

Material Appearance Modeling: Radiometry, Surface Scattering, and Subsurface/Volume Scattering

Yuhao Zhu

Department of Computer Science
Department of Brain and Cognitive Sciences
University of Rochester

yzhu@rochester.edu
<https://yuhaozhu.com/>
<https://horizon-lab.org/>

Contents

1 Overview	3
2 Observed Reflection and Transmission	4
3 A Little Bit of Radiometry and Photometry	6
3.1 Energy and Power	6
3.2 Irradiance	7
3.3 Solid Angle	7
3.4 Radiance	9
3.5 Lambertian Emitter, Radiant Intensity, and Lambert's Cosine Law	11
3.6 The Measurement Equation in Camera Imaging	12
3.7 Light Field	13
3.8 Photometric Quantities	13
4 Surface Scattering	14
4.1 BRDF	15
4.2 Reflectance and Albedo	16
4.3 The Rendering Equation	19
4.4 Specular vs. Glossy vs. Diffuse Materials	21
4.5 BRDF Parameterization with Microfacet Models	24
5 Measuring Spectral Reflectance and BRDF	26
5.1 Measuring Spectral Reflectance	27
5.2 Measuring BRDF	31

6 Subsurface and Volume Scattering: Informal	34
6.1 General Intuitions	35
6.2 Transparent vs. Opaque vs. Translucent Materials	36
6.3 Equilibrium	37
7 Absorption	38
7.1 A Simple Case: Collimated Illumination on Uniform Medium	38
7.2 A Few Important Quantities	40
7.3 General Case	42
7.4 Nature and Applicability of the Model	43
8 Scattering	44
8.1 Scattering by a Particle vs. a Collection of Particles	45
8.2 A Single Scattering Event	47
8.3 Common Models and General “Rules”	52
9 Radiative Transfer Equation and Volume Rendering	56
9.1 Radiative Transfer Equation	56
9.2 Volume Rendering Equation	58
9.3 Discrete VRE and Scientific Volume Visualization	60
9.4 Discrete VRE in Neural Rendering	64
9.5 Integrating Surface Scattering with Volume Scattering	68
10 The Kubelka–Munk Model	69
10.1 Deriving the Model	70
10.2 The Model and Its Interpretation	75
10.3 N-Stream Model	76
10.4 Correction for Surface Reflection	77

So far we have assumed that objects have colors because they emit lights. In the real world, the vast majority of objects have colors not because they emit lights by themselves, but because they scatter lights from a light source to our eyes. For instance, if the illuminant is white-ish and a paint scatters only long-wavelength lights (while other photons are either absorbed or pass through the paint), the paint would look red-ish. Similarly, if we see through an optical filter such as a sunglass, the sunglass might look, say, blue-ish, because while the light entering the sunglass might be white-ish, the sunglass allows only short-wavelength lights to go through and so the transmitted light would look blue-ish, which we mentally equate with the color of the sunglass itself.

In both cases, the material modifies the energy spectrum of the light illuminating it, and the modified light that enters our eyes dictates the color of the material. We have to go back to physics and first principles to model the interaction between light and the material, from which we can then model the material color.

1 Overview

When a beam of photons hit a material surface, some of the photons will be scattered directly back to your eyes, others will penetrate into the material. These surface phenomena are governed by **surface scattering**. We use the word “scattering” here to generally refer to lights coming back from the surface. Depending on the material, some of the scattered photons are along the perfect mirror-reflection directions and others might be more diffuse. You might sometimes see the word “reflection” used. Reflection sometimes is used in the same way as scattering, which will be our use, but other times is reserved for the perfect, mirror-like reflection. Usually what the word means is self-evident given the context, but we will err on the side of verbosity when we want to mean a specific form of reflection.

Photons that penetrate the surface will further interact with particles in the material, which absorb, scatter, or might even emit photons. This is called **subsurface scattering** (SSS) in computer graphics. Even though we use the term “scattering”, you should know that the actual SSS processes involve not only scattering but also absorption and emission. It turns out that the principles that govern SSS are exactly the same as those that govern the interactions between photons and particles in the so-called “participating media”, such as clouds, fogs, and smokes. In computer graphics, light transport in participating media is called **volume scattering**, and again, even though we use the term “scattering”, absorption and emission are usually involved in the most general cases.

The way to model SSS/volume scattering is different from the way to model surface scattering: we no longer consider the material as a continuous surface and the light-matter interaction as photons bounding off of the surface; instead, we break a material down into small particles and model how photons interact with individual particles.

Very importantly, the difference in the modeling methodology does *not* imply that there somehow is a fundamental difference between surface scattering and volume scattering. Ultimately, both are caused by the light, an oscillating electromagnetic field, exciting discrete electric charges. The differences lie in how the charges are arranged in space and in relation to

one another. The laws that govern how photons interact with the charges are described by the electromagnetic theories in the classical regime and, in the quantum regime, by the quantum electrodynamics (QED)¹. In fact, using the electromagnetic theories, we can show that surface reflection/refraction is nothing more than the coherent scattering of incident light waves by the surface particles.

Since there is no fundamental differences in the underlying physics, the only meaningful distinction is one between different phenomenological approximations, or “models”, of the same underlying physics. We totally could invoke the electromagnetic theories or QED, and if we did, we would have one single unified model that explains both surface scattering (reflection and refraction) and volume scattering. Doing so, however, is not only unnecessary (because many, not all, real-world material color phenomena could be modeled without them) and too computationally expensive, but also, perhaps more importantly, blinds us from the relatively simple intuitions in each scenario. Instead, each phenomenological model is based on a set of high-level guiding principles, which are approximates of the underlying physical process but are sufficient to quantitatively describe light-matter interactions in each scenario.

[Johnsen \[2012\]](#) is a great reference, which has some equations but generally focuses on building intuitions and mostly uses the electromagnetic language rather than the quantum language. If you want to get to the nuts and bolts of the mathematical modeling, [Bohren and Clothiaux \[2006\]](#) is a phenomenal text, whose models are also built in the electromagnetic land. [Feynman \[1985\]](#) has an accessible and breathtaking introduction of QED that I highly recommend. [Dorsey et al. \[2010\]](#) is a classic text on material appearance modeling in graphics that covers a range of topics, including modeling, measurements, and various implementation issues in practice. [Johnston-Feller \[2001\]](#) is specifically concerned with paintings; it has many interesting discussions of pigments and pigment mixtures, and has many real-world data and insights that are rarely found elsewhere.

2 Observed Reflection and Transmission

Regardless of the details of surface scattering and volume scattering, a material appears to have some color because some photons leaving the material enter our eye. If we observe the material from the same side of the light source, it is the lights reflected from the material that matter. If we observe the material from the other side of the light source, it is the lights transmitted through the material that matter.

So at the highest level of abstraction, we can model the material color in the real world by modeling the *observed* reflection and transmission from an outside observer: how much of the incident power is reflected/transmitted back to the eye?

We can quantify the observed reflection and transmission using the **spectral reflectance function** $r(\lambda)$ and the **spectral transmittance function** $t(\lambda)$, respectively. These two functions spare us the details of how lights interact with a material, but describe, at each wavelength

¹The electromagnetic theories do not explain everything in light-matter interactions. Famously, they do not explain how the interference pattern in the double-slit experiment still arises even if the photons are delivered sequentially.

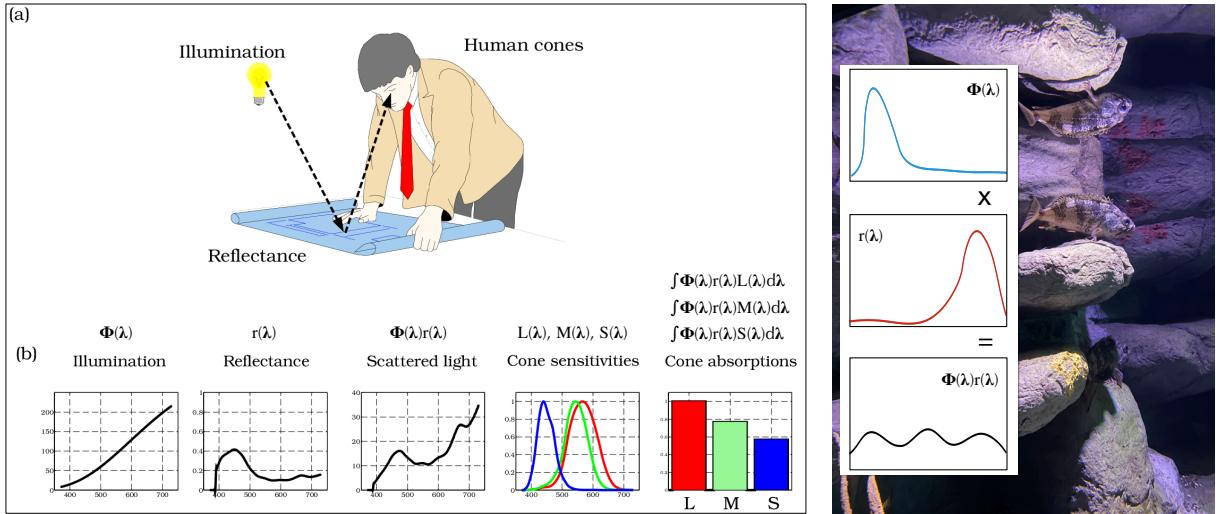


Figure 1: Left: apparent spectral reflectance modifies the illumination spectrum and dictates the observed color; adapted from Wandell [1995, Fig. 9.2]. Right: a photo of Acadia Redfish I took in the Ripley’s Aquarium of Canada. The fish ordinarily looks red-ish under a white-ish light, but appears colorless in the aquarium, which simulates the lighting environment in the deep sea where lights are predominately blue/violet. The spectral data are not accurate and for the illustration purpose only.

λ , the percentage of optical power that is reflected back to the eye or transmitted through the material and enters the eye, respectively. The left (a) panel in Figure 1 illustrates this modeling setup.

Given a light source $\Phi(\lambda)$, the light reflected toward the eye is then $\Phi(\lambda)r(\lambda)$ and the light transmitted through the material is $\Phi(\lambda)t(\lambda)$. We can then calculate the color of these lights using the cone fundamentals or some set of CMFs, the same way as if the lights were directly emitted. The math is illustrated in the left (b) panel in Figure 1.

The right panel in Figure 1 is a photo of Acadia Redfish I took in the Ripley’s Aquarium of Canada. The fish ordinarily looks red-ish under a white-ish light, which suggests that its spectral reflectance $r(\lambda)$ peaks at longer wavelengths: it scatters more long-wavelength, i.e., red-ish, lights than short-wavelength lights. But the fish appears colorless in the aquarium, which simulates the lighting environment in the deep sea where lights $\Phi(\lambda)$ are predominately blue/violet². As a result, the scattered lights have a rather uniform spectral power distribution, resulting in a gray-ish appearance.

The left image in Figure 1 makes an important simplification: the reflectance of a point p on the material is simplified to only a single spectrum. In reality, the reflectance of a point p depends on both ω_i , the direction of the light incident on p , and ω_s , the outgoing direction (leaving p) through which one observes the material. In certain materials where SSS contributes

²which results from a combination of water selectively absorbing medium-to-long wavelengths of light and increasing scattering of short wavelengths in the Rayleigh regime (Chapter 8.3).

to the material appearance (e.g., translucent materials like jade), the reflectance can also depend on lights incident on *other* points of the material surface. So when we use a single reflectance spectrum to model material colors, what we have implicitly assumed is that the reflectance spectrum has been calculated in such a way that when you multiply it with the incident illumination you get the scattered light power that is actually observed.

How can such a reflectance spectrum be obtained in measurement (to the extend that it is a useful high-level abstraction) will be discussed in Chapter 5. The reflectance is a “quick-and-dirty” abstraction that we often use to give a rough estimation/explanation of a material’s color, but it is so high-level that it hides lots of the low-level details: what exactly are the mechanisms in light-matter interactions that give rise to the scattering behavior of a particular material? The rest of this Chapter essentially answers this question by building physically-based mathematical models.

3 A Little Bit of Radiometry and Photometry

To be more formal about surface and volume scattering, we need to scientifically define a few physical properties pertaining to light propagation spatially and angularly. This is called **radiometry**, which operates completely at the geometric optics level, so we will be describing lights as a collection of photons, each of which can travel along a particular direction with certain energy associated with it. Reinhard et al. [2008, Chpt. 6] and Bohren and Clothiaux [2006, Chpt. 4] have more rigorous treatments of radiometry, and we will here just introduce the language and a few important radiometric quantities that are relevant to our discussion.

3.1 Energy and Power

Each photon carries a certain amount of energy that is determined by its wavelength governed by:

$$Q = \frac{hc}{\lambda}, \quad (1)$$

where c is the speed of light, λ is the photon wavelength, and h is the Planck’s constant.

Power or, more formally in radiometry, **radiant flux** is the total amount of energy passing through some surface in spatial per unit time. Or taking a calculate perspective, power Φ is defined as:

$$\Phi = \lim_{\Delta t \rightarrow 0} \frac{\Delta Q}{\Delta t} = \frac{dQ}{dt}. \quad (2)$$

The way to think about this is that each photon carries a certain amount of energy so if you monitor photons passing across a surface over a period of time Δt , you can calculate the average power of that period by dividing the total energy passed by by Δt . As Δt approaches 0, we get the instantaneous power.

Of course, energy/power is a function of wavelength, so more rigorously we should be talking about *spectral* power $\Phi(\lambda)$, which has a unit of W/nm :

$$\Phi(\lambda) = \lim_{\Delta\lambda \rightarrow 0} \frac{\Delta\Phi}{\Delta\lambda} = \frac{d\Phi}{d\lambda}, \quad (3)$$

where $\Delta\Phi$ is the total power within a wavelength interval $\Delta\lambda$.

3.2 Irradiance

Our power calculation is done with respect to a surface area, but how about the power at each point on the surface area? You can imagine that some points get more photons and others get fewer, so it is useful to characterize the power at any given point. Technically, the answer to the question “how many photons hit a particular point” is *zero*, since the area of a single point is 0^3 . The meaningful question is: what the power *density* of a particular point p ? **Irradiance** is such a quantity.

Imagine again that you are monitoring photons crossing a surface for a period Δt ; you can calculate the average power received per unit area by dividing the average power by the surface area, and when you shrink the surface area to an infinitesimal point p , we can calculate the power density, i.e., the irradiance, of p by:

$$E(p) = \lim_{\Delta A \rightarrow 0} \frac{\Delta\Phi(p)}{\Delta A} = \frac{d\Phi(p)}{dA}. \quad (4)$$

Irradiance is a more primitive measure than power, because we can derive the power of a surface by integrating the irradiance over the surface area:

$$\Phi = \int^A E(p)dA. \quad (5)$$

Irradiance has a unit of Wm^{-2} , and *spectral* irradiance has a unit of $Wm^{-2}nm^{-1}$.

3.3 Solid Angle

Irradiance is concerned with the power of all the photons incident on a point, but photons hit a point from all directions, so how do we quantify the amount of light coming from a direction?

A direction is a vector, which is invariant to translational transformations, so the two parallel “arrows” r_1 and r_2 in Figure 2 (left) represent the same vector/direction. Therefore, conceptually it is easier if we translate all the arrows so that they start from the same origin when we want to reason about a collection of directions.

³A similar question is: imagine you are throwing darts at a wall; what is the probability of hitting a particular point p ? The answer is 0. The meaningful question to ask is: what is the probability density of hitting p ?

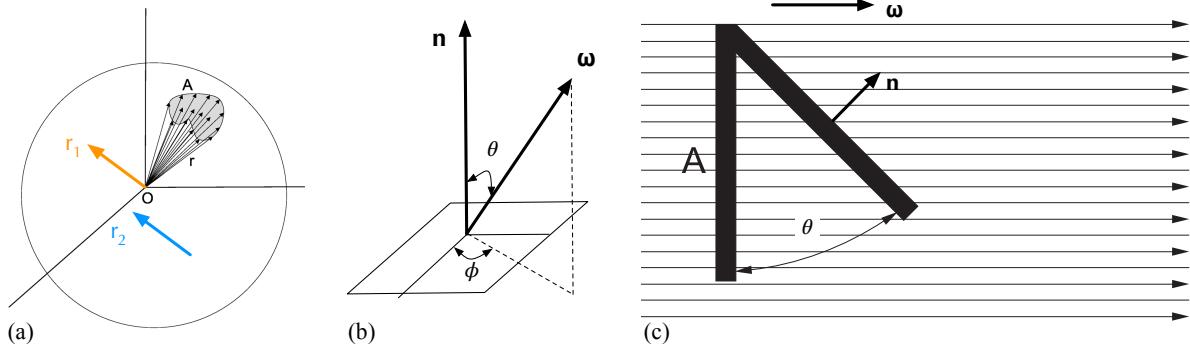


Figure 2: (a): a solid angle is measure of the size of a collection of directions in 3D. A direction is a vector, which is translational invariant, so r_1 and r_2 refer to the same direction. (b): in spherical coordinate systems, a 3D direction can be parameterized by two angles, a polar angle θ and an azimuthal angle ϕ . (c): radiance is an intrinsic property of the radiation field, but we can measure it differently. Adapted from [Bohren and Clothiaux \[2006, Fig. 4.1, 4.2\]](#).

How do we count the amount of directions? In 2D, we use a *planar angle* to measure the amount of directions. Given an origin O and a vector, we rotate it to generate an arc. The angle subtended by the arc and O is a measure of the amount of directions we have just covered. The angle can also be mathematically given by the ratio s/r , where s is the arc length and r is the radius of the circle. This matches our intuition that if we increase the radius of the circle we would get a longer arc but the same angle. A full circle has a planar angle of 2π .

We can similarly define the size of a set of directions in 3D. We draw a sphere around O , and imagine that we have some area on the spherical surface. Connecting O to every point on that area represents a direction in 3D. So the spherical surface area is a measure of the amount of 3D directions. Like in the 2D case, we want the measure to be invariant to the spherical radius, so we define **solid angle**, a measure of the size of a set of 3D directions, as:

$$\Omega = \frac{A}{r^2}, \quad (6)$$

where A is an area on a spherical surface and r is the radius. The unit of a solid angle is **steradian (sr)**, and the entire sphere subtends a solid angle of 4π .

Sometimes we want to know the size of the set of directions from a point O to an arbitrary surface. We would project that surface to a sphere and get a projected spherical area A , using which we can invoke Equation 6 to estimate the solid angle subtended by the surface. One useful trick that might help sometimes is to project the surface to unit sphere (i.e., $r = 1$), and the solid angle is mathematically equivalent to the projected area on the unit sphere. But the most useful intuition I use whenever I am confused about what a particular solid angle means is to always think of the set of directions/vectors that are represented by that solid angle.

3.4 Radiance

We can now ask, what is the amount of flux received by a point from a particular direction? Photons travel in all sort of directions. Let's consider an imaginary detector with an area A that is able to receive light from only one direction ω , as illustrated in Figure 2 (right). We can measure the total flux received by the detector Φ , from which we know that the power per unit area along the direction ω is simply $\frac{\Phi}{A}$.

Now imagine that we place the detector so that its normal subtends an angle θ with respect to the light direction ω . Figure 2 (b) explicitly illustrates this angle, where the tilted detector lies in the xy -plane, and the z direction is the normal n . In a spherical coordinate system, a direction ω can be parameterized by two angles: a polar angle θ and an azimuthal angle ϕ .

The total flux received by the detector has changed to $\Phi \cos \theta$, because the area that is available to receive photons is now $A \cos \theta$. We call this the “effective area”. As a result, the power per area at the direction ω remains the same, i.e., $\frac{\Phi}{A}$. This is not surprising, because we are not changing the radiation field, only how we measure it. When the effective area reaches 0 (i.e., the detector is completely parallel to the light direction), the detector collects no photon, but it certainly does not mean that there is no light in the field.

