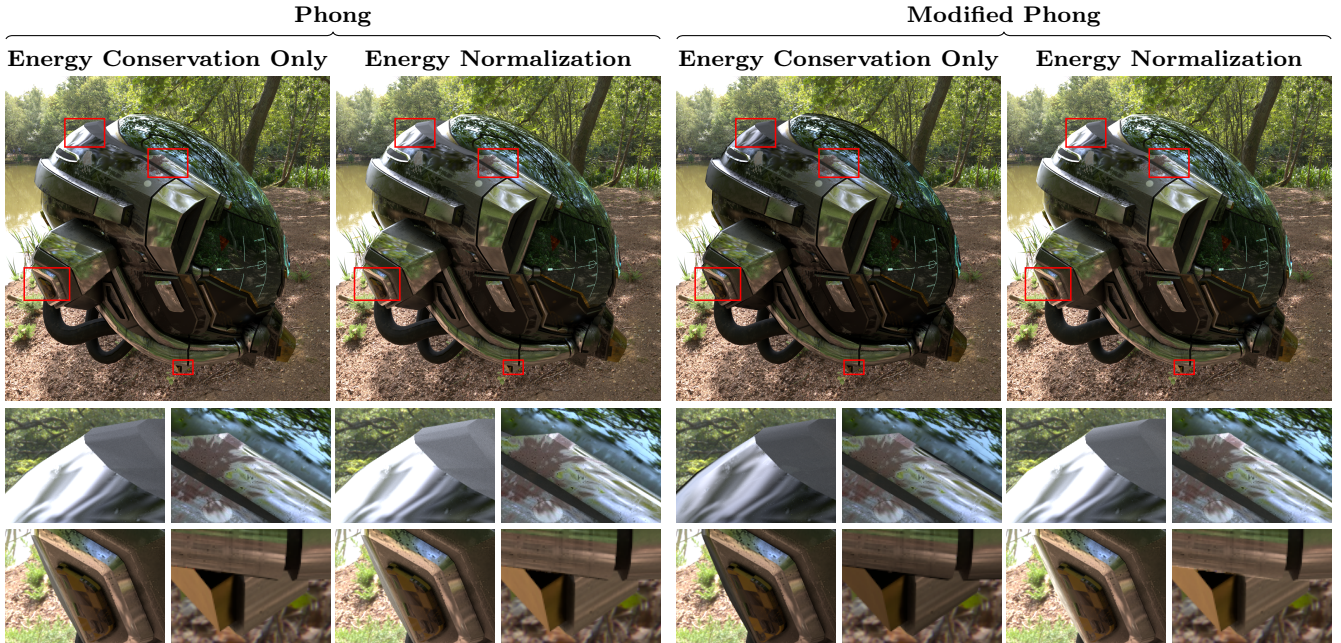


Fig. 5 Effect of energy normalization: unlike mere conservation, energy is not lost, leading to noticeably brighter reflection at glancing angles and on more-diffuse surfaces. Model is the [Damaged Helmet](#) GLTF sample and environment is [Epping Forest](#).

5. Heitz, E., Hanika, J., d’Eon, E., Dachsbacher, C.: Multiple-scattering microfacet bsdfs with the smith model. *ACM Transactions on Graphics (TOG)* **35**(4), 1–14 (2016)
6. Kajiya, J.T.: The rendering equation. In: *Proceedings of the 13th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH ’86*, pp. 143–150. ACM, New York, NY, USA (1986). DOI 10.1145/15922.15902. URL <http://doi.acm.org/10.1145/15922.15902>
7. Kulla, C., Conty, A.: Revisiting physically based shading at imageworks. *SIGGRAPH Course, Physically Based*

Shading (2017)

8. Lecocq, P., Dufay, A., Sourimant, G., Marvie, J.E.: Analytic approximations for real-time area light shading. *IEEE transactions on visualization and computer graphics* **23**(5), 1428–1441 (2017)
9. Lee, J.H., Jarabo, A., Jeon, D.S., Gutierrez, D., Kim, M.H.: Practical multiple scattering for rough surfaces. *ACM Transactions on Graphics (TOG)* **37**(6), 1–12 (2018)
10. Lewis, R.R.: Making Shaders More Physically Plausible. *Computer Graphics Forum* **13**(2), 109–120 (1994). DOI

- 10.1111/1467-8659.1320109
11. Meta.Numerics: Meta.numerics. <http://www.meta-numerics.net/> (2009)
 12. Nicodemus, F.E.: Directional reflectance and emissivity of an opaque surface. *Applied optics* **4**(7), 767–775 (1965)
 13. Parker, S.G., Bigler, J., Dietrich, A., Friedrich, H., Hoberock, J., Luebke, D., McAllister, D., McGuire, M., Morley, K., Robison, A., Stich, M.: Optix: A general purpose ray tracing engine. In: ACM SIGGRAPH 2010 Papers, SIGGRAPH '10, pp. 66:1–66:13. ACM, New York, NY, USA (2010). DOI 10.1145/1833349.1778803. URL <http://doi.acm.org/10.1145/1833349.1778803>
 14. Phong, B.T.: Illumination for computer generated pictures. *Communications of the ACM* **18**(6), 311–317 (1975). DOI 10.1145/360825.360839. URL <http://doi.acm.org/10.1145/360825.360839>
 15. Torrance, K.E., Sparrow, E.M.: Theory for off-specular reflection from roughened surfaces. *Josa* **57**(9), 1105–1114 (1967)
 16. Turquin, E.: Practical multiple scattering compensation for microfacet models. ILM (2019)