If we now want to measure light power coming from another direction, we would change the detector so that it receives lights from only that direction. In reality, this is of course not possible. No detector can screen lights only from one direction. If we place a detector in a radiation field, it is going to receive photons from all sorts of directions. We can limit the directions of photons that the detector collects by placing a baffle that allows only certain directions to hit the detector.

This setup is illustrated in Figure 3 (left). The total flux collected by the detector is $\Delta\Phi$, the detector size is ΔA , and the solid angle subtended by the baffle is $\Delta\omega$. The average power collected per unit “effective area” per unit direction by the detector is then:

$$\frac{\Delta\Phi}{\Delta A \cos \theta \Delta\omega}. \quad (7)$$

The baffle does a good job of rejecting many directions that are outside $\Delta\omega$, but unless it is infinitely long, the detector will still collect some photons traveling through directions outside $\Delta\omega$. But as we reduce the detector size and the baffle size, the baffle becomes a very thin cylinder over a very small detector, which collects light from a very small area along a very small solid angle, visualized in Figure 3 (right)⁴. In calculus terms, when we let the detector size and baffle’s solid angle approach 0, we obtain the quantity called **radiance**:

⁴It is just a visualization convention, but visualizing $d\omega$ as a cylinder rather than a cone makes it easier to imagine what $dA \cos \theta$ is like.

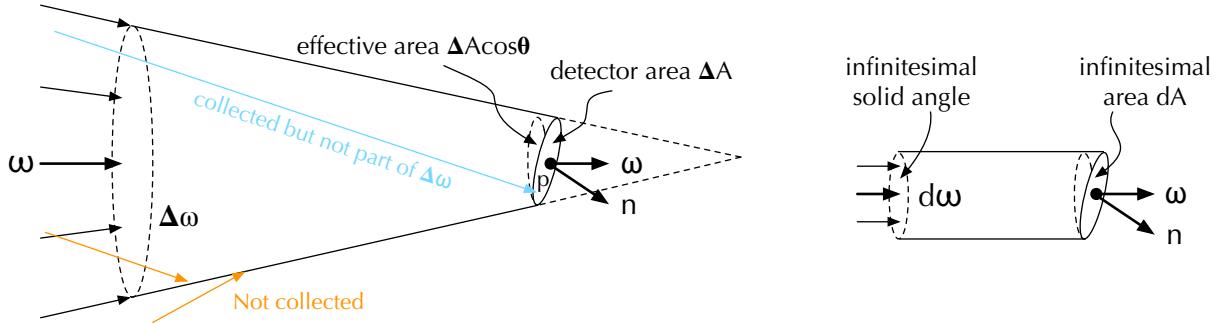


Figure 3: Left: the baffle limits the directions through which incident photons can be collected by the detector. As we reduce the solid angle of the baffle $\Delta\omega$ and the detector ΔA , the average power per unit “effective area” per unit solid angle approaches $L(p, \omega)$, the radiance at position p along direction ω . Right: intuitively we can think of a point (an infinitesimal area) receiving lights from a single direction (an infinitesimal solid angle) as just a tiny area intercepting a tiny cylinder.

$$L(p, \omega) = \lim_{\Delta\omega \rightarrow 0} \lim_{\Delta A \rightarrow 0} \frac{\Delta\Phi}{\Delta A \cos \theta \Delta\omega} \quad (8a)$$

$$= \frac{d}{d\omega} \frac{d\Phi(p)}{dA \cos \theta} = \frac{d^2\Phi(p)}{d\omega dA \cos \theta} \quad (8b)$$

$$= \frac{dE(p)}{d\omega \cos \theta} \quad (8c)$$

$$= \frac{dE_{\perp}(p)}{d\omega} \quad (8d)$$

Equation 8b is the definition of radiance, and it can be re-written to Equation 8c given the definition of irradiance (see Equation 4). Radiance is an intrinsic property of the radiation field, and the reason we have the $\cos \theta$ term in the definition is merely due to the way we have chosen to measure the property (using a detector that is θ -oriented). Radiance has a unit of Wm^2sr^{-1} , and *spectral* radiance has a unit of $Wm^2sr^{-1}nm^{-1}$.

Radiance is a density function: the density of power at a point along a direction. As with any density function, it is useful when it gets integrated to compute some other quantities. For instance, given the radiance $L(p, \omega)$, the irradiance at p is given by:

$$E(p, \Omega) = \int^{\Omega} L(p, \omega) \cos \theta d\omega. \quad (9)$$

Here we write the irradiance as $E(p, \Omega)$ to explicitly mean that the irradiance depends not only on the specific position p and the solid angle Ω over which the lights are coming.

Looking at the effective area in Figure 3, if the irradiance at the infinitesimal area p is $dE(p)$, the irradiance at the (infinitesimal) effective area, which projected from dA along ω , is $\frac{dE(p)}{\cos \theta}$.

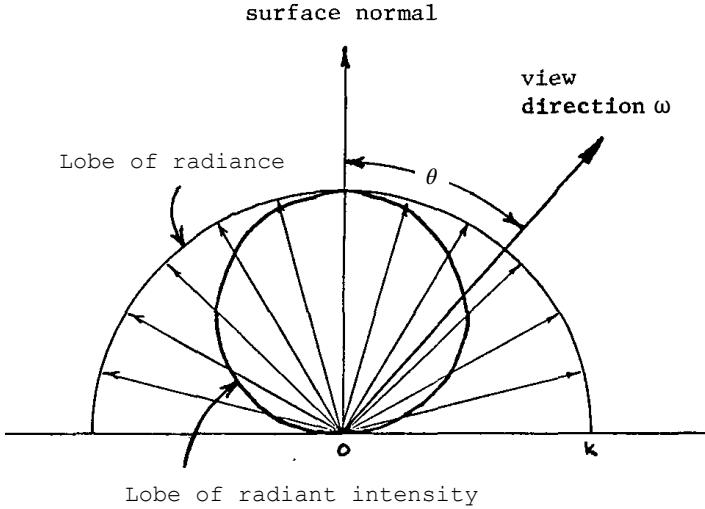


Figure 4: Comparison between the radiance distribution (constant w.r.t. viewing direction ω) and radiant intensity distribution (weakens by a factor of $\cos \theta$) of a Lambertian emitter/scatterer. Adapted from [Goral et al. \[1984\]](#), Fig. 2].

Combine this with Equation 8c, we can interpret the radiance $L(p, \omega)$ in a different way: it is the ratio between $dE_{\perp}(p)$, the infinitesimal irradiance defined at the surface perpendicular to the light direction, and the infinitesimal solid angle. This is shown in Equation 8d.

Using this interpretation of radiance, we can also give a more operational interpretation of Equation 9: we first calculate the infinitesimal irradiance $dE_{\perp}(p) = L(p, \omega)d\omega$ made by lights at the direction ω , then “transfer” that to the infinitesimal irradiance at the detector surface through the $\cos \theta$ factor, and then repeat this for all the directions and accumulate the contributions.

3.5 Lambertian Emitter, Radiant Intensity, and Lambert's Cosine Law

A **Lambertian emitter** or an ideal *diffuse emitter* is a flux-emitting point whose emitted radiance is constant regardless of the outgoing direction. A related concept is a **Lambertian scatterer** or an ideal *diffuse surface*, which is a surface point where the scattered radiance is independent of the scattering direction.

It might come as a surprise that the flux emitted by a Lambertian emitter through a fixed solid angle is different for different emission directions. Consider a setup where a Lambertian emitter has an infinitesimal area dA . The power emitted by dA toward its normal direction in an infinitesimal solid angle of $d\omega$ is $d\Phi_0 = L d\omega dA$, where L is the radiance. The power emitted toward an oblique direction ω through the same solid angle is $d\Phi_\theta = L d\omega \cos \theta dA$.

In radiometry, the ratio of infinitesimal power and infinitesimal solid angle is called the

radiant intensity⁵, denoted I :

$$I(\omega) = \frac{d\Phi}{d\omega}. \quad (10)$$

I is a meaningful measure only for a point source (e.g., our infinitesimal Lambertian emitted here). We can see that for a Lambertian emitter, the radiant intensity decays by a factor of $\cos \theta$: $\frac{d\Phi_\theta}{d\omega} = \frac{d\Phi_0}{d\omega} \cos \theta$. This is usually called the **Lambert's cosine law**, named after Johann Heinrich Lambert, from his Photometria [Lambert, 1760]. Similarly, if we have a Lambertian scatterer, its scattered radiant intensity will also decay by $\cos \theta$ as the polar angle θ of the viewing direction ω increases. Figure 4 compares the radiance distribution and radiant intensity distribution of a Lambertian emitter/scatterer. Both distributions are over the entire hemisphere but we show only a cross section. The distributions are visualized as two lobes, and the distance of a point on the lobe to the origin is proportional to the value at that point. The radiance distribution is constant regardless of ω but the radiant intensity is proportional to $\cos \theta$.

3.6 The Measurement Equation in Camera Imaging

We will study this more carefully in the imaging lecture, but given that we have the basic understanding of radiometry, it is probably a good time to show you how radiometry is of fundamental importance to camera imaging. For simplicity, let's just consider one single pixel.

Each pixel is very small, but it has a finite area, say A_p . Each pixel is constantly being bombarded by lights that enter the aperture, which has a size V . The raw pixel value is roughly proportional to the energy it receives during the exposure time⁶. So using the basic radiometry, we can write the total energy received by a pixel during the exposure time T as:

$$Q = \int^T \int^{A_p} \int^{\Omega(p,V)} L(p, \omega) \cos \theta \, d\omega \, dp \, dt, \quad (11)$$

where $\Omega(p, V)$ explicitly expresses that a solid angle is determined by the aperture V and a point p on the pixel surface. Of course this quantity changes with p . We sometimes omit p and V when it is clear what Ω refers to, but here since the solid angle changes with the dummy variable p in the integral equation we express it explicitly. This equation is sometimes called the **measurement equation** of an image sensor in computer graphics [Kolb et al., 1995; Reinhard et al., 2008, Chpt. 6.8.1].

We can see, again, that radiance is the most fundamental quantity. Knowing the radiance distribution in camera, we can in theory synthesize the value of any pixel and, thus, an entire image. How do we know the radiance distribution inside a camera? We can calculate it if we 1) know the radiance distribution just outside the camera and 2) the material properties of the

⁵Or simply, the “intensity”, which is an extremely overloaded term, so we will be verbose and use “radianc intensity.”

⁶Assuming there is no noise and there is no quantization error in converting analog signals to digital signals.

camera optics (e.g., lenses, filters, etc.). The effects of the camera optics are nothing more than surface scattering and volume scattering, so using the principles introduced in this Chapter we can in theory convert the radiance distribution outside the camera (in the scene) to that inside the camera.

3.7 Light Field

There is a name for the distribution of the radiance in the space — it is called the **light field**, which refers to the complete set of all the possible radiances flowing through every possible direction. Knowing the light field of a scene, we can in theory synthesize any image captured by any camera — limited only by the limitations of geometric optics. This is fundamentally because given the light field information, we can estimate the irradiance, power, and energy of any thing at any time at will.

The field of **light-field imaging** is concerned with capturing the light field of a scene. **Light-field display** is a 3D display technology that attempts to reproduce the light field of a scene, which is usually captured beforehand by some sort of light-field imaging technique. Reproducing the light field provides the depth information of a scene that is missing in conventional 2D displays, is one of the technologies to an immersive experience (other technologies include varifocal displays, multi-focal displays, and holographic displays).

Light-field rendering is concerned with rendering a new image/photo at a novel perspective given a set of photons taken at other perspectives. It is a form of *image-based* rendering. The idea is that each image capture is a sample of the light field followed by a low-pass filter (i.e., the integration in Equation 11). Rendering an image at a new perspective is a classic signal sampling and reconstruction problem, where the new image is nothing more than another sample of the light field. In this sense, many familiar tasks such as interpolating between video frames, panoramic photography, (stereoscopic) 360° video rendering are all light-field rendering in disguise.

As with any signal resampling task, the ideal solution to light-field rendering is to first estimate the underlying light field from a set of samples, and then re-sample the light field given the new perspective. Signal filtering is necessary for both signal reconstruction and anti-aliasing, and the name of the game used to designing good filters that are practically useful and computationally tractable. Of course, modern-day *image-based* rendering treats the whole problem as a deep learning/neural network problem and learns the light field from massive amount of data. The key seems to be figuring out a good representation of the 3D geometry of the scene.

3.8 Photometric Quantities

Spectral radiant flux (power), irradiance, radiant intensity, and radiance are all radiometric quantities. They all have a **photometric** counterpart, which weighs the radiometric quantity by the luminous efficiency function (LEF). The LEF, as we have discussed in the Color Vision Chapter, at a particular wavelength is inversely proportional to the radiometric quantity at each wavelength needed to produce the same level of perceptual brightness.

For instance, given a spectral radiant flux $\Phi(\lambda)$, the corresponding photometric counterpart is then:

$$\Phi_v(\lambda) = K\Phi(\lambda)V(\lambda), \quad (12)$$

where $\Phi_v(\lambda)$ is the spectral **luminous flux**, $V(\lambda)$ is the LEF, and K is a constant that, for historical reasons, takes the value of 683.002. The total luminous flux is then:

$$\Phi_v = \int_{\lambda} K\Phi(\lambda)V(\lambda)d\lambda. \quad (13)$$

Luminous flux has a unit of **lumen** (lm), so K has a unit of lm/W. We can also weigh the radiant power by the scotopic LEF, in which case $K = 1700$ (lm/W).

Other radiometric quantities can be similarly converted to the photometric counterparts. Specifically,

- the photometric counterpart of irradiance is **illumination**, which has a unit of lx = lm/m², which is also called the **lux**;
- the photometric counterpart of radiance intensity is **luminous intensity**, which has a unit of cd = lm/sr, which is called the **candela**;
- the photometric counterpart of radiance is **luminance**, which has a unit of lm/(m²sr) = cd/(m²), which is also called the **nit**.

Sometimes radiometric vs. photometric quantities are also called the radiant vs. luminous quantities. The way to interpret the photometric quantities is that they take into account the spectral sensitivity of a particular photodetector, which in our case is the photoreceptors on the retina. But if we use other detectors, such as an image sensor, we will have a different spectral sensitivity, and the corresponding photometric measurements will be different. We will study the spectral sensitivity of image sensors in later chapters.

A **radiometer** measures the absolute radiometric quantities, whereas a **photometer** reports photometric quantities. An image sensor and our retina can both be thought of as a photometer but the spectral sensitivities in the two cases are different, so the raw pixel readings and the photoreceptor responses are different even under an identical illumination.

4 Surface Scattering

When a group of photons arrive at a material surface, some of the photons might be immediately turned away (i.e., reflected) while others might penetrate into the material (i.e., refracted). The refracted portion causes SSS, which we will discuss next. In surface scattering, what we care about is the reflected portion. There are two properties we care about in surface scattering: the directions of the reflection and the energy along each direction. These two properties are captured by what is known as the BRDF of the surface-scattering point.

4.1 BRDF

Generally, the energy distribution of the surface scattering is captured by the Bidirectional Reflectance Distribution Function (**BRDF**) [Nicodemus et al., 1977]. Informally, it tells us how the incident energy from a particular direction is distributed to different exiting directions. The BRDF is parameterized by three parameters, a surface point p , the direction of light incident on p , denoted ω_i , and the direction of light leaving p , denoted ω_s . So the BRDF is usually written as $f_r(p, \omega_s, \omega_i)$.

The way to understand BRDF $f_r(p, \omega_s, \omega_i)$ is to consider the following. $L(p, \omega_s)$, i.e., the radiance leaving p toward ω_s , is dependent on the light incident on p . When the incident light on p comes from only the direction ω_i , the irradiance at p is zero, since the solid angle of a single direction ω_i is zero, so naturally $L(p, \omega_s)$ is 0 (assuming there is no other light hitting p). When p receives lights from a non-zero solid angle of directions $\Delta\omega_i$ (centered around ω_i), the irradiance of p is increased by $\Delta E(p, \omega_i)$. At the same time due to this increase in incident light, $L(p, \omega_s)$ is no longer zero; the increase in the radiance leaving p over ω_s is denoted $\Delta L(p, \omega_s)$.

As we increase $\Delta\omega_i$, both $\Delta E(p, \omega_i)$ and $\Delta L(p, \omega_s)$ increase. BRDF is defined as the ratio of the two *increments* when $\Delta\omega_i$ approaches 0 (when the radiance along all directions in $\Delta\omega_i$ can be thought of as a constant):

$$f_r(p, \omega_s, \omega_i) = \lim_{\Delta\omega_i \rightarrow 0} \frac{\Delta L(p, \omega_s)}{\Delta E(p, \omega_i)} = \frac{dL(p, \omega_s)}{dE(p, \omega_i)} = \frac{dL(p, \omega_s)}{L(p, \omega_i) \cos \theta_i d\omega_i}. \quad (14)$$

A Useful Approximation

Now assume that we illuminate p through a finite, but small, solid angle Ω_i . Turning the differential equation to an integral equation:

$$L(p, \omega_s) = \int_{\Omega_i} f_r(p, \omega_s, \omega_i) dE(p, \omega_i) \quad (15a)$$

$$= \int_{\Omega_i} f_r(p, \omega_s, \omega_i) L(p, \omega_i) \cos \theta_i d\omega_i \quad (15b)$$

$$\approx f_r(p, \omega_s, \omega_i) \int_{\Omega_i} L(p, \omega_i) \cos \theta_i d\omega_i \quad (15c)$$

$$= f_r(p, \omega_s, \omega_i) E(p, \Omega_i), \quad (15d)$$

$$f_r(p, \omega_s, \omega_i) \approx \frac{L(p, \omega_s)}{E(p, \Omega_i)}. \quad (15e)$$

The derivation proceeds as follows:

- The simplification from Equation 15b to Equation 15c assumes that the BRDF is a constant over all the directions in Ω_i .

- The integration in Equation 15c has no analytical solution, since we do not know the analytical form of $L(p, \omega_i)$, but we know the integration is just another way of expressing the total irradiance incident upon p over Ω_i , which is denoted as $E(p, \Omega_i)$. This gets us to Equation 15d.
- To be more rigorous, the integration in Equation 15c evaluates to $E(p, \Omega_i) + C$, where C is a constant. Given the boundary condition that $L(p, \omega_s) = 0$ when $E(p, \Omega_i) = 0$, we know $C = 0$, so C is omitted.

Ultimately, we can see from Equation 15e that the BRDF $f_r(p, \omega_s, \omega_i)$ can also be calculated as the ratio between the *absolute* radiance $L(p, \omega_i)$ and the absolute irradiance $E(p, \Omega_i)$ illuminated from a very small, but finite solid angle Ω_i . Another way to interpret this is that the so-calculated BRDF is the average BRDF over Ω_i . This derivation is useful for actually measuring a BRDF, which we will discuss in Chapter 5.2, where we will *have* to use a non-zero solid angle for illumination, because physically we just cannot illuminate a point through an infinitesimal solid angle $d\omega_i$.

Isotropic Material

A 3D direction ω expressed in the Cartesian coordinate system can also be expressed by two 2D planar angles in the spherical coordinate system: the polar angle θ and the azimuthal angle ϕ . So BRDF can also be parameterized as $f_r(p, \theta_s, \phi_s, \theta_i, \phi_i)$. A material is **isotropic** if its BRDF satisfies $f_r(p, \theta_s, \phi_s, \theta_i, \phi_i) = f_r(p, \theta_s, \phi_s + x, \theta_i, \phi_i + x)$ for any x . An intuitive way to think of an isotropic material is that its color does not change if you pick a point p and rotate the material about the normal vector at p , the color of p does not change. This is because rotation about the normal vector keeps θ_i and θ_s unchanged and varies ϕ_i and ϕ_s by the same amount.

The nice thing about an isotropic BRDF is it can be parameterized with one fewer degree of freedom: $f_r(p, \theta_s, \phi_s - \phi_i, \theta_i)$. This is because it is $(\phi_i - \phi_s)$ rather than the specific values of ϕ_s or ϕ_i that matter.

4.2 Reflectance and Albedo

The BRDF does not have to be a value between 0 and 1. Let's say that there is a 100 J of energy incident on a point coming from a solid angle $\Delta\omega_i$. That amount of energy is distributed across all the outgoing directions in the hemisphere, which forms a solid angle of $4\pi/2 = 2\pi$. So on average the energy exiting per direction is $\frac{100}{2\pi} J$, which clearly is greater than 1. This is not surprising, since BRDF is ultimately a *density* measure, a *distribution*, which is most meaningful when it is integrated to calculate some quantity. Integrating the BRDF gives a percentage/fraction measure between 0 and 1, i.e., reflectance, which we will discuss next.

For the energy to be conserved, the total outgoing energy at any point must not exceed that of the incident energy received by that point. Assume that a point p receives an irradiance dE_i from a direction ω_i over an infinitesimal solid angle $d\omega_i$, the outgoing radiance along the direction ω_s due to that irradiance is $f_r(p, \omega_s, \omega_i)dE_i$. Then the outgoing irradiance leaving p over an infinitesimal solid angle $d\omega_s$ around ω_s would be $f_r(p, \omega_s, \omega_i)dE_i \cos \theta_s d\omega_s$. If we

integrate all the outgoing directions, we get the total outgoing irradiance dE_o , which must not exceed the incident irradiance dE_i :

$$dE_o = \int_{\Omega} dE_i f_r(p, \omega_s, \omega_i) \cos \theta_s d\omega_s \quad (16a)$$

$$\Rightarrow \int_{\Omega} f_r(p, \omega_s, \omega_i) \cos \theta_s d\omega_s = \frac{dE_o}{dE_i} = \rho_{dh}(p, \omega_i) \leq 1. \quad (16b)$$

dE_i is independent of ω_s , so it can be hoisted out of the integration, which gets us Equation 16b, which holds for any arbitrary incident direction ω_i . ρ_{dh} is defined as the ratio between dE_o and dE_i . When Ω is the hemisphere, ρ_{dh} is called the **directional-hemispherical reflectance** in the computer vision and graphics literature, and is interpreted as the percentage of energy scattered by a point over the entire hemisphere given the incident light from a particular direction. Clearly, ρ_{dh} is a function of both p and ω_i and takes a value between 0 and 1.

Since we are dealing with geometric optics, the Helmholtz reciprocity holds:

$$f_r(p, \omega_s, \omega_i) = f_r(p, \omega_i, \omega_s), \quad (17)$$

which means the energy conservation can also be expressed as:

$$\int_{\Omega} f_r(p, \omega_s, \omega_i) \cos \theta_i d\omega_i = \rho_{hd}(p, \omega_s) \leq 1, \quad (18)$$

where ρ_{hd} is called the **hemispherical-directional reflectance** when Ω is the hemisphere. ρ_{hd} , a function of p and ω_s , is interpreted as the percentage of energy reflected toward a particular direction ω_s given the incident energy over the entire hemisphere.

Equation 18 can be derived by first re-writing Equation 16b as $\int_{\Omega} f_r(p, \omega_i, \omega_s) \cos \theta_s d\omega_s \leq 1$ (using the reciprocity) followed by switching ω_s and ω_i (simply a change of notation). This derivation suggests that $\rho_{hd}(p, \omega_i) = \rho_{dh}(p, \omega_s)$, a natural consequence of the reciprocity⁷.

We can also describe the relationship between all the outgoing irradiance E_o of a point over a solid angle Ω_s due to all the incident irradiance E_i over a solid angle Ω_i :

$$E_o = \int_{\Omega_s} (\int_{\Omega_i} f_r(p, \omega_s, \omega_i) L(p, \omega_i) \cos \theta_i d\omega_i) \cos \theta_s d\omega_s \leq E_i = \int_{\Omega_i} L(p, \omega_i) \cos \theta_i d\omega_i. \quad (19a)$$

$$\rho_{hh}(p) = \frac{E_o}{E_i} \leq 1. \quad (19b)$$

Equation 19b defines ρ_{hh} , which is called the **hemispherical-hemispherical reflectance** when both Ω_1 and Ω_2 are hemispheres. ρ_{hh} has another name: **albedo**. When $f_r(p, \omega_s, \omega_i)$

⁷For instance, if $\rho_{hd}(p, \omega_i) = \frac{1}{1+\omega_i^2}$, then $\rho_{dh}(p, \omega_s)$ must take the form $\rho_{dh}(p, \omega_s) = \frac{1}{1+\omega_s^2}$

is independent of (invariant to) ω_i and ω_s , i.e., when p is an ideal **Lambertian** surface (see Chapter 4.4), Equation 19a can be re-written as:

$$E_o = E_i \rho_{hh}(p) = \int^{\Omega_s} f_r(p, \omega_s, \omega_i) \left(\int^{\Omega_i} L(p, \omega_i) \cos \theta_i d\omega_i \right) \cos \theta_s d\omega_s \quad (20a)$$

$$= \int^{\Omega_s} f_r(p, \omega_s, \omega_i) E_i \cos \theta_s d\omega_s \quad (20b)$$

$$= E_i \int^{\Omega_s} f_r(p, \omega_s, \omega_i) \cos \theta_s d\omega_s \quad (20c)$$

$$= E_i \rho_{dh}(p, \omega_i), \quad (20d)$$

where Equation 20b is derived using the definition of E_i in Equation 19a, Equation 20c is derived since E_i is independent of ω_s , and Equation 20d is derived by using the definition of ρ_{dh} in Equation 16b.

We can see that for a Lambertian surface the albedo (ρ_{hh}) is equivalent to ρ_{dh} and ρ_{hd} , but this relationship is not true in general. We can also show that for a Lambertian surface, the BRDF is the constant $\frac{\rho_{hh}}{\pi}$. Starting from Equation 20c:

$$E_i \rho_{hh} = E_i \int^{\Omega_s} f_r(p, \omega_s, \omega_i) \cos \theta_s d\omega_s = E_i f_r(p, \omega_s, \omega_i) \int^{\Omega_s} \cos \theta_s d\omega_s = E_i f_r(p, \omega_s, \omega_i) \pi \quad (21a)$$

$$\Rightarrow f_r(p, \omega_s, \omega_i) = \frac{\rho_{hh}}{\pi} \quad (21b)$$

The derivation uses the integral results that:

$$d\omega = \sin \theta d\theta d\phi, \quad (22a)$$

$$\int^{\Omega} \cos \theta d\omega = \int_0^{2\pi} \int_0^{\pi/2} \cos \theta \sin \theta d\theta d\phi, \quad (22b)$$

$$= 2\pi \int_0^{\pi/2} \cos \theta \sin \theta d\theta \quad (22c)$$

$$= \pi, \quad (22d)$$

when Ω is the hemisphere.

You might be thinking that, mathematically, for Equation 20a to hold f_r just needs to be independent of ω_i , but not ω_s , so do we really need to assume a Lambertian surface here? It turns out that if f_r is independent of ω_i it must also be independent of ω_s (can you prove this?⁸) and, thus, must be a constant (i.e., $\frac{\rho_{hh}}{\pi}$ in Equation 21b) for a given p .

⁸One informal way to do so is the following. Since $f_r(p, \omega_s, \omega_i)$ is independent of ω_i , let's re-write it as $g(p, \omega_s)$. Now we invoke the reciprocity and rewrite $f_r(p, \omega_s, \omega_i)$ as $g(p, \omega_i)$. The only way for $g(p, \omega_s) = g(p, \omega_i)$ is for g to be dependent only on p .

Finally, one can also define the **directional-directional reflectance**, which is naturally a function of both the incident direction and outgoing direction and can be defined as the ratio between the incident irradiance and the outgoing irradiance when both the incident and outgoing solid angles approach 0.

To compare the BRDF and directional-directional reflectance, both are sensitivity to both the incident and outgoing directions. But the former is a density measure whereas the latter is a fraction/percentage measure. Integrating BRDF over a finite set of directions gives us some measure reflectance. This is why the BRDF is defined as the radiance/ irradiance ratio rather than radiance/radiance or irradiance/irradiance ratio; it is to reflect the fact that the energy of a small cone of incident directions is distributed over all the directions over the hemisphere and what we care to characterize is the *distribution* of the incident energy over all outgoing directions.

4.3 The Rendering Equation

Given the BRDF, we can estimate the outgoing radiance of a point given its illumination using the well-known **Rendering Equation**.

The setup is that we have a surface on which there is a point p that is receiving lights from a solid angle Ω . We are interested in calculating the exiting radiance leaving p toward an arbitrary direction ω_s . The rendering equation formulates this calculation by:

$$L(p, \omega_s) = \int_{\Omega} f_r(p, \omega_s, \omega_i) L(p, \omega_i) \cos \theta_i d\omega_i, \quad (23)$$

where $L(p, \omega_s)$ is the outgoing radiance from p toward the direction ω_s ; Ω is usually a hemisphere in surface scattering, since lights hitting a surface point can come from anywhere in the hemisphere, in which case Equation 23 is also called the *reflection equation*, indicating the fact that the equation governs surface reflection/scattering.

The rendering equation is exactly the same equation in Equation 15b, so there is nothing more profound about the rendering equation than the definition of the BRDF: we are simply following the BRDF's definition and turning the differential equation into an integral one. Intuitively, the way to understand this equation is that every ray that hits p makes some contribution toward the outgoing radiance $L(p, \omega_s)$, and the integration just accumulates all the contributions. In particular:

- $L(p, \omega_i) d\omega_i$ is the incident irradiance of a differential solid angle $d\omega_i$; note that the irradiance calculated here is defined with respect to a surface perpendicular to the direction of ω_i .
- $L(p, \omega_i) \cos \theta d\omega_i$ applies the Lambert's cosine law and calculates the irradiance at the surface where p lies.
- $f_r(p, \omega_s, \omega_i) L(p, \omega_i) \cos \theta d\omega_i$ “transfers” the differential incident irradiance to the differential outgoing radiance toward ω_s through the BRDF function.

- The integration over all the incident direction calculates the total outgoing radiance given all the incident lights.

The rendering equation in theory allows us to calculate the entire light field, i.e., the radiance distribution in space, given an arbitrary p and ω_s . Why is knowing the light field important? Recall Equation 11: knowing the light field allows us to synthesize any image or calculate the color of any object from any perspective.

It is of course much easier said than done when it comes to solving the rendering equation, which itself is worth multiple chapters in a computer graphics textbook. We will not get into it here; let's just consider the following challenges. First, the integrand in Equation 23 generally has no analytical form, so we will not be able to get an analytical solution to the integral equation. A common method is Monte-Carlo integration, which samples the integrand at different points and estimates the integral from the samples.

Second, in a realistic environment we need to solve the rendering equation *recursively*. Note how the radiance function shows up on both sides of the equation. Put it in another way, when using Monte-Carlo integration to solve Equation 23 we need to sample the value of $L(p, \omega_i)$ for a specific ω_i — how? We evaluate Equation 23 again, but this time treating ω_i as the ω_s , which means we invoke Monte Carlo integration again. You can see how this can quickly blow up the computation: the number of rays whose radiances we need to calculate exponential increases as long as we need to sample more than ray at each point. A big chunk of physically-based graphics is devoted to addressing this issue; the most commonly used strategy is called **path tracing**, for which Pharr et al. [2023, Chpt. 13] is a great reference.

Another way to think of this is that there are infinitely many paths through which lights can propagate and incident on a point. A **global illumination** method for rendering would attempt to track all these paths (e.g., through Monte Carlo methods). In contrast, a **local illumination** method is concerned with only a small subset of these paths, in which case we might be able to evaluate the rendering equation as a single-pass integration while avoiding recursion. For instance, we might consider lights only from direct light sources. We will see the counterpart of this exact situation in surface scattering/volume rendering in Chapter 9. For this reason, the rendering equation is sometimes called the **light transport equation** (LTE), because it in principles captures how light is transported in space⁹.

Finally, we also need to somehow know the BRDF of the material. There are generally two methods of going about it. We can of course measure it, but we have no realistic way of measuring the complete BRDF for a material, because we would have to measure infinitely many points and, for each point, infinitely many incident and outgoing directions. We can only sample the BRDF, but there is still a massive amount of samples we need to take and to store. We will discuss the technical instrument and calculations that go behind measuring and calculating the BRDF in Chapter 5.2.

Another approach is to parameterize the BRDF so that we can evaluate the BRDF on-demand rather than storing all the BRDF data, and this is what we will discuss next.

⁹To be exact, the LTE sometimes has an emission term at the right-hand side to denote the spontaneous emission from a surface point.

4.4 Specular vs. Glossy vs. Diffuse Materials

In everyday life material surfaces are usually classified into being specular, glossy, or diffuse. Figure 5 shows examples of the three materials. We can now give a more rigorous treatment of these material types using BRDF, which will, in turn, give us some inspirations for parameterizing the BRDF.

Perfectly Specular Material

If a surface is perfectly smooth, like a mirror, it is called a **perfectly specular** material. Such materials follow the **Snell's law**, which governs the angles of reflection and refraction, and the **Fresnel equations**, which govern the energy of reflection and refraction.

In the plane of incidence (the plane uniquely determined by the incident direction and the surface normal), the reflection direction is symmetric about the surface normal as the incident direction. More precisely, if the incident direction is ω_i (parameterized by the polar angle θ_i and azimuthal angle ϕ_i) and the reflection direction is ω_s (θ_s, ϕ_s), we have:

$$\theta_s = \theta_i, \quad (24a)$$

$$\phi_s = \phi_i + \pi. \quad (24b)$$

The refraction/transmitted direction ω_t (θ_t, ϕ_t) follows:

$$n_1 \sin \theta_i = n_2 \sin \theta_t, \quad (25a)$$

$$\phi_t = \phi_i + \pi, \quad (25b)$$

where n_1 is the refractive index of the medium where lights come and n_2 is that of the medium that reflects/refracts the lights.

The energy of the reflected and refracted light is governed by the Fresnel equations. We will spare you the details, but it suffices to say that the fractions of reflected/refracted light are dependent on the incident angle, refractive indices of the two interface media, and the polarization states of the light. If you work out the math and assume that the incident light is unpolarized, the percentage of reflected energy $F_r(\omega_i)$ for an incident direction ω_i is given by:

$$F_r(\omega_i) = \frac{r_a + r_e}{2}, \quad (26a)$$

$$r_a = \left(\frac{n_2 \cos \theta_i - n_1 \cos \theta_t}{n_2 \cos \theta_i + n_1 \cos \theta_t} \right)^2, \quad (26b)$$

$$r_e = \left(\frac{n_1 \cos \theta_i - n_2 \cos \theta_t}{n_1 \cos \theta_i + n_2 \cos \theta_t} \right)^2. \quad (26c)$$

We call $F_r(\omega_i)$ the **specular reflectance**, which not only varies with ω_i but also is also a spectral term; we omit the wavelength for simplicity. Assuming no loss of energy, the specular transmittance, i.e., the fraction of the transmitted energy, is given by $1 - F_r$.

Fresnel's equations are best understood in the context of the electromagnetic theory and are derived by treating lights as waves in an *electric field* (the fact that we need to consider polarization states of a light is a giveaway). While F_r cannot be derived from radiometry, it is fundamentally about the energy transfer of surface scattering, which radiometry is also concerned with. So F_r can be integrated into the radiometry framework. One good example is to express the BRDF of a specular material using F_r :

$$f_r(p, \omega_s, \omega_i) = F_r(\omega_i) \frac{\delta(\theta_s - \theta_i)\delta(\phi_s - \phi_i - \pi)}{\cos \theta_i}, \quad (27)$$

where $\delta(x)$ is the Dirac delta function, which is 0 everywhere except when $x = 0$ and has the property $\int \delta(x)dx = 1$.

We can verify that this BRDF makes sense. First, the BRDF is non-zero only when Equation 24 holds because of the double-delta term. Second, the energy conservation is followed. For instance, if we calculate the directional-hemispherical reflectance by plugging the BRDF into Equation 16b and assuming Ω is hemisphere, we get:

$$\frac{E_o}{E_i} = \rho_{dh}(p, \omega_i) = \int^{\Omega} F_r(\omega_i) \frac{\delta(\theta_s - \theta_i)\delta(\phi_s - \phi_i - \pi)}{\cos \theta_i} \cos \theta_s d\omega_s \quad (28a)$$

$$= F_r(\omega_i) \int^{\Omega} \frac{\delta(\theta_s - \theta_i)\delta(\phi_s - \phi_i - \pi)}{\cos \theta_i} \cos \theta_s d\omega_s \quad (28b)$$

$$= F_r(\omega_i). \quad (28c)$$

Since $F_r(\omega_i)$ is independent of ω_s , Equation 28a evaluates to Equation 28b. The integration in Equation 28b evaluates to 1. This is because, informally, the integrand is non-zero only when Equation 24 holds, at which point $\theta_s = \theta_i$, so the cosine terms cancel out. So the integration is just sort of a hugely complicated way of writing $\int \delta(x)dx$, which is 1.

We can see that the specular reflectance F_r is equivalent to ρ_{dh} , the directional-hemispherical reflectance. This makes sense, because in specular materials the scattering is directional if the incident light is directional. So the directional-hemispherical reflectance reduces to the “directional-directional” reflectance, which is essentially the specular reflectance.

The specular reflectance is also equivalent to the hemispherical-directional reflectance ρ_{hd} . We can show this either by simply invoking the reciprocity that $\rho_{hd} = \rho_{dh}$, or by plugging the BRDF into Equation 18 and obtaining (assuming Ω is hemisphere):

$$\rho_{hd}(p, \omega_s) = \int^{\Omega} F_r(\omega_i) \frac{\delta(\theta_s - \theta_i)\delta(\phi_s - \phi_i - \pi)}{\cos \theta_i} \cos \theta_i d\omega_i \quad (29a)$$

$$= F_r(\hat{\omega}_s) = F_r(\omega_s), \quad (29b)$$

where $\hat{\omega}_s(\theta_s, \phi_s - \pi)$ is the mirror-reflection direction of $\omega_s(\theta_s, \phi_s)$. The integral evaluates to $F_r(\hat{\omega}_s)$ because, informally, the integrand is non-zero only when Equation 24 holds, at which

point $\omega_i = \hat{\omega}_s$ so $F_r(\omega_i) = F_r(\hat{\omega}_s)$; the integral is a complicated way of writing $\int F_r(\omega_i) \delta(\hat{\omega}_s - \omega_i) d\omega_i$, which evaluates to $F_r(\hat{\omega}_s)$. The result has an intuitive explanation: for a specular surface, the scattered energy along ω_s given a hemispherical illumination is the same as when the illumination comes only from $\hat{\omega}_s$. We can then show that $F_r(\hat{\omega}_s) = F_r(\omega_s)$, which is not surprising given reciprocity; you can also verify it by going through the equations in Equation 26.

Interestingly, the specular reflectance F_r in general is *not* equivalent to the hemispherical-hemispherical reflectance ρ_{hh} . To see this, plug the specular BRDF into Equation 19a (assuming Ω_i and Ω_s are hemispheres):

$$E_o = \int^{\Omega_s} \left(\int^{\Omega_i} f_r(p, \omega_s, \omega_i) L(p, \omega_i) \cos \theta_i d\omega_i \right) \cos \theta_s d\omega_s \quad (30a)$$

$$= \int^{\Omega_s} (F_r(\omega_s) L(p, \omega_s)) \cos \theta_s d\omega_s \quad (30b)$$

$$= \int^{\Omega_i} F_r(\omega_i) L(p, \omega_i) \cos \theta_i d\omega_i, \quad (30c)$$

$$E_i = \int^{\Omega_i} L(p, \omega_i) \cos \theta_i d\omega_i. \quad (30d)$$

We can see that only when $F_r(\omega_i)$ is a constant do we get $F_r(\omega_i) = \frac{E_o}{E_i} = \rho_{hh}$. Interestingly, when $F_r(\omega_i)$ is constant, the specular material is isotropic (can you prove it?). Since $F_r(\omega_i)$ do not have to be a constant, specular materials could be anisotropic. That is, it is theoretically possible that a material always reflects specularly, but the reflected energy depends on the incident direction.