A - Derivation of Energy Normalization Criterion

The following informal proof demonstrates the well-known BRDF normalization criterion. We start with 1:

$$1 = \frac{L_o}{L_o} = \frac{1}{L_o} \int_{S^2} dL_o = \int_{S^2} \frac{dL_o}{L_o (\mathbf{N} \cdot \boldsymbol{\omega}_o)} (\mathbf{N} \cdot \boldsymbol{\omega}_o) d\boldsymbol{\omega}_o$$

Then, since the incoming irradiance $E_e = d\Phi_i/dA$ must match the outgoing radiant exitance $M_e = d\Phi_o/dA$:

$$\begin{aligned} 1 &= \int_{S^2} \frac{dL_o}{dM_e} (\mathbf{N} \cdot \boldsymbol{\omega}_o) d\boldsymbol{\omega}_o = \int_{S^2} \frac{dL_o}{dE_e} (\mathbf{N} \cdot \boldsymbol{\omega}_o) d\boldsymbol{\omega}_o = \\ &= \int_{S^2} f_s(\boldsymbol{\omega}_i \rightarrow \mathbf{x} \rightarrow \boldsymbol{\omega}_o) (\mathbf{N} \cdot \boldsymbol{\omega}_o) d\boldsymbol{\omega}_o \end{aligned}$$

B - Derivation of Energy Conservation

Unlike energy *normalization*, which is the subject of this paper, the derivation of energy *conservation* terms for the Phong BRDFs is widely known.

Consider the Modified Phong specular component BRDF. In [Equation 5](#), first assume that $\boldsymbol{\omega}_i = \mathbf{N}$:

$$I_M = \int_{\Omega} \max(0, \mathbf{R}(\boldsymbol{\omega}_o) \cdot \mathbf{N})^n (\mathbf{N} \cdot \boldsymbol{\omega}_o) d\boldsymbol{\omega}_o$$

Then, since the dot product of the reflected vector must be nonnegative and the same as that of the original vector:

$$\begin{aligned} I_M &= \int_{\Omega} (\boldsymbol{\omega}_o \cdot \mathbf{N})^n (\mathbf{N} \cdot \boldsymbol{\omega}_o) d\boldsymbol{\omega}_o \\ &= \int_{\Omega} (\mathbf{N} \cdot \boldsymbol{\omega}_o)^{n+1} d\boldsymbol{\omega}_o \\ &= \int_0^{2\pi} \int_0^{\pi/2} \cos^{n+1}(\phi) \sin(\phi) d\phi d\theta \\ &= \int_0^{2\pi} \frac{1}{n+2} d\theta \\ &= \frac{2\pi}{n+2} \end{aligned}$$

The computation of I_P with [Equation 4](#) is analogous.

C - Example Code

Although the normalization terms given by [Equations 7 and 8](#) may look intimidating, we provide a listing of the key source code, with simple optimizations applied, to demonstrate that our approach can be implemented in only a few lines:

```
#include <cmath>

#define TWO_PI  6.2831853f
#define SQRT_PI 1.77245385f

//Incomplete Beta function
float ibeta( float x, float a, float b );

//Computes Γ(a) / Γ(b)
inline static float gamma_quot( float a, float b ) {
    return std::exp( std::lgamma(a) - std::lgamma(b) );
}

inline static float calc_I_P( float NdotV, float n ) {
    float const& costerm = NdotV;
    float        sinterm_sq = 1.0f - costerm*costerm;
    float        halfn      = 0.5f * n;

    return (
        TWO_PI -
        SQRT_PI*gamma_quot( halfn+1.0f, halfn+0.5f ) *
        ibeta( sinterm_sq, halfn+0.5f, 0.5f )
    ) / (n+1.0f);
}

inline static float calc_I_M( float NdotV, float n ) {
    float const& costerm = NdotV;
    float        sinterm_sq = 1.0f - costerm*costerm;
    float        halfn      = 0.5f * n;

    float negterm = costerm;
    if (n>=1e-18f)
        negterm*=halfn*ibeta( sinterm_sq, halfn, 0.5f );

    return (
        TWO_PI*costerm +
        SQRT_PI*gamma_quot( halfn+0.5f, halfn+1.0f ) *
        ( std::pow(sinterm_sq, halfn) - negterm )
    ) / (n+2.0f);
}
```

The incomplete beta function is the only omission from the above, since it is not provided by most implementations of the C/C++ standard library. However, there is no shortage of available implementations to choose from. For example, the `boost::math::beta(...)` function from Boost [\[3\]](#) will be available to many users; the additional implementation would look like:

```
#include <boost/math/special_functions/beta.hpp>
float ibeta( float x, float a, float b ) {
    return boost::math::beta( a, b, x );
}
```

In our experiments, we chose to adapt the implementation of [\[11\]](#), because it seems to be based on a newer method. Such choices also in principle affect the epsilon at which the incomplete beta term must be replaced with its limit. However, empirically, our implementation was stable down to $n \approx 5 \times 10^{-37}$, so a very small epsilon, as-above, is still adequately conservative.