Diffuse Material

When the surface is rough, the energy of surface reflection deviates away from the perfect mirror-like reflection and, instead, distributes across the hemisphere. When the surface becomes rough enough, the distribution of outgoing energy can become uniform across all outgoing directions over the entire hemisphere. Such a surface is called a **diffuse** or an ideal **Lambertian** surface. The perfect Lambertian surface does not exist, but many things in the real world come close, such as paper, marble, or wood.

The BRDF is a Lambertian surface is a uniform function. As we have seen in Equation 21b, $f_r(p, \omega_s, \omega_i) = \frac{\rho_{hh}}{\pi}$ when ρ_{hh} is the surface albedo and is between 0 and 1. It is easy to see that diffuse materials are always isotropic.

Glossy Material

The surface scattering in most materials is in-between being perfectly specular and perfectly diffuse. These materials scatter lights to a small cone of directions, usually centered around the direction of a perfect reflection. These materials are usually called **glossy** or sometimes, confusingly, “specular” too. The energy distribution of a glossy material is neither a Delta

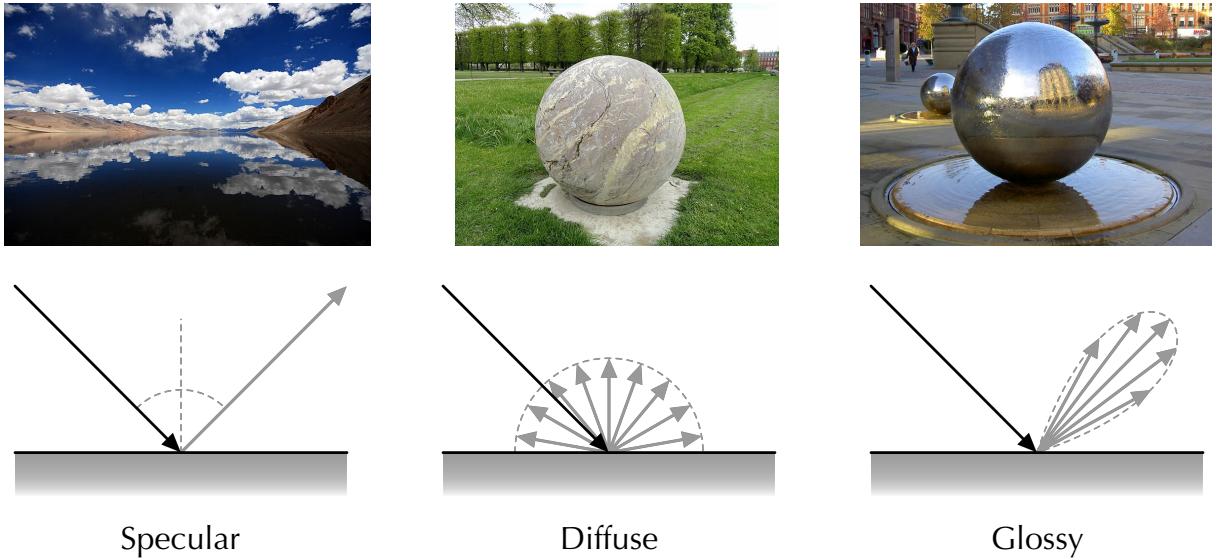


Figure 5: Left: a specular material and its BRDF. Middle: a diffuse material and its BRDF. Right: a glossy material and its BRDF. From [Prabhu B Doss \[2007\]](#); [User:Daderot \[2012\]](#); [Steve Fareham \[2007\]](#); [VonHaarberg \[2018c,a,b\]](#).

function (as in the perfectly specular case) not a uniform function (as in the diffuse case). It usually is a function that peaks at the mirror-reflection direction and gradually decays as we move away from that direction.

The bottom figures in Figure 5 illustrate an example of the BRDF for each of the three surface types — under a given incident direction. An actual BRDF (for a given surface point and a given incident direction) would be a 3D shape, and what we are showing here is the cross section. The shape of the locus is drawn to be proportional to the magnitude of the BRDF; the locus in graphics literature is sometimes called the *specular lobe*.

The spectral-lobe visualization give us a hint: we can parameterize a BRDF by mathematically describing the shape of the specular lobe. In fact, the BRDFs for the Lambertian surface and for the specular materials are two such examples; see Equation 27 and Equation 28. A glossy BRDF is more difficult to parameterize. Many BRDF parameterizations have been proposed; some are empirical while others attempt to be physically plausible. The most popular and widely used is based on the microfacet model, which we will discuss next.

4.5 BRDF Parameterization with Microfacet Models

The assumption of the microfacet model is that the surface scattering behavior of a point depends on its local roughness: the rougher the surface, the more diffuse the surface scattering becomes. To model the roughness, the surface is modeled as a collection of small microfacets, each of which acts like a perfect mirror. A specular surface is one where all the microfacets have the exact same orientation. As the surface becomes rougher, the mirrors become more randomly

oriented. When the mirrors are completely randomly oriented, the resulting surface scattering becomes diffuse.

To derive a microfacet model, we need to first define the orientation of each microfacet. Given a beam of incident lights from a particular direction, we can then trace, following the laws governing specular reflection, how the lights are scattered by the collection of the microfacets given their orientations. In the end, we obtain the collection of outgoing directions, from which we can derive the BRDF.

There are many variants of the microfacet model. They have one thing in common: they do not explicitly model the scattering of each ray at each microfacet but, rather, model the scattering of the microfacets *statistically* given the *distribution* of the microfacet orientations. In the end, they can either have an analytical form of the BRDF (Lambertian surface being an extreme example), have a close approximation of the analytical form, or can numerically estimate the BRDF efficiently (mostly through sampling).

Without going into the details, we will refer you to [Pharr et al. \[2023, Chpt. 9.6\]](#) for a mathematically treatment of the general idea and to [Torrance and Sparrow \[1967\]](#); [Cook and Torrance \[1982\]](#); [Ward \[1992\]](#); [Oren and Nayar \[1995\]](#); [Walter et al. \[2007\]](#) for the classical models.

Microfacets Models are Discrete Models Applied to a Continuous Domain

If the microfacet theory does not sound weird to you, it should!

In a microfacet model, we are still modeling surface scattering using discrete objects (microfacets) and events (perfect mirror-like reflection on each microfacet). Is it surprising that we can use the discrete microfacet model to reason about the behavior of a *continuous* surface? Given any point p on a surface, wouldn't p correspond to one single microfacet and the behavior of p simply be the result of a perfect mirror reflection there? If so, how can the microfacet model describe non-specular surface scattering of glossy and diffuse materials?

An intermediate answer is that the microfacet theory is just a modeling methodology. We use a set of discrete microfacets to derive the surface-scattering statistics of that set of microfacets, but then simply assume that the so-derived statistics applies anywhere on a continuous surface of interest. Still, does this methodology reflect the physical reality?

Well, the physical world is fundamentally not continuous; when we break down the surface into finer and finer scales we eventually get to molecules and atoms, so the surface property undergoes wild fluctuations depending on whether a small area contains molecules or not. If that is the level of detail you want to get into, you have to model things at the molecular and atomic levels (or even lower). Figure 6 illustrates this idea.

Fortunately for many real-world use-cases, we do not have to go there. Our eyes have a resolution limit, so we cannot resolve the details of a tiny surface area anyways. The just-resolvable area δA , set by the spatial resolution limit of our visual system, is more than large enough that it contains many microfacets, so the aggregated behavior of those microfacets can effectively model the observed scattering of δA , which is all that matters to our vision (and to computer graphics and imaging, which is concerned only with satisfying human vision). So effectively what the microfacet theory does is to assume that the small δA (which contains a

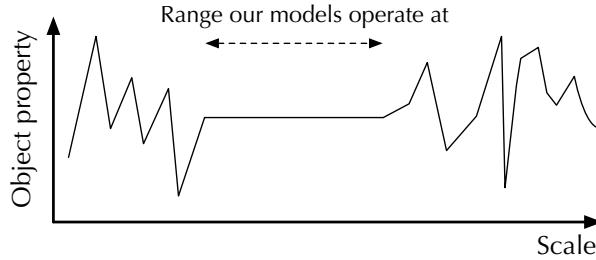


Figure 6: Triphasic profile of object property. Object property at both the macroscopic scale and at the atomic/molecular scale fluctuates wildly, but there is a scale where the property does not fluctuate very much. Models based on radiometry operate at this scale. This scale is sufficiently small (smaller than the spatial resolution of human vision and typical cameras) so our calculus machinery can be applied, but still larger than individual molecules and atoms so that we do not have to worry about the wild fluctuations at that scale.

distribution of microfacets) is just within the range where the surface scattering property is stable.

This way of modeling and thinking is pervasive in radiometry, which uses differential and integral equations and thus has inherently assumed that the radiation field under modeling is continuous. That is not true. Take irradiance as an example. The average irradiance of a surface changes dramatically at the microscopic level when we initially reduce the surface area, because the photon distribution over a large area is likely very non-uniform. When the surface area is sufficiently small, the number of photons hitting the surface will change proportionally with the surface area, because at that scale the photon distribution is roughly uniform. This is the scale at which irradiance is defined. But if we keep reducing the area smaller and smaller, the amount of photons hitting a tiny area will, again, undergo wild fluctuations depending on whether there are photons in the area or not — photons are discrete packets of energy. We will see another example shortly in volume scattering, where we use a small volume of discrete particles to build a model for radiative energy transfer, which we then apply to any given point in a continuous volume.

Orthogonal to the discussion above is the limitation that microfacet models do not account for the surface roughness on the scale of the light wavelength. In the regime where the length of each microfacet is comparable with the light wavelength, diffraction takes place. As a result, reflection does not follow the Snell's law and is wavelength dependent. In fact, this is how we get iridescence; in engineering people make diffraction gratings that take advantage of the wavelength dependency to disperse lights of different wavelengths.

5 Measuring Spectral Reflectance and BRDF

This section talks about measuring the spectral reflectance or spectral BRDF. It is absolutely important to note that the measured reflectance is not necessarily attributed only to surface

scattering, because the measurement setup does not care what the material being measured is. If SSS plays a role (e.g., translucent materials), the resulting reflectance data would include the contribution from volume scattering, too.

Worse, for these materials not all the SSS influences are captured by this measurement geometry, since some back-scattered photons will exit at other surface points, which will not be captured by the detector. So the measurement is *neither complete nor sound* for materials where back-scattered photons contribute to their reflectance.

5.1 Measuring Spectral Reflectance

How do we know the spectral reflectance (transmittance) of a material? We measure it. This is easier said than done. We will focus on the reflectance measurement here, but transmittance is measured similarly, except you are not measuring from the same side of the illuminant, but from the other side. [Sharma \[2003, Chpt. 1.11.4\]](#), [Trussell and Vrheil \[2008, Chpt. 8.7\]](#), and [Reinhard et al. \[2008, Chpt 6.8\]](#) have overviews of various measurement devices that might be helpful.

The Importance of Measurement Geometry

Consider Figure 1 (a) again. The illuminant emits lights everywhere, but what matters is the light incident on the point p the viewer is currently gazing; of course the incident lights could come from everywhere else in the space, not just a particular illuminant. Similarly, p could potentially scatter lights everywhere over the hemisphere (through surface scattering and/or SSS), but it is the small beam of light that enters the viewer's eye that matters. In order to measure the reflectance that is relevant to this particular illumination-viewing geometry, we need to 1) measure all the illuminating power that hits p and 2) measure the scatter light from p only along the viewing direction.

You can imagine that if we change the illumination to be, say, a diffuse lighting where there is an equal amount of light hitting p from all directions, the reflectance would be different, and it would be a perfectly relevant reflectance measure to report. If you have not, next time when you visit an art museum, pay attention to how the lighting system is carefully set up to bring out the best viewing experience (while also considering conservation); you ideally want the reflectance measurement of an artifact to simulate the viewing lighting.

Single Reflectance Measurement

In general there really is no single reflectance number we can associate with a material. There are two ways to approach this. A common approach is to set up the measurement geometry so that it is close to an actual viewing experience. Figure 7 (a) shows four common settings. Some might illuminate the material from 0° (assuming the direction of the surface normal has an angle of 0°) and then measure the scattered lights at 45° ; others can illuminate the material using diffuse illumination and measure the reflectance at 0° (see [Judd and Wyszecki \[1975, p. 122-125\]](#), [Reinhard et al. \[2008, Chpt. 6.8.2\]](#), and [Li \[2003, Chpt. 2.2.2\]](#)).

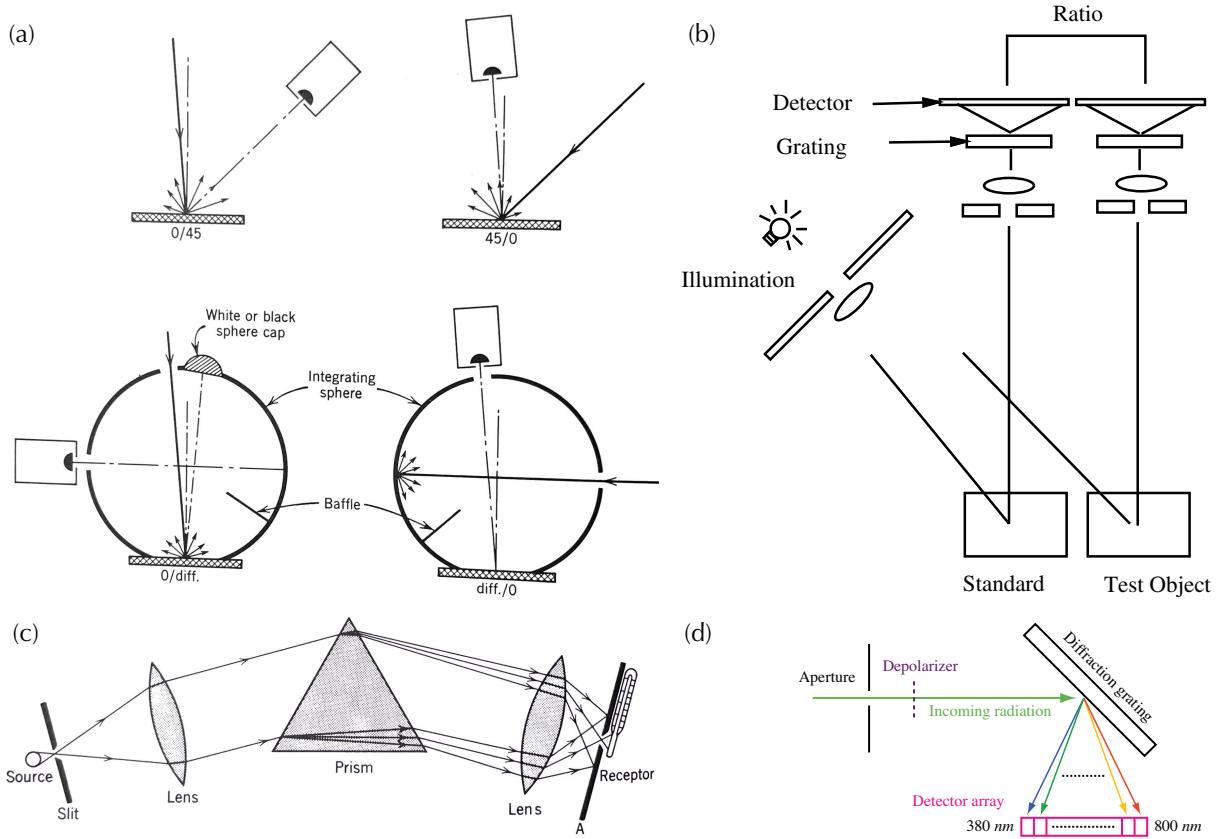


Figure 7: (a): Four different illumination-viewing geometries to measure the reflectance of a material; from [Judd and Wyszecki \[1975, Fig. 2.11\]](#). (b): A spectrophotometer, which takes two spectroradiometric measurements of the standard material with a known reflectance and a test material to calculate the spectral reflectance of the test material; from [Sharma \[2003, Fig. 1.33\]](#). (c): A spectroradiometer design, which measures the spectral power distribution of a light source (self-luminous or scattering) using a prism; from [Judd and Wyszecki \[1975, Fig. 2.1\]](#). (d): Another way to implement the spectroradiometer that uses diffraction grating to disperse incident light; from [Reinhard et al. \[2008, Fig. 6.22\]](#).

To get a reflectance spectrum, we need to know the reflectance at each sampled wavelength. There are multiple ways to go about measuring the spectral information. For instance, we can place a monochromator or a set of optical filters between the illuminant and the material so that we can control the wavelength of the light that is incident on the material.

Alternatively, we can change the detector to measure spectral information. We can use a dispersive medium such as a prism, shown in Figure 7 (c), or a diffraction grating, shown in Figure 7 (d), to separate the scattered light into different wavelengths and measure them individually. A detector that is capable of measuring the spectral radiometric quantities (e.g, the spectral power distribution) is called a **spectroradiometer**.

The raw detector readings of a spectroradiometer are usually not the absolute radiometric quantity of interest. The raw recording is, instead, roughly proportional to radiometric quantity up to a constant scaling factor $k(\lambda)$, which is usually called the detector's spectral sensitivity or responsivity function (Chapter 3.8), which we will study carefully in the image sensor lecture. $k(\lambda)$ can be calibrated offline, and that allows us to turn a detector's raw recording to the corresponding absolute radiometric quantity.

We take a spectroradiometric measurement of the illumination hitting the material and that of the scattered light of interest; the ratio is the spectral reflectance $\rho(\lambda)$:

$$\rho(\lambda) = \frac{\Phi_s(\lambda)K(\lambda)}{\Phi_i(\lambda)K(\lambda)} = \frac{\Phi_s(\lambda)}{\Phi_i(\lambda)}, \quad (31)$$

We can see that for reflectance measurement, the exact values of $k(\lambda)$ are immaterial. A curious question is that, while the detector can measure $\Phi_s(\lambda)$, what measures $\Phi_i(\lambda)$? One strategy is to, offline, place the same detector where the material is and directly measure $\Phi_i(\lambda)$ there.

Another, perhaps much more common and standard, way to measure spectral reflectance is to use something called **spectrophotometer**. This method does not need to know $\Phi_i(\lambda)$, but requires a reference sample with a known spectral reflectance. This is shown in Figure 7(b). It takes two spectroradiometric measurements under the identical illumination: one for the test material and the other for the standard/reference sample. The spectral reflectance of the test material $\rho_t(\lambda)$ is given by:

$$\rho_t(\lambda) = \frac{m_t(\lambda)}{m_s(\lambda)} \rho_s(\lambda), \quad (32)$$

where $\rho_s(\lambda)$ is the known spectral reflectance of the standard/reference sample, $m_s(\lambda)$ and $m_t(\lambda)$ refer to the raw detector readings of the standard and the test material at wavelength λ , respectively. We can see that the spectrum of the illumination does not matter. Sometimes $\frac{m_t(\lambda)}{m_s(\lambda)}$ is called the **spectral reflectance factor** of the test material if the reference material is perfectly diffuse [Judd and Wyszecki, 1975, p. 93].

In practice, the reference measurement can be done separately rather than simultaneously with the test material to reduce the device form factor, and the reference measurement data can be tabulated to save measurement time.

One note on terminology: while a spectroradiometer is used to measure the spectral radiometric quantities (e.g., spectral radiance), a spectrophotometer does not measure the spectral photometric quantities (e.g., spectral luminance); instead, it measures the spectral reflectance. This is standardized in American Society for Testing and Materials (ASTM) E284-13b [ASTM International](#) (along with other terminologies related to material properties and measurement instruments).

The nice thing about the approach described so far is that you get a single reflectance spectrum, but be very careful under what measurement geometry is the spectrum obtained. There is no guarantee that a particular measurement geometry corresponds to the illumination/observation geometry of an actual viewing experience, so use the reported reflectance data with that caveat in mind.

Goniometric Measurements

A more general approach is to measure the reflectance at every illumination-vs-viewing direction combination. For that we need what is called a **goniospectrophotometer**¹⁰. Figure 8 shows one such setup. The illuminant/light source incident on the material comes through the small aperture I , and the scattered light from the material is captured by a detector (e.g., a photodiode or, essentially, a single-pixel image sensor) through another aperture V . Transmittance can be similarly measured by placing the detector at the other side of the material.

The idea is to simultaneously sample, say, N illumination directions (parameterized by the azimuth ϕ_i and polar angle θ_i) and M scattering directions (parameterized by the azimuth ϕ_s and polar angle θ_s), and obtain $M \times N$ measurements, each of which corresponds to one particular combination of the illuminant and scattering directions. For convenience, commercial goniometric measurements usually use a beam splitter to simultaneously measure the illumination and scattering flux [[Lanevski et al., 2022](#); [Rabal et al., 2012](#)].

Denote the area on the material being measured A_r . The size of the area is dictated by the illumination aperture I . Assuming the power received by A_r from the illuminant through I is $\Phi_i(\lambda, A_r, I)$, and the power scattered by A_r and collected by the detector through the aperture V is $\Phi_s(\lambda, A_r, V)$, the reflectance of the small area A_r is simply given by:

$$\rho(\lambda, A_r) = \frac{\Phi_s(\lambda, A_r, V)}{\Phi_i(\lambda, A_r, I)}. \quad (33)$$

As the two apertures become very small, A_r becomes very small and the incident and outgoing solid angles become very small, too. The resulting reflectance measurement can be thought of as estimating the **directional-directional reflectance** (Chapter 4.2). But in general you can see how the reflectance number can easily change when we slightly vary the hardware setup. For instance, if we increase the detector aperture V , the detected power will increase, and that would increase the resulting reflectance. If we increase the illumination aperture I , the resulting reflectance would be for a larger material area A_r .

¹⁰“gonio-” comes from the Greek word $\gamma\omegaνία$ (gōnia), which means angle. There are also **gonioradiometers**, which measure the spectral radiometric quantities from different viewing directions.

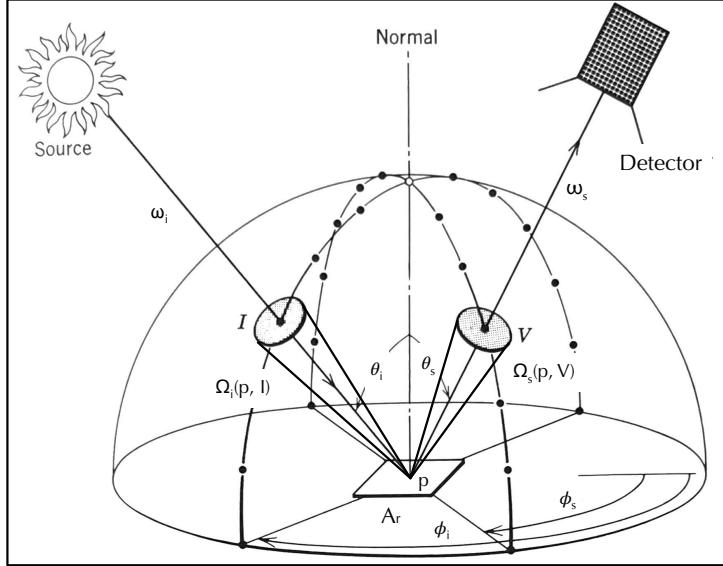


Figure 8: A setup for measuring goniometric reflectance and BRDF. Both the illuminant (source) and the detector (photometer) can vary in two degrees of freedom, (θ_i, ϕ_i) for the source and (θ_s, ϕ_s) for the detector, covering different illuminant-scattering combinations. Adapted from [Judd and Wyszecki \[1975, Fig. 3.4\]](#).

One can also use a reference material (with known reflectance spectra at the same measurement geometries) so as to avoid measuring $\Phi_i(\lambda, A_r, I)$, similar to how a spectrophotometer is operated.

5.2 Measuring BRDF

Reflectance is integrated from the BRDF, which suggests that the latter is a more fundamental measure of material property. The same setup shown in Figure 8 can also be used to measure the BRDF, in which case the setup is called a **goniospectroreflectometer**. We will take the same measurements, but with a bit more calculations we can estimate the BRDF of the material, rather than just the (goniometric) reflectance spectra.

Let us be precise about the setup (omitting the λ term in all relevant quantities).

- We are illuminating a small area A_r through the illumination aperture I .
- The center of A_r is an infinitesimal point p , which along with I subtends a solid angle $\Omega_i(p, I)$.
- ω_i is the direction between p and the center of I .
- A_r scatters lights toward the detector through the detector aperture V , which subtends a solid angle of $\Omega_s(p, V)$ with p .

- ω_s is the direction between p and the center of V .
- The power incident on A_r is $\Phi_i(A_r, I)$, and the portion of the power scattered by A_r and collected by the detector is $\Phi_s(A_r, V)$.
- We are interested in calculating the BRDF $f_r(p, \omega_s, \omega_i)$.

Recall that $f_r(p, \omega_s, \omega_i)$ is defined as the ratio of the difference in radiance leaving p toward ω_s over the difference in irradiance incident on p due to the lights coming from an infinitesimal solid angle $d\omega_i$ (omitting λ in all equations for simplicity):

$$f_r(p, \omega_s, \omega_i) = \frac{dL_s(p, \omega_s)}{dE_i(p, \omega_i)} \quad (34a)$$

$$\approx \frac{L_s(p, \omega_s)}{E_i(p, \Omega_i(p, I))}. \quad (34b)$$

There is no way we can illuminate a point p through an infinitesimal solid angle $d\omega_i$; all we could do is to illuminate a small cone of directions $\Omega_i(p, I)$. We can then calculate the average BRDF of all the directions in $\Omega_i(p, I)$ using Equation 34b (i.e., assuming the BRDF is the same for all the outgoing directions in $\Omega_i(p, I)$), whose derivation is shown in Equation 4.1.

How do we calculate $E_i(p, \Omega_i(p, I))$? There is no way we can illuminate and measure the irradiance of an infinitesimal point p ; all we can do is to illuminate a small area A_r and assume that the irradiance received is constant anywhere inside A_r , so we have:

$$E_i(p, \Omega_i(p, I)) \approx \frac{\Phi_i(A_r, I)}{A_r}. \quad (35)$$

Now how do we get $L_s(p, \omega_s)$? For this we turn to the detector side. Using basic radiometry, $\Phi_s(A_r, V)$ is expressed in Equation 36a, where p' and ω'_s are dummy variables, θ'_s is associated with ω'_s , and $\Omega_i(p', V)$ is associated with p' (c.f., p refers to a specific point on A_r , and ω_s and $\Omega_s(p, V)$ refer to physical quantities associated specifically with p):

$$\Phi_s(A_r, V) = \int_{A_r}^{A_r} \int_{\Omega_s(p', V)}^{\Omega_s(p', V)} L_s(p', \omega'_s) \cos \theta'_s d\omega'_s dp' \quad (36a)$$

$$\approx \int_{A_r}^{A_r} \int_{\Omega_s(p, V)}^{\Omega_s(p, V)} L_s(p, \omega_s) \cos \theta_s d\omega_s dp' \quad (36b)$$

$$= L_s(p, \omega_s) \cos \theta_s \int_{A_r}^{A_r} \int_{\Omega_s(p, V)}^{\Omega_s(p, V)} d\omega_s dp' \quad (36c)$$

$$= L_s(p, \omega_s) \cos \theta_s (A_r(\Omega_s(p, V) + C_1) + C_2) \quad (36d)$$

$$\approx f_r(p, \omega_s, \omega_i) E_i(p, \Omega_i(p, I)) \cos \theta_s A_r \Omega_s(p, V) \quad (36e)$$

$$\approx f_r(p, \omega_s, \omega_i) \frac{\Phi_i(A_r, I)}{A_r} \cos \theta_s A_r \Omega_s(p, V). \quad (36f)$$

The rest of the derivation proceeds as follows:

- We assume that the radiance of any ray between A_r and the detector aperture V is constant and takes the value of $L_s(p, \omega_s)$; this gets us Equation 36b.
- Since $L_s(p, \omega_s)$ and $\cos \theta_s$ are invariant to ω_s' and p' , they can be taken out of the two integrations, and this gives us Equation 36c.
- Calculating the two integrals in Equation 36c gives us Equation 36d, where C_1 and C_2 are constant. Given the boundary condition that $\Phi_s(\cdot)$ has to be 0 when $\Omega_s(\cdot)$ or A_r is 0 (if the detector aperture is closed or the illumination area vanishes, no scattered light will be detected), we know $C_1 = C_2 = 0$.
- Plugging Equation 34b we get Equation 36e.
- Plugging Equation 35 we get Equation 36f.

Equation 37a gives the final BRDF:

$$f_r(p, \omega_s, \omega_i) = \frac{\Phi_s(A_r, V)}{\Phi_i(A_r, I) \cos \theta_s \Omega_s(p, V)} \quad (37a)$$

$$= \frac{[\Phi_s(A_r, V)/(A_r \cos \theta_s)]/\Omega_s(p, V)}{\Phi_i(A_r, I)/A_r}. \quad (37b)$$

While the derivation seems excruciating, Equation 37b gives a simple interpretation. The denominator is the average irradiance incident on p through a small solid angle $\Omega_i(p, I)$ (see Equation 35), and the nominator is the average radiance leaving p . Taking the ratio of the two matches our intuition of the average BRDF: radiance over irradiance (received over a small solid angle). In particular, $\Phi_s(A_r, V)/(A_r \cos \theta_s)$ gives us the average irradiance leaving p (note that this radiance is defined at the surface perpendicular to A_r , hence the $\cos \theta_s$ term).

If we assume the surface to be Lambertian, the BRDF is then $1/\pi$ for any ω_s (under a given p and ω_i ; see Equation 28) assuming no loss of energy. This means:

$$\Phi_s(A_r, V) \propto \cos \theta_s. \quad (38)$$

That is, the flux reading weakens as the incident direction θ by a factor of $\cos \theta$. Is this surprising? It should not be if you recall our discussion of radiant intensity (Equation 10). If we assume that every point on A_r emits the same amount flux to the same solid angle (through the aperture V), the radiant intensity of p toward ω_s is to $\frac{\Phi_s(A_r, V)}{A_r \Omega_s(p, V)}$ and, thus, proportional to $\cos \theta$, which matches our earlier conclusion of how the radiant intensity of a Lambertian emitter/scatterer decays with θ .

Anytime you measure something the measurement is subject to noise and uncertainty. For instance in the case of gonioreflectometer measurement, the angular positioning of the illuminant and detector might not be accurate, the detector itself is subject to all sorts of measurement noise (which we will study in the image sensor lecture), and there might be stray lights that

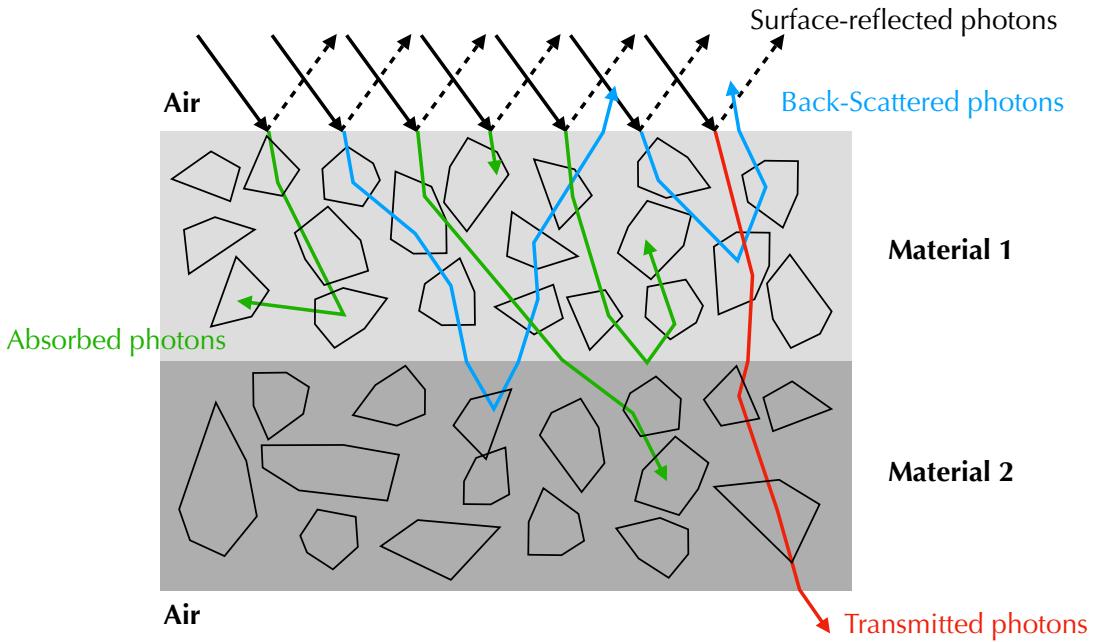


Figure 9: At the Air-Material 1 interface, photons are either reflected directly back or penetrate into the material through refraction. The refracted photons interact with the material particles through the volume scattering processes, where some photons are absorbed and others penetrate into Material 2. For someone observing from the outside, a portion of the photons would eventually leave the material composite altogether and re-enter the air. Some of these leaving photons are called the back-scattered photons that contribute the apparent surface reflectance; others transmit through the materials and contribute to the apparent transmittance of the material composite.

enter the detector. Quantifying the sources of uncertainty and, even better, correcting for them is an important part of reflectance/BRDF measurement [Lanevski et al., 2022; Rabal et al., 2012].

6 Subsurface and Volume Scattering: Informal

Once inside the material (through surface refraction), a photon roams about until it meets a particle. The interactions between photons and particles are governed by the subsurface scattering (SSS) or volume scattering processes.. As noted before, photon emission, absorption, and scattering all take place during the SSS/volume scattering processes, not just scattering, even though the names suggest otherwise. We will generally ignore emission in our discussion unless otherwise noted, but just note that emission does happen, and is correlated with absorption, since emission is the result of absorbed photons having (e.g., chemical) reactions with the particles.

Also a reminder that SSS and volume scattering are governed by exactly the same principles, because they are exactly the same thing. In computer vision and graphics literatures they might be used to refer to superficially different phenomena. Volume scattering is concerned with materials that can be modeled as a volume of particles like fog, clouds, and smoke; they are given the name **participating media** in computer graphics. SSS is, instead, more commonly used to refer to solids where subsurface-scattered photons contribute to their observed colors.

Subsurface scattering is so termed to distinguish itself from surface scattering, but what is beneath the surface is nothing more than a volume of particles. In fact, what is above the surface is also a volume of particles. Looking at Figure 9, the air, Material 1, and Material 2 can all be thought of as participating media. We usually model the air as vacuum so photons traverse in straight lines undisturbedly, but if we were to be exact we would want to model the particles in the air, which becomes a participating medium. So “above-surface scattering” is as different from as surface scattering as is subsurface scattering.

6.1 General Intuitions

We will use Figure 9 as a running example to discuss the life of photons inside the material. At the Air-Material 1 interface, photons are either reflected directly back or penetrate into the material through refraction. When a refracted photon meets a particle, the particle might absorb the photon or scatter it away. If absorbed, the photon is “dead” and can be removed from the discussion. If scattered, the photon might appear to change its direction and continue to travel on a straight line until it meets another particle, so in principle a photon can be scattered multiples.

There are three fates a photon eventually has to accept: 1) it might be absorbed along the way, 2) it might re-emerge from Material 1 back to the air, or 3) it might emerge to the air from the bottom of Material 2. Absorption is easy to understand: a photon has a certain probability of being absorbed when it meets a particle so the longer it travels and more likely it will be absorbed. Let’s examine the other two cases where a photon escapes the media.

- After multiple scattering, some of the initial photons enter Material 1 from the air will reach the Material 1-air boundary again, but this time from the material side. At that point, the photons necessarily go through another round of reflection-refraction governed by the surface scattering processes. The refracted photons will re-emerge from Material 1. This is called **back-scattering**, because these photons are scattered back to where they come. As a consequence, when we observe the material from the same side of the illumination, the lights that enter our eye come from two sources: the initial surface scattering and the back-scattering.
- Some photons might leave Material 1 from the other side and enter Material 2, in which photons go through the same volume scattering processes, where some are absorbed, some can be turned back to Material 1, and some, critically, can hit the Air-Material 2 interface. Just like what happens at the Air-Material 1 interface, some of the photons will eventually emerge from Material 2. These photons essentially survive the absorption of all the parti-

cles in the media. When you observe the material from opposite side of the illumination, it is these transmitted photons that dictates the color of the material.

Sometimes people will also say “sub-surface scattering is caused by photons exiting at a point different from the incident point.” It points to the fact that a photon can re-emerge anywhere from the material after SSS, whereas surface scattering is *modeled* to be taking place only at the incident point (although we will see later that this is just a useful macroscopic abstraction or, rather, modeling strategy).

6.2 Transparent vs. Opaque vs. Translucent Materials

We often hear materials being described as opaque, translucent, and transparent. We can now more scientifically approach these terms given the intuitions we have built so far.

Transparent Materials

Transparent materials either scatter lights predominately in forward directions or they scatter very little lights (other than surface scattering). Either way, most photons traveling through the material are either absorbed or go through without changing much of their directions. So if you hold a transparent material against a light source, you can clearly see through the material and see the light on the other side. This does not mean transparent materials always have the same color as the light source — absorption could be wavelength-selective. An example is aqueous/dye solutions where dye molecules are very small ($\sim nm$ range) and, thus, scatter little lights so they look transparent, but depending on the absorption spectrum (which depend on how the dye molecules interact with molecules in the solvent), most dye solutions are not colorless.

Opaque Materials

In many materials, photons arriving at the material surface are either reflected back right away at the surface or, for those that do penetrate into the materials, are all absorbed by the subsurface particles. Examples include conductors like metals, whose subsurface absorption is very strong, or sufficiently thick dielectrics. These materials are **opaque** in two senses. First, their transmittance is practically 0. Because of strong absorption, no photon re-emerges at the other side of the material. If you hold a, say, brick (dielectric) against a light bulb, the brick would completely block the light. Second, their reflectance is independent of the substrate or the material beneath them, so they completely hide the color of the substrate¹¹. Painters know that if they want to cover a layer in their painting, they will need to apply a very thick layer of paint on top.

¹¹Technically speaking having a zero transmittance requires the material to have a stronger absorption than hiding the substrate, because in the latter case photons have to make a round trip so have more opportunities to be absorbed.

Translucent Materials

Translucent materials such as jade, wax, and human skin are neither opaque nor transparent. If you hold a wax against a light bulb, the wax will not completely block the light so you will see some light, but you will not be able to see clearly the other side through the wax, since photons from the light bulb are very much volume-scattered after passing through the wax. Clearly modeling SSS is critical for accurately estimating the color of translucent materials. In fact, in graphics literature we sometimes see things like “modeling translucent material must consider sub-surface scattering.” In this sense, we might be tempted to classify participating media as translucent materials, because their colors certainly very much depend on volume scattering. While it is technically correct, people rarely do that, perhaps just because of the weirdness of calling say, smokes, a material rather than a medium?

It is *not* true that SSS is important only in modeling translucency. Modeling SSS can be important for opaque materials. Consider the wax case: what if we make the wax very thick? The thick wax will eventually become opaque in that it will completely hide the material behind it. But that does not mean volume scattering does not matter here; the back-scattered photons do contribute to the apparent color of a thick wax.

Oil Painting Example

To put things together, consider a painting. One way paintings are characterized is by how they were painted, and we might see things like “oil on canvas”. Oil means the paint is oil paint, where paint pigments are dispersed into (usually linseed) oil, which is usually called the binder or the vehicle. Canvas is the substrate, which is nothing more than another material, that is right beneath the painting.

The oil itself is somewhat transparent especially when you just apply a thin layer on the canvas. But with the paint pigments, the entire oil paint becomes a translucent material. When photons leave the oil paints, they immediately interact with the canvas. If the paint layer is thick enough, virtually no photon can ever reach the canvas. But if the paint is relatively thin, the property of the substrate will contribute to the overall color of the paint. For instance, if the canvas is white-ish, a good percentage of the photons will be reflected back. The same paint would look much darker if the canvas is black, which absorbs a lot of photons.

6.3 Equilibrium

We can view the light-material interaction as a dynamical system under an equilibrium. To appreciate this, consider again Figure 9. Some photons entering Material 1 are back-scattered and hit the Air-Material 1 interface and some of those photons will re-enter Material 1 through internal reflection. Those photons will then go through multiple scattering, and as a result some will be back-scattered again and hit the Air-Material 1 interface. The cycle goes on. The secondary back-scattering is weaker in power than the first back-scattering, and the third-order back-scattering is even weaker, and so on. So eventually you can imagine that the total number of photons back-scattered at the surface will reach a constant.

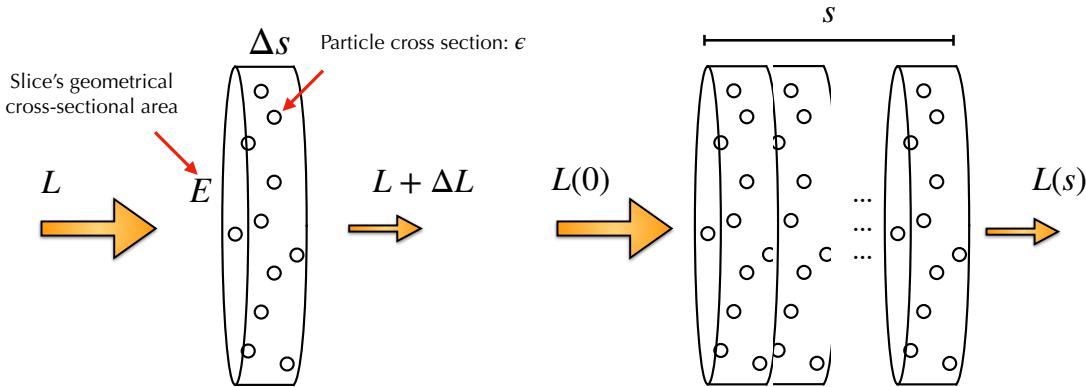


Figure 10: Conceptual model to help reason about photon absorption; see text. Adapted from Max [1995, Fig. 1].

In fact, this sort of dynamics takes place everywhere inside the material along every direction. If you pick a point \mathbf{p} in the material (or at the surface) and a direction ω starting at the point, the radiance at (\mathbf{p}, ω) is a constant under equilibrium. In other words, the spatial radiance distribution (a.k.a., the light field) is not changing over time.

The equilibrium is reached almost instantaneously, since lights propagate incredibly fast. So the equilibrium discussion is probably of no practical impact in modeling or actual measurement, but it is still important to keep this in mind. The (spectral) reflectance/BRDF modeling/measurement is done assuming equilibrium, and later when we model volume scattering we will set up the differential equations under the equilibrium assumption, too.

7 Absorption

We will focus on modeling absorption in this chapter, and the way we build the models is fundamental to how scattering will be dealt with later.

7.1 A Simple Case: Collimated Illumination on Uniform Medium

Imagine that a beam of light hits a volume of particles. The light is **collimated** in that all photons travel along the same direction. We take a slice of the material perpendicular to the incident direction. The slice is so thin that no particles in that material cover each other from the direction of the incident light. This is shown in Figure 10 (left). We also for now assume that the medium is *uniform* in that the **number concentration** c (i.e., the number of particles per unit volume) of each slice is exactly the same.

Say the slice has a depth of Δs and a geometrical cross-sectional area of E . All the particles have the same geometrical cross-sectional area of ϵ_g . In the simplest model a photon is absorbed whenever it hits a particle. In reality, the chance of absorption can be higher or lower. The *effective* area available for absorption is

$$\epsilon = \epsilon_g Q_a, \quad (39)$$

where Q_a is called the **absorption efficiency** and is usually smaller than 1 for molecules (which have small ϵ_g) and greater than 1 for large particles (whose ϵ_g can be large). Q_a is wavelength dependent so we should have written it as $Q_a(\lambda)$, but we will omit the wavelength in our notations for simplicity's sake. In physics, ϵ is called the **absorption cross section** of the particle; it characterizes the intrinsic capability of a particle to absorb photons. Mind the subtle but important difference between the geometrical cross-sectional area and the cross section of a particle.

The question we are interested in is, if the incident radiance is L , what is the radiance leaving the slice $L + \Delta L$? By convention, the difference between the exitant and incident radiance is always ΔL , which in this case has to be negative. The percentage of photons that are absorbed by this slice of particles ($-\frac{\Delta L}{L}$) is equivalent to the cross-sectional area of the slice that is covered by the total cross sections of the particles:

$$-\frac{\Delta L}{L} = \frac{cE\Delta s\epsilon}{E}, \quad (40a)$$

$$\frac{\Delta L}{\Delta s} = -c\epsilon L = -\sigma_a L, \quad (40b)$$

where c is the particle concentration of the slice, and $E\Delta s$ is total volume of the slice. So $cE\Delta s$ is the number of particles in this thin slice, and $cE\Delta s\epsilon$ is the total cross section of all the particles. Given the assumption that no particle is covering each other, $\frac{cE\Delta s\epsilon}{E}$ is then the percentage of the thin slice's cross-sectional area that is available for photon absorption and, thus, the percentage of the incident photons that are absorbed. The negative sign on the left-hand side of Equation 40a signals the fact that ΔL is negative.

We re-write Equation 40a as Equation 40b, which shows that the amount of photon absorption per unit length ($\frac{\Delta L}{\Delta s}$) is proportional to the current amount of photons up to a scaling factor $c\epsilon$. In the computer graphics literature, $c\epsilon$ is called the **absorption coefficient**, denoted σ_a .

When Δs approaches infinity, we can re-write Equation 40b as a differential equation in Equation 41a. This equation is a classic case of exponential decay, and its solution is given by Equation 41b, which allows us to calculate the remaining radiance after the light travels a length s :

$$\frac{dL}{ds} = \lim_{\Delta s \rightarrow 0} \frac{\Delta L}{\Delta s} = -\sigma_a L, \quad (41a)$$

$$L(s) = L_0 e^{-\sigma_a s}, \quad (41b)$$

where $L_0 = L(0)$ is the initial radiance of the light before interacting with the particles, as visualized in Figure 10 (right), $L(s)$ denotes the radiance at a particular length s .

Equation 41b is called the **Bouguer-Beer-Lambert's law** (BBL), which is a geometrical optics' simplification of the electromagnetic theory of light-matter interaction where the matter is purely absorptive [Mayerhöfer et al., 2020].

Absorption Coefficient

The absorption coefficient is an important measure of the medium's ability to absorb photons. It has a unit m^{-1} , which means it is not bound by 0 and 1. One way to interpret the absorption coefficient is to observe that $\sigma_a ds = dL/L$, which is the fraction of the radiance absorbed or the probability of light absorption by an infinitesimal slice. So $\sigma_a = (dL/L)/ds$ can be interpreted as the probability *density* of photon absorption, i.e., the probability of absorption per unit length traveled:

$$\sigma_a = \lim_{\Delta s \rightarrow 0} \frac{\Delta L}{L}/\Delta s = \frac{dL}{L ds}. \quad (42)$$

Like any density measure, absorption coefficient is most useful when it is integrated: when we integrate σ_a over the length that light travels, we get the fraction/percentage of the light absorbed. One can also show that $1/\sigma_a$ is the mean length a photon can travel before being absorbed [Bohren and Clothiaux, 2006, Chpt. 5.1.3], so this quantity is given the name **mean free path**.

An Alternative Derivation

An equivalent way of deriving the BBL law is the following. We divide the entire volume (with a total length of s) into N thin slices, each with a length of Δs . After the first slice, the surviving portion of the initial radiance is $L = L_0(1 - \sigma_a \Delta s)$, so after going through all the N slices, the remaining radiance is given by Equation 43a:

$$L_N = L_0(1 - \sigma_a \Delta s)^N = L_0(1 - \sigma_a \frac{s}{N})^N, \quad (43a)$$

$$L(s) = \lim_{N \rightarrow \infty} L_0(1 - \sigma_a \frac{s}{N})^N = L_0 e^{-\sigma_a s}. \quad (43b)$$

Now when ds becomes infinitesimal small, N approaches infinity, so the limit of the remaining radiance as a function of the total length s is given in Equation 43b, which is the same as Equation 41b.

7.2 A Few Important Quantities

We can now define a few other commonly used quantities (omitting the wavelength dependence for simplicity). The **transmittance** T of a volume with a total thickness of s is defined as the percentage of the transmitted/unabsorbed photons after traveling the length of s (Equation 44a). The **absorbance** A is the product of $\sigma_a s$. The **absorptance** a of a volume is defined as the percentage of the absorbed photons by the volume, which relates to T and A by Equation 44b.

$$T = \frac{L(s)}{L_0} = e^{-\sigma_a s}, \quad (44a)$$

$$A = -\ln T = \ln \frac{L(s)}{L_0} = \sigma_a s, \quad (44b)$$

$$a = 1 - T = 1 - e^{-A}. \quad (44c)$$

We have seen these definitions in the Photoreceptor Chapter, where we said that one very nice thing about the absorbance A is that it is approximately equivalent to absorptance a when A is small (which would be true when, e.g., the length s is very small, as is the case when discussing how a photoreceptor absorbs photons when illuminated transversely).

Another nice thing about absorbance is that absorbances add — because *absorption coefficients add*. Imagine you have n kinds of particles mixed up in a medium, each with a different absorption coefficient σ_a^i , the overall absorbance of the medium is the sum of the individual absorbance A^i derived as if the medium is made up of only one kind of particles. That is:

$$A = \sum_i^n A^i = s \sum_i^n \sigma_a^i = s \sum_i^n c^i \epsilon^i, \quad (45)$$

where c^i and ϵ^i are the concentration and absorption cross section of the i^{th} particles. This is not a surprising result. As long as particles in a thin slice of this new heterogeneous medium do not cover each other, we can easily extend Equation 40a and the rest of the derivation to consider multiple kinds of particles; eventually Equation 45 would be a natural conclusion.

Equation 45 is a nice conclusion to have, because usually we *are* dealing with hybrid media. For instance, a paint is a mixture of binder particles and pigment particles, and a mist is a mixture of water droplets and air particles. If we do not want to model individual matters, we can use a single absorption coefficient to describe the aggregate behavior of the mixture. That absorption coefficient does have a physical meaning: it is the concentration-weighted sum of the individual absorption coefficients.

There are a bunch of other quantities defined in the literature. The state of the definitions is a bit of a mess, largely because different communities use different definitions.

- In visual neuroscience people sometimes use a quantity called **specific absorbance** (see, e.g., [Bowmaker and Dartnall \[1980\]](#)), which is the absorbance per unit length $\frac{A}{s}$. Whenever you see a quantity that starts with the word “specific”, chances are that the quantity is defined per unit length. You can see that specific absorbance is actually just our absorption coefficient.
- In scientific communities especially chemistry and spectroscopy, people define ϵ , rather than $c\epsilon$, to be the absorption coefficient. You can see the appeal of doing that — ϵ is a more fundamental measure of a medium’s ability to absorb photons, independent of the particle concentration c (and certainly independent of the traversal length s).

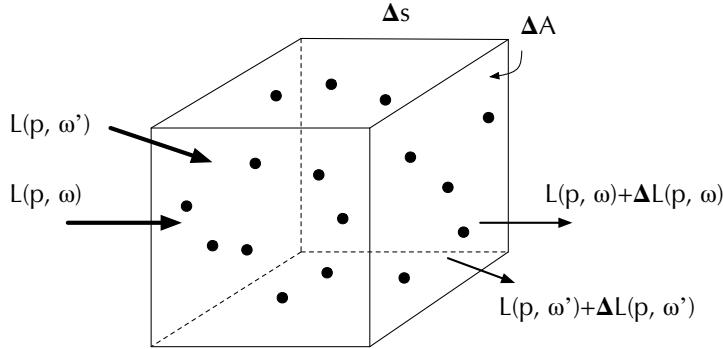


Figure 11: A conceptual model to help reason about photon absorption in the general case, where the absorption coefficient can vary spatially and directionally. The medium is divided into many tiny elemental volumes, each of which is so small that particles do not cover each from any direction.

- The absorbance defined in Equation 44b is technically called the **Naperian absorbance**, because we take the natural logarithm of T . Sometimes people also use the **decadic absorbance**, which is defined as $-\log T$. This quantity is also called the **optical density**.
- Finally, the number concentration c here is defined in terms of the absolute quantity per unit volume, but sometimes people want to define c as the **molar concentration**, which is the number of moles per unit volume. If so, all other derived quantities are then prefixed with “molar”. Next time when you see something like the **molar decadic absorption coefficient**, you know what it is!

The annoying thing is that people do not always tell you which definition they use. The plea I have to you is to be specific about which definition *you* use in your writing and tell me when I am being vague!

7.3 General Case

So far we have assumed that the absorption coefficient $\sigma_a = c\epsilon$ is a constant regardless of the position p in the medium and along any direction ω . The former property assumes that the medium is uniform, and the latter property is called **isotropic**¹² in that the medium’s ability to absorb photons is independent of the light direction.

Both assumptions are problematic in practice. The concentration can change spatially and should be denoted $c(p)$, where p is an arbitrary position in space. ϵ can also change with p and, more importantly, change with the direction of light incidence ω . For instance, the particles

¹²“Isotropic” is a very overloaded term; it just means some physical property is invariant when measured from different directions. So depending on what physical property you care about, “isotropic” can mean different things. The property we care about here is a volume’s ability to absorb photons, which is different from our earlier use of isotropy, which is concerned with the ability of a surface to scatter photons.

might not be spherical so its geometrical cross-sectional area and, thus, the cross section ϵ available for absorbing photons can depend on ω . As a result, the absorption coefficient should generally be denoted $\sigma_a(p, \omega)$.

Effectively, our conceptual model, shown in Figure 11, has to be changed to one where the entire body of particles is divided into many equally-sized volumes (with a length Δs and an area ΔA), each of which is so small that particles do not cover each other from any direction. The radiance reduction per unit length in a small volume is then expressed as:

$$\frac{\Delta L(p, \omega)}{\Delta s} = -\sigma_a(p, \omega)L. \quad (46)$$

Given this model, we can calculate the exitant radiance after light travels a length s through the medium:

$$L(p + s\omega, \omega) = L(p, \omega)e^{-\int_0^s \sigma_a(p + t\omega, \omega)dt}, \quad (47)$$

where ω is the (unit) direction of the incident radiance, $L(p, \omega)$ is the incident radiance, $L(p + s\omega, \omega)$ is the exitance radiance (radiance toward ω leaving the entire medium after traveling s).

You would notice that for a beam with an oblique incident direction, the distance traveled, say $\Delta s'$, can be different (longer or shorter than) from Δs . Our model can account for this by folding the factor $\Delta s'/\Delta s$ specific to a particular direction ω' into the absorption coefficient $\sigma_a(p, \omega')$. Note that the $\Delta s'/\Delta s$ factor should be the average for all the incident photons with the same direction ω' across the entire ΔA .

7.4 Nature and Applicability of the Model

The absorption model (the BBL law) derived before (Equation 41b and Equation 47) is a continuous one, but it is derived based on modeling discrete particles and events. It is another example of the modeling methodology discussed on p. 25.

Equation 47 seems to suggest that absorption coefficient $\sigma_a(p, \omega)$ is continuously defined at any position p in the medium along any direction ω . It is not true. For starters, concentration c is not continuous. Rather, it exhibits the triphasic profile shown in Figure 6. As we keep shrinking the size of the volume to the molecular scale, eventually the concentration depends on whether the tiny volume contains any molecules or not, so it becomes wildly discontinuous, not to mention the headache of dealing with a partial molecule in a volume — should it be counted or not? In general the absorption coefficient can be an arbitrary discontinuous function that is not integrable.

What about Equation 41b where the absorption coefficient is uniform so we do not have to take the integral? Well, that is a lie too: concentration is not continuous, so it cannot be uniform everywhere and, by extension, the absorption coefficient cannot be a constant everywhere either. So Equation 41a is technically wrong when we let $\Delta s \rightarrow 0$ (i.e., $N \rightarrow \infty$), which is necessary for us to construct the differential equation (or take the limit in Equation 43b). For Equation 41a

to be true, the concentration/absorption coefficient must be a constant everywhere, which can be true only if the volume is continuous.

What has to happen is that the limit of Δs cannot be literally 0 and the limit of N cannot be infinity. What we do is to keep reducing Δs to the point where the concentration (and thus absorption coefficient) is insensitive to slight perturbation of Δs (i.e., operating in the stable range in Figure 6), and call it the concentration/absorption coefficient of that specific Δs . And we repeat this for all the Δs . This certainly applies to the general-case models in Equation 46 and Equation 47, where we iterate over not the thin slices Δs but all the tiny volumes ($\Delta A \times \Delta s$). So all the integral symbols are secretly summing over an extremely fine-grained grid.

How big of an error are we introducing here? Technically, we should sum all N slices across the total traversal length s in Equation 43a. If we assume Δs to be very small (even thought not infinitesimal) compared to s , N would be large, so taking the integration (equivalent to letting $N \rightarrow \infty$) would be very close to summing over N . Similarly, the integral in Equation 47 should have been a summation of the concentration in each of the N slices. If you want to be pedantic, however, the integration there is exact: we can model c as a piece-wise function, where the value at each piece is the concentration of the corresponding volume. Integrating over a piece-wise function is the same as summing all the pieces. Only the exponential expression in Equation 47 is inexact.

The discontinuity of the medium is of course orthogonal to the discontinuity and non-uniformity in the light field itself. For instance, the fact that we use $\frac{cE\Delta l\epsilon}{E}$ as the percentage of photon absorption in Equation 40a (and implicitly in Equation 41b) assumes that the irradiance of the incident illumination is continuous and uniform in the small volume. This is technically not true because photons are discrete packets of energy. But in practice this is not a concern because we can assume that there is an enormous amount of photons incident on the small volume and these photons are randomly distributed.

In essence, we are using the aggregated behavior of a large number of photons to model the behavior of a small volume. This is similar to the microfacet models, where we use the aggregated behavior of a large number of microfacets to statistically model the behavior of a small macro-surface.

This sort of modeling strategy is a weird case where the discrete model provides the “ground truth”, which is approximated by a continuous model. I say ground truth — that is to the extent that the geometrical optics can approximate the electromagnetic theory of light-matter interaction. The BBL law fails when the wave nature of photons have to be considered [Mayerhöfer et al., 2020].

8 Scattering

Scattering is much more difficult to reason about than absorption, primarily because a scattered photon is not “dead” and continues to participate in light-matter interaction. The way to study scattering is to first understand the behavior of a single scattering event and then consider the overall behavior of a large collection of particles.

This chapter focuses on discussing a single scattering event (Chapter 8.2), and the next

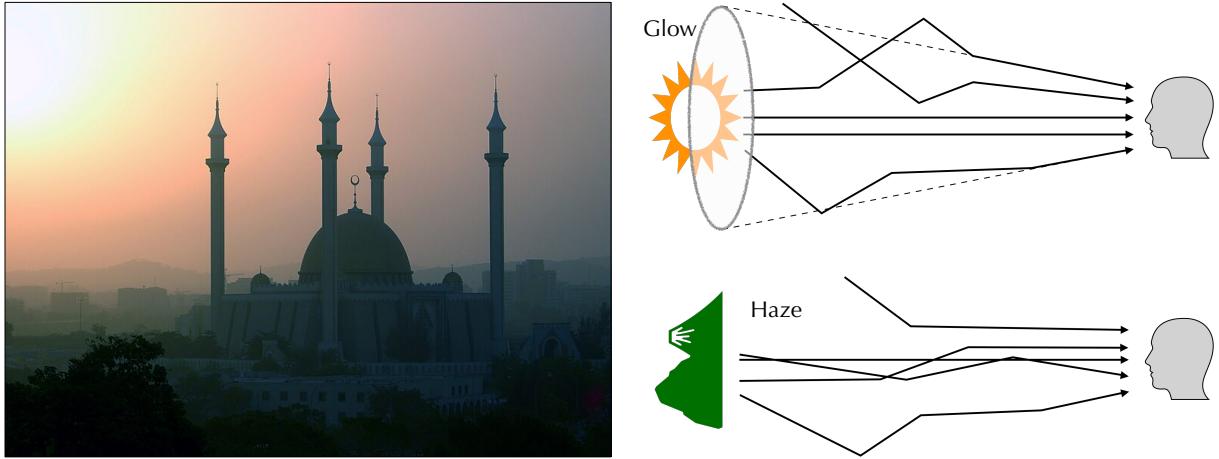


Figure 12: Left: The atmospheric scattering of the sand coming from the Sahara during Harmattan glows the sun and gives a hazy view of the remote mountains. Nigeria's National Mosque is in the foreground; from [Kipp Jones \[2005\]](#). Right: illustrations of the glow of the sun and the haze. Both are due to scattering and the difference is purely visual but not fundamental.

chapter discusses the general case where a large collection of particles interact with photons. Before all these though, it is useful to first build some intuitions as to why there is a distinction between a single scattering event and scattering by a particle collection and explicitly lay out the assumptions made for the rest of our discussions (Chapter 8.1).

8.1 Scattering by a Particle vs. a Collection of Particles

In geometric optics terms, scattering can be thought of as an event that takes place between a photon and a particle. In the real world, however, objects and media are usually made of a large collection of particles, which introduces two complications: multiple scattering and interference.

Multiple Scattering

First, it is possible that a scattered photon, after traveling a certain distance, meets another particle and gets scattered again. This makes it considerably more difficult to analyze the effect of scattering by a medium than does the scattering of a single particle.

Look at Figure 12 (left) taken during Harmattan, where the atmosphere is full of sands and dusts blown from the Sahara. The large collection of particles in the atmosphere scatter lights, glowing the sun and giving the remote mountains a hazy view. The right panel illustrates the scattering events that give rise to the glow and the haze.

Without scattering, sun lights enter the eye directly. With scattering, some photons from the sun first knocked out of the view and could potentially be then scattered again back to the eye. Some photons that enter the eye might even come from nearby objects other than the sun. The scattering creates a glow around the sun and, for the observer, the sun appears larger

than it actually is. The hazy view of the mountains is created by exactly the same scattering processes. The photons enter the eyes are mixed up from different parts of the mountains and from other objects. The mountains appear hazy rather than glowing as the sun does simply because the sun has a higher brightness contrast against the background than does a region on the mountain. So the distinction between “glow” and “haze” is nothing more than a visual difference at a superficial level rather than anything deeper in physics.

You can see why multiple scattering by large collections of particles poses challenges to our analysis. If a photon is scattered once in the medium, the only effect of scattering would be to knock photons out of our line of sight and, thus, remote objects would only look dimmer rather than hazy. In this case, scattering would function exactly like absorption, for modeling purpose at least. With multiple scattering, we have to track not only photons that are scattered out but also photons that are scattered into the rays that enter our eyes¹³. This is a daunting task considering that we are usually dealing with millions of particles and billions of photons, if not more.

If you want to be absolute pedantic, we can distinguish the following cases:

1. a single scattering event, where a photon meets a particle and is scattered away;
2. single scattering, where a photon is scattered *once* by a medium (a large collection of photons), which is under
 - (a) a collimated illumination, so the radiance of a ray can only be weakened because photons are scattered to other directions,
 - (b) an arbitrary illumination, so the radiance of a ray can be both weakened and augmented (by photons scattered from other directions);
3. multiple scattering, where a photon is scattered *multiple times* by a medium, so the radiance of a ray can both be weakened and augmented.

Chapter 8.2 studies Case 1 and Case 2(a) together, because the latter is the statistical consequence of the former. Chapter 9 studies Case 2(b) and Case 3 together, because they have the same observable effects and, thus, are modeled in the same way.

Interference and Coherence

Second, when there is a large collection of particles, the scattered radiation fields of individual particles can interfere with each other. The exact impact of interference can only be calculated by considering the wave nature of the light. But to the first order, the inference depends on how densely packed the particles are.

In fact, the specular surface scattering we discussed in Chapter 4.4 is just a macroscopic approximation of the microscopic volume scattering where particles interfere non-randomly. In a mirror or a glass of water, the particles/molecules are very densely packed to the point that

¹³Technically photons from other objects can enter our eye through a single-scattering event; see the discussion at the end.

the distance between two particles is smaller than the wavelength of the light. As a result, the scattering is *coherent*, which gives rises to the *illusion* of a specular surface. One can show that the Fresnel equations are the solution to the Maxwell's equations when surface particles are densely packed.

Why would the particle density matter? If particles are very close to each other, their radiation fields are close too, so the interference is stronger and cannot be ignored. More importantly, when particles are close to each other, their spatial positions can no longer be treated as random, so the interference can become coherent. Imagine you drop particles into a vast empty space; the particle sizes are much smaller relative to the space so their spatial distribution can be roughly described as random. But if the particles are very densely packed, where the next particle can be is very much restrained, so their positions are highly correlated, leading to coherent scattering.

We will generally assume **incoherent scattering** unless otherwise noted, where individual scattering events interfere each other in random ways, so we are spared of the complication of thinking of the wave nature of the photons. Under this assumption, the total power scattered by a collection of particles is the same as the sum of the power scattered by the individual particles. This happens when the particles are sufficient sufficiently distant (separated by more than multiple wavelengths) and their spatial arrangements are uncorrelated.

8.2 A Single Scattering Event

Scattering Efficiency and Coefficient

Intuitively, scattering has a similar effect as absorption: it weakens the radiance by taking photons away from a beam of light. The difference is that scattered photons are not dead; they are re-directed to other directions. We can define two important quantities, one to characterize a *particle*'s ability to scatter photons and the other to characterize a *medium*'s ability to scatter photons.

Similar to the situation in absorption, the intrinsic capability of a particle to scatter photons is defined by the particle's **scattering cross section** ϵ_s , which itself is the product of the geometrical cross-sectional area of the particle ϵ_g and the **scattering efficiency** Q_s . We can then define the **scattering coefficient** σ_s of a medium (a large collection of particles), which characterizes the ability of the medium to scatter photons away from its incident radiance. σ_s is the product of the particle concentration of the medium c and the particle's scattering cross section ϵ_s . Again, σ_s has a unit m^{-1} and is not bound by 0 and 1; it is best interpreted as the probability density (i.e., probability per unit length) of light being scattered away. Of course both the scattering efficiency and scattering coefficient can vary spatially, angularly, and spectrally.

The effects of scattering and absorption add up, because they both weaken a radiance. We can extend Equation 47 to consider scattering (again omitting the wavelength from the equations):

$$L(p + s\omega, \omega) = L(p, \omega) e^{-\int_0^s (\sigma_a(p + t\omega, \omega) + \sigma_s(p + t\omega, \omega)) dt}, \quad (48a)$$

$$T(p \rightarrow p + s\omega) = \frac{L(p + s\omega, \omega)}{L(p, \omega)} = e^{-\int_0^s (\sigma_a(p + t\omega, \omega) + \sigma_s(p + t\omega, \omega)) dt}, \quad (48b)$$

where $\sigma_s(p, \omega)$ is the scattering coefficient at p toward the direction ω , and $T(p \rightarrow p + s\omega)$ is defined as the **transmittance** between p and $p + s\omega$ along the direction ω .

Equation 48a can be derived using the same idea as that used for deriving the absorption equation in Chapter 7.1 by modeling a thin layer Δx — with an additional assumption that Δx is so thin that a photon is scattered at most once before leaving Δx . Therefore, scattering by a single particle has the same effect of absorption: they both take the photon out of the radiance and that is why the absorption equation (Equation 47) can be directly extended here.

Think of the applicability of Equation 48a: it says that the radiance of a ray can only be weakened. If the incident light has only one direction (e.g., a collimated beam), Equation 48a is true when a photon is scattered at most once in the medium. This is because a scattered photon will not have a chance to get back to the ray. If the incident light is not monodirectional, e.g., diffuse illumination, Equation 48a in general does not apply — even if we consider only single scattering. This is because photons originally not along the direction ω can be scattered toward it through just one single scattering event. We can see how limited Equation 48a is: it applies only when the illumination is collimated and we assume only single scattering. We will relax this constraint later.

Just like the absorption case (Equation 45), if a medium is mixed with different particles, each with a different scattering coefficients, the overall scattering coefficient is the sum of the individual scattering coefficients as if the medium is made up of a particular kind of particles.

The sum of the scattering coefficient and absorption coefficient is called the **extinction coefficient** or **attenuation coefficient**, denoted $\sigma_t(p, \omega)$:

$$\sigma_t(p, \omega) = \sigma_a(p, \omega) + \sigma_s(p, \omega). \quad (49)$$

The ratio between the scattering coefficient and the attenuation coefficient is called the **single-scattering albedo** of the medium:

$$\rho = \frac{\sigma_s(p, \omega)}{\sigma_t(p, \omega)}. \quad (50)$$

This albedo can be seen as the volumetric counterpart of the surface albedo discussed in Equation 19. The two forms of albedo have the same physical meaning: the fraction of the incident energy that is scattered away (i.e., not absorbed). A dark medium (e.g., smoke) has a lower albedo and a bright medium (e.g., mist) has a higher albedo.

The sum of the scattering and absorption cross sections is called the **extinction cross section** or **attenuation cross section**, denoted $\epsilon_t = \epsilon_a + \epsilon_s$. And of course $1/\sigma_t$ is the mean free path in a medium where both absorption and scattering take place, i.e., the mean distance a photon can travel without being absorbed or scattered away.

Scattering Direction Distribution: Phase Function

While the scattering efficiency (coefficient) characterizes how well a particle (medium) is able to scatter photons, it tells us nothing about the *direction* of scattering. The direction of a single scattering event is characterized by the **phase function** $f_p(p, \omega_s, \omega_i)$, which can be interpreted as the probability *density* function that a photon incident from a direction ω_i is scattered toward a direction ω_s . We will write the phase function as $f_p(\omega_s, \omega_i)$, omitting p when the discussion is unconcerned of p .

Like all density functions, the meaning of a phase function is most clear when it is integrated to compute some other quantity. For instance, consider that a point p receives an infinitesimal irradiance dE_i from a direction ω_i (over an infinitesimal solid angle $d\omega_i$); these photons are scattered across the entire sphere (a solid angle of 4π). The irradiance (fraction of photons) scattered toward ω_s is $dE_i f_p(p, \omega_s, \omega_i)$; integrating all the scattered radiance gives us the total scattered irradiance dE_o :

$$dE_o = \int^{\Omega=4\pi} dE_i f_p(\omega_s, \omega_i) d\omega_s. \quad (51)$$

Now, of course some of the photons in dE_i might not be scattered; they could be absorbed or they could simply not hit the cross section of any particle. So technically $dE_o \leq dE_i$, just like how energy conservation is expressed in surface scattering in Equation 16b. By the convention in the volume scattering literature, however, the phase function is defined such that dE_i refers to only the portion of the incident irradiance that does get scattered. Therefore, $dE_o = dE_i$ and, thus, we have:

$$\int^{\Omega=4\pi} f_p(\omega_i, \omega_s) d\omega_s = \int^{\Omega=4\pi} f_p(\omega_i, \omega_s) d\omega_i = 1. \quad (52)$$

That is, the phase function integrates to 1; the second integral can be derived using the reciprocity (since we are still dealing with geometrical optics):

$$f_p(\omega_i, \omega_s) = f_p(\omega_s, \omega_i). \quad (53)$$

One way to interpret the fact that the phase function integrates to 1 is that the phase function is the *conditional* probability density function of scattering: given that a photon is scattered what is the probability (density) of scattering to a particular direction? Another way to interpret the phase function is that it describes the *percentage* or fraction of the incident flux/irradiance/radiance from the direction ω_i that is scattered toward the direction ω_s , so it is akin to the (directional-directional) reflectance quantity discussed in Chapter 4.2.

The phase function can be seen as the volumetric counterpart (in the sense that we are talking about volume scattering) of BRDF (Equation 4.1) — with two differences. First, the definition of the BRDF accounts for absorption, so the BRDF integrates to *at most* 1. Second, there is no $\cos \theta$ term when using the phase function; see, e.g., Equation 51, unlike how

BRDF is used to turn incident irradiance into outgoing radiance (e.g., Equation 16b). This is because the phase function is a fraction quantity, like the (directional-directional) reflectance discussed in Chapter 4.2 rather than a distribution quantity like the BRDF. Note that these differences between the BRDF and the phase function are purely out of convention; there is nothing fundamentally different going on here; it is all about how we choose to define things.

Given the normalization in the phase function, the scattering efficiency should actually be parameterized as $\bar{Q}_s(p, \omega_s, \omega_i)$:

$$\bar{Q}_s(p, \omega_s, \omega_i) = Q_s(p, \omega_i) f_p(p, \omega_s, \omega_i), \quad (54)$$

where $Q_s(p, \omega_i)$ should be interpreted as the *total* scattering efficiency at p over all outgoing directions for a given incident direction ω_i . Similarly, the scattering coefficient would be expressed as:

$$\bar{\sigma}_s(p, \omega_s, \omega_i) = \sigma_s(p, \omega_i) f_p(p, \omega_s, \omega_i), \quad (55)$$

where $\sigma_s(p, \omega_i)$ is interpreted as the *total* scattering coefficient at p over all outgoing directions for a given incident direction ω_i .

Figure 14 visualizes a few common phase functions. An implicit assumption made in these visualizations is that while $f_p(\cdot)$ is technically a 4D function parameterized by ω_s and ω_i , the phase function of many natural media is 1D and depends only on the angle θ subtended by ω_s and ω_i . Consider under what conditions can this simplification be true.

- First, it says that the phase function does not depend on the absolute incident direction ω_i but the relative angle between ω_i and ω_s . To get a visual intuition, see Figure 13; if the phase function is invariant to the photon incident direction ω_i , we can, without losing any generality, assign ω_i to the z -axis; the scattered direction ω_s is parameterized by θ and ϕ .
- Second, it also says the phase function depends on only θ but not ϕ . That is, the phase function is rotational symmetric about the incident direction ω_i . So $f_p(\omega_s, \omega_i) = f_p(\omega_s', \omega_i) \neq f_p(\omega_s'', \omega_i)$.

Intuitively, the phase function has the following two properties:

- If you fix the incident direction, no matter how you rotate the particle, the phase function distribution is the same. Alternatively, if you change the incident direction, the phase function distribution moves along with the incident direction.
- Given an incident direction, the phase function distribution is axially symmetric about the incident direction.

The two conditions above are met only when the medium consists of randomly distributed spherically symmetric particles¹⁴, in which case 1) there is no reason to think that any incident

¹⁴or when the medium consists of randomly distributed and oriented spherically asymmetric particles, in which case the medium is *statistically* spherically symmetric.

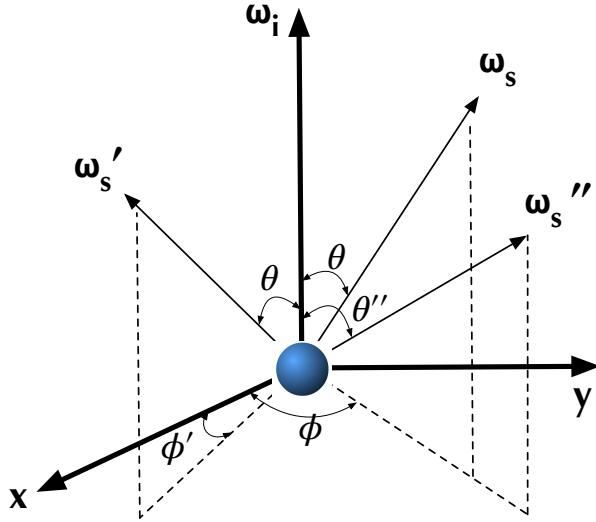


Figure 13: The phase function of a spherical particle is 1) invariant to the incident direction ω_i , which, without losing generality, is taken to be the z -axis here, and 2) also invariant to the azimuthal angle ϕ of the outgoing direction ω_s but depends on the polar angle θ . So $f_p(\omega_s, \omega_i) = f_p(\omega_s', \omega_i) \neq f_p(\omega_s'', \omega_i)$. Media consisting of such particles are called isotropic media, but it does not mean the particle itself is an isotropic scatterer, which does not exist but if it did its phase function would be a constant (invariant to both θ and ϕ).

direction is special, so the phase function certainly is invariant to ω_i , and 2) there is no reason to think ω_s and ω_s' are any different since one should not expect the scattering behavior to change if we rotate the sphere about the incident direction (z -axis).

A medium consisting of spherically symmetric particles is called a *symmetric* or an **isotropic medium**, which is what Figure 14 assumes, where the distance of a point on the contour to the center represents the magnitude of the phase function at that particular θ . Usually when we refer to an isotropic medium, not only is the phase function rotationally invariant to the incident direction, but also total scattering coefficient $\sigma_s(p, \omega_i)$ (Equation 55), too¹⁵.

As we said earlier, “isotropic” is an unbelievably overloaded term. People also call a particle an **isotropic scatterer** if its phase function is a constant, i.e., invariant to ω_s ; such a phase function is sometimes called an *isotropic phase function*. An isotropic scatterer does not exist; it is a purely theoretical construction and, if it existed, its phase function would take the value of $\frac{1}{4\pi}$ given Equation 52, as shown in the first graph in Figure 14.

¹⁵In theory, it is certainly possible to have a medium whose total scattering coefficient/efficiency varies with the incident direction but not the angular distribution/probability of the scattered photons.

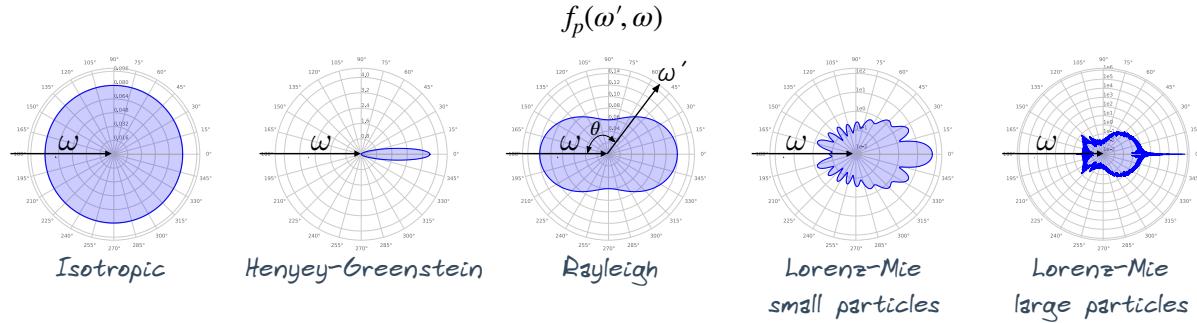


Figure 14: Visualizations of common phase functions; adapted from Novák et al. [2018], where ω is the incident direction ω_i and $\bar{\omega}$ is the outgoing direction ω_s in our notation. For an isotropic medium, the phase function depends on only the angle θ subtended by ω and ω' and is axially symmetric about ω .

8.3 Common Models and General “Rules”

There are many factors that determine the exact scattering efficiency and scattering direction, which can be calculated by solving the Maxwell’s equations. We will talk about a few common models here; we focus on the intuitions while omitting the exact mathematical expressions, which can be found in standard texts. From the models, we can identify a few general “rules” or, rather, approximations under certain assumptions.

The main theory or model for a single scattering event is called the **Mie scattering** theory, which, strictly speaking, applies only when the particle is spherical; see Sharma [2003, Chpt. 3.5.2], Bohren and Clothiaux [2006, Chpt 3.5], and Melbourne [2004, Chpt. 3]. Mie scattering is *not* somehow a different scattering process from any other scattering, and the Mie theory is nothing more than the solution to the Maxwell’s equations under certain conditions¹⁶.

The Mie theory predicts that the overall scattering efficiency Q_s is:

$$Q_s = \frac{8}{3} \gamma^4 \left(\frac{m^2 - 1}{m^2 + 2} \right) \left[1 + \frac{6}{5} \left(\frac{m^2 - 1}{m^2 + 2} \right) \gamma^2 + \dots \right], \quad (56a)$$

$$\gamma = \frac{r}{\lambda_m}, \quad (56b)$$

$$m = \frac{n}{n_m}, \quad (56c)$$

where m is the relative refractive index between the particle and the medium surrounding the particle, and γ is the ratio between the particle radius r and the incident light wavelength in the surrounding medium λ_m . The notion of surrounding media might come across as a little surprising: doesn’t the material consist merely of its particles? Hardly. For instance in paints, pigments are surrounded by binders (e.g., linseed oil in oil paints, egg yolk in tempera paints,

¹⁶The modern form of the solution is summarized, not invented, by Gustav Mie but the solution had been developed by many predecessors such as Ludvig Lorenz.

and beeswax in encaustic paints) and usually some amount of water (except oil paints). When paints dry, some water might be evaporated, leaving pockets of air, which also contributes to the surrounding media.

We can draw a few general conclusions from the model.

Small-Particle (Rayleigh) Scattering

For small particles where $\gamma \ll 1$ (generally when the radius is ten times smaller than the wavelength of the incident light), only the first term in Equation 56a’s bracket matters, so the scattering efficiency is inversely proportional to λ_m^{-4} . The inverse proportionality to λ_m^{-4} *largely* (but apparently not entirely) explains why sky is blue and why the sun is red [Bohren and Clothiaux, 2006, Chpt. 8.1]. Why? First of all, recognize that individual molecules, such as air molecules, are usually sub-nm in size so they scatter in this small-particle regime. Short wavelength lights from the sun are scattered by the atmospheric molecules more toward the sky and eventually enter your eyes, so if you look at the sky (against the sun) it would appear blue; when you look at the sun directly, the photons entering your eyes are mostly those unscattered ones that transmit directly through the atmosphere, and they are mostly longer-wavelength photons.

By then water molecules are also similarly small, so why would water look so different from the air? It is because water molecules are very densely packed so their scatterings are coherent. In fact, the end result of such coherent scatterings by a collection of water molecules is that water appears specular.

The photopigments in a photoreceptor are very small in size compared to the wavelengths of visible lights (each rhodopsin has a cross-section area of about 10^{-2} nm^2 [Milo and Phillips, 2015, p. 144]), so they almost do not scatter lights at all, only absorption. That is why we could use microspectrophotometry (MSP) to measure a photoreceptor’s (transverse) absorption rate: MSP measures the amount of light transmitted through a photoreceptor, and if there is little scattering then all the photons that are not measured must be absorbed by the photoreceptor.

Scattering in the small-particle regime is also called **Rayleigh scattering**, which, again, is *not* somehow a fundamentally different scattering process, and the Rayleigh scattering theory (worked out by Lord Rayleigh, who won the Nobel Prize in Physics in 1904) is nothing more than a special case of the Mie scattering theory; see Sharma [2003, Chpt. 3.5.1] and Bohren and Clothiaux [2006, Chpt. 3.2].

The phase function in the Rayleigh regime is proportional to $1 + \cos^2 \theta$, which looks like the third graph shown in Figure 14, where backward and forward scatterings are roughly equally probable. Taking the phase function into account, the scattering efficiency (in the form defined in Equation 54) in Rayleigh scattering is proportional to:

$$Q_s \propto \left(\frac{r}{\lambda_m}\right)^4 \frac{m^2 - 1}{m^2 + 2} (1 + \cos^2 \theta). \quad (57)$$

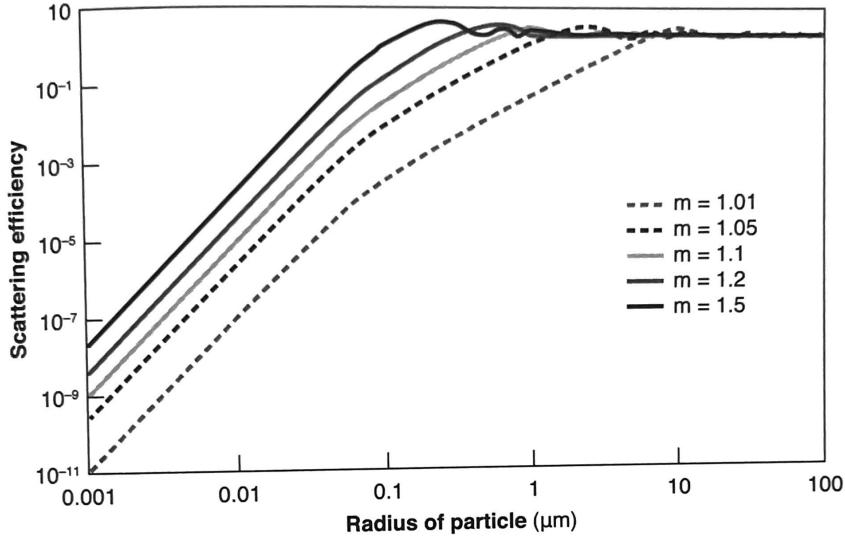


Figure 15: Scattering efficiency as a function of particle radius under different relative refractive index m ; the incident light wavelength is 500 nm. From [Johnsen \[2012\]](#), Fig. 5.4.

Impact of Particle Size

When the particle size increases, the scattering efficiency increases, initially very quickly, but eventually saturates. In fact, the Mie theory predicts that when the particle size is much larger than the wavelength (e.g., more than 100 times larger), the scattering efficiency approaches a constant 2 (regardless of m and λ_m). This is evident in Figure 15, which shows the scattering efficiency of a kind of particle as a function of particle radius (x -axis) under different m (different curves); the incident light wavelength is 500 nm.

The particle size also affects the phase function. As we have discussed above, small particles in the Rayleigh regime tend to scatter photons equally in the forward and backward directions while large particles primarily scatter photons in the forward directions. The last two graphs in Figure 14 show the phase functions predicted by the Mie scattering theory under different particle sizes (both are larger than that in the Rayleigh regime). The forward fraction increases as the particle size increases.

Consider the scenario in Figure 9, where Material 1 sits on top of Material 2, and our goal is to hide Material 2 so that the color of Material 1 is dependent only on the illumination (not the property of Material 2). There is an interesting trade-off between scattering efficiency and scattering direction here. If we want Material 1 to hide Material 2, we want the particles in Material 1 to scatter a lot of lights (high scattering efficiency) backwards. If the scattering efficiency is low (so photons march on and are hindered only by absorption) or the scattering is heavy in the forward directions, photons penetrate through Material 1 and reach Material 2, which would then contribute to the overall color.

Now, to scatter a lot of lights, we need the particles to be large, but then the scattering

will be mostly in the forward directions. So there exists a sweet spot of the particle size that provides the highest “hiding power” for a material per unit volume. If we work out the math, we will see that the sweet spot falls roughly in the visible wavelength range. That is why most paint pigments have a diameter between 100 nm and 1 μm [Bruce MacEvoy, 2015]. Of course, no matter how poor the hiding power is for a particular paint, if you apply enough of it, it would eventually hide whatever is behind it. Dye pigments are rather small in size ($\sim \text{nm}$ range), so they scatter little photons and that is why dye solutions look relatively transparent.

Impact of Refractive Index

Figure 15 also shows the impact of the relative refractive index m (between the particle and the surrounding media) on scattering efficiency. Generally, the scattering efficiency increases with m at all particle sizes until when the particles are so large that the scattering efficiency becomes a constant. This is supported by Equation 56a, too ($\frac{m^2-1}{m^2+2}$ monotonically increases and has a limit of 1).

For large particles, while m does not affect the scattering efficiency, it influences the scattering directions. When m is small, the scattering tends to be more forward whereas when m is large, the scattering tends to be toward large angles (i.e., more photons will be back-scattered). This is why wet objects look darker (recall the unpleasant experience of accidentally spilling water on your pants). In dry paints the medium surrounding the textile particles is air and in wet paints it is water. m becomes smaller when the material is wet (i.e., the relative refractive-index difference becomes smaller between the textile particles and water), so most of the scattering will be forward, increasing the traversal length of photons and essentially giving photons more opportunities to be absorbed.

Aspherical Particles

What if the particle is not spherical? The Mie theory does not apply. Analytical or even numerical solutions to the Maxwell’s equations would be difficult, so a perhaps better approach is just to parameterize a model and fit it with the experimental data.

One popular one-parameter parameterization of the phase function is the **Henyey–Greenstein** phase function; see [Pharr et al., 2018, Chpt. 11.3.1] and [Bohren and Clothiaux, 2006, Chpt. 6.3.2], which takes the form:

$$p(\theta) = \frac{1}{4\pi} \frac{1 - g^2}{(1 + g^2 - 2g \cos \theta)^{3/2}}, \quad (58)$$

where g is the free parameter and is usually called the **asymmetry parameter**.

We hasten to emphasize that the Henyey–Greenstein function has absolutely zero physical meaning; it is designed for fitting experimental phase function data, so in modern deep learning era you might as well try a deep neural network. The second graph in Figure 14 shows one instantiation of the Henyey–Greenstein function.

9 Radiative Transfer Equation and Volume Rendering

So far we have assumed that a ray can only be attenuated, which can happen only when the illumination is collimated and we assume single scattering. Under this assumption, Equation 48a allows us to calculate any radiance in the medium (by weakening the initial radiance). General media are much more complicated: illumination can be from anywhere and multiple scattering must be accounted for. As a result, external photons can be scattered into a ray of interest, as we have intuitively discussed in Chapter 8.1.

In the realm of geometric optics and radiometry, the general way to model lights going through a material/medium amounts to solving the so-called **Radiative Transfer Equation** (RTE), whose modern version was established by [Chandrasekhar \[1960\]](#), who won the Nobel prize in physics in 1983 (not for the RTE). The RTE provides a mathematical way to express an arbitrary radiance in a medium.

9.1 Radiative Transfer Equation

The basic idea is to setup a differential equation to describe the (rate) of the radiance *change*. Given an incident radiance $L(p, \omega_s)$, we are interested in $L(p + \Delta s \omega_s, \omega_s)$, the radiance after the ray has gone a small distance Δs . The radiance can be:

- attenuated by the medium because of absorption;
- attenuated by the medium because photons are scattered out into other directions; this is called **out-scattering** in graphics;
- augmented by photons that are scattered into the ray direction from all other directions — because of multiple scattering¹⁷; this is called **in-scattering** in graphics;
- augmented because particles can emit photons.

The attenuation (reduction) of the radiance over Δs is:

$$-L(p, \omega_s)\sigma_t(p, \omega_s)\Delta s. \quad (59)$$

The radiance augmentation due to in-scattering is given by:

$$\int_{\Omega=4\pi} f_p(p, \omega_s, \omega_i)\sigma_s(p, \omega_s)\Delta s L(\omega_i)d\omega_i = \sigma_s(p, \omega_s)\Delta s \int_{\Omega=4\pi} L(p, \omega_i)f_p(p, \omega_s, \omega_i)d\omega_i. \quad (60)$$

The way to interpret Equation 63 is the following. $L(p, \omega_i)$ is the incident radiance from a direction ω_i , $L(p, \omega_i)d\omega_i$ is the irradiance from ω_i over a small solid angle, of which $\sigma_s(p, \omega_i)\Delta s L(p, \omega_s)d\omega_i$ is the irradiance scattered in all directions after a distance Δs . That portion of the scattered

¹⁷Technically, even single scattering can lead to augmentation if there is illumination coming from anywhere outside the ray direction.

irradiance is multiplied by $f_p(\omega_s, \omega_i)$ to give us the amount of radiance that is scattered toward ω_s . We then integrate over the entire sphere, accounting for the fact that lights can come from anywhere over the space, to obtain the total augmented radiance toward ω_s .

If we consider emission, the total radiance augmentation is:

$$\sigma_a(p, \omega_s) \Delta s L_e(p, \omega_s) + \sigma_s(p, \omega_s) \Delta s \int^{\Omega=4\pi} L(p, \omega_i) f_p(p, \omega_s, \omega_i) d\omega_i. \quad (61)$$

where $L_e(p, \omega_s)$ is the emitted radiance at p toward ω_s so the first term represents the total emission over Δs . If we let:

$$L_s(p, \omega_s) = \sigma_a(p, \omega_s) L_e(p, \omega_s) + \sigma_s(p, \omega_s) \int^{\Omega=4\pi} L(p, \omega_i) f_p(p, \omega_s, \omega_i) d\omega_i, \quad (62)$$

the total augmentation can be simplified to:

$$L_s(p, \omega_s) \Delta s, \quad (63)$$

where the L_s term is sometimes called the **source term** or **source function** in computer graphics, because it is the source of power at p ¹⁸.

Combining Equation 59 and Equation 63, the net radiance change is¹⁹:

$$\Delta L(p, \omega_s) = L(p + \Delta s \omega_s, \omega_s) - L(p, \omega_s) \quad (64a)$$

$$= -L(p, \omega_s) \sigma_t(p, \omega_s) \Delta s + L_s(p, \omega_s) \Delta s. \quad (64b)$$

As Δs approaches 0, we get (assuming ω_s is a unit vector as in Equation 47 and Equation 48a):

$$\begin{aligned} \omega_s \cdot \nabla_p L(p, \omega_s) &= \frac{dL(p, \omega_s)}{ds} = \lim_{\Delta s \rightarrow 0} \frac{L(p + \Delta s \omega_s, \omega_s) - L(p, \omega_s)}{\Delta s} \\ &= -\sigma_t(p, \omega_s) L(p, \omega_s) + L_s(p, \omega_s), \end{aligned} \quad (65)$$

where ∇_p denotes the gradient of L with respect to p , and $\omega_s \cdot \nabla_p$ denotes the directional derivative, which is used because technically p and ω_s are both defined in a three-dimensional space, so what we are really calculating is the rate of radiance change at p along ω_s .

¹⁸Some definitions do not include emission in the source term while in other definitions the source term is what is defined here divided by σ_t .

¹⁹A subtlety you might have noticed is that not all the out-scattering of $L(p, \omega_s)$ attenuates the radiance; some of the scattering could be toward ω_s so should augment the radiance. This is not a concern since our augmentation term Equation 63 integrates over the entire sphere, so it considers $L(p, \omega_s)$ again as part of in-scattering and accounts for the forward-scattered portion of $L(p, \omega_s)$.

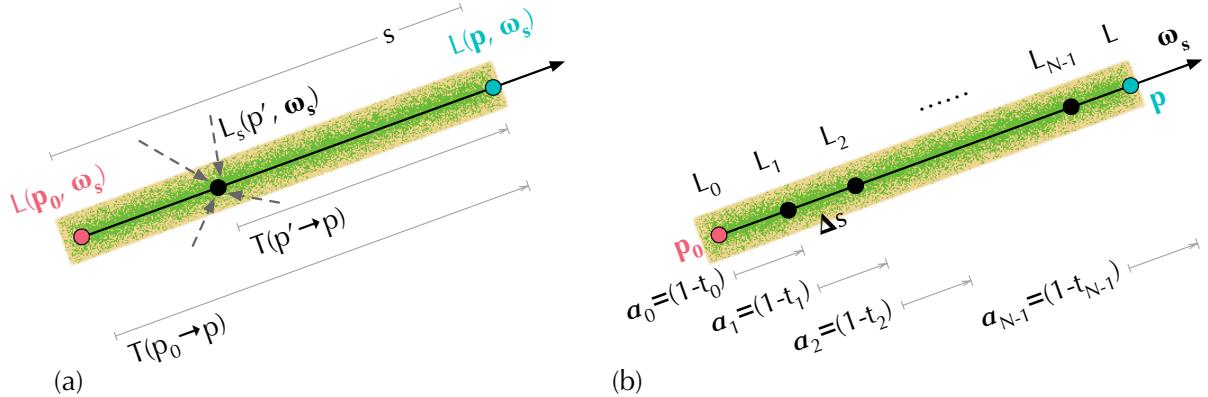


Figure 16: (a): Illustration of the continuous VRE (Equation 66). (b): Illustration of a discrete VRE (Equation 68a), where the integral in the continuous VRE is replaced by a summation between p_0 and p at an interval of Δs ; t_i is the total transmittance between p_i and p_{i+1} ; L_i is a shorthand for $L_s(p_i, \omega_s)$, the source term of at p_i toward ω_s .

Equation 65 is the RTE, which is an integro-differential equation, because it is a differential equation with an integral embedded. The RTE has an intuitive interpretation: if we think of radiance as the power of a ray, as a ray propagates its power is attenuated by the medium but also augmented by “stray photons” from other rays. The latter is given by $L_s(p, \omega_s)$, which can be thought of as the augmentation of the radiance per unit length.

The RTE describes the rate of change of an arbitrary radiance $L(p, \omega_s)$. But our ultimate goal is to calculate the radiance itself? Generally the RTE has no analytical solution. There are two strategies to solve it. First, we can derive analytical solutions under certain assumptions and simplifications.

- For instance, the integral in Equation 65 can be approximated by a summation along N directions; then we can turn Equation 65 into a system of N differential equations to be solved. This is sometimes called the **N-flux theory**. We will see one such example in Chapter 10 where $N = 2$.
- Another assumption people make is to assume that volume scattering is isotropic and can be approximated as a *diffusion* process. This is called the **diffusion approximation** [Ishimaru, 1977; Ishimaru et al., 1978], which is widely used in both scientific modeling [Farrell et al., 1992; Eason et al., 1978; Schweiger et al., 1995; Boas et al., 2001] and in rendering [Stam, 1995; Jensen et al., 2001; Dong et al., 2013, Chpt. 7]; see Bohren and Clothiaux [2006, Chpt. 6.2] for a theoretical treatment.

9.2 Volume Rendering Equation

The second approach, which is particularly popular in computer graphics, is to first turn the RTE into a purely integral equation and then *numerically* (rather than analytically) estimate

the integral using Monte Carlo integration, very similar to how the rendering equation is dealt with for surface scattering (Chapter 4.3).

The way to think of this is that in order to calculate any given radiance $L(p, \omega_s)$, we need to integrate all the changes along the direction ω_s up until p . Where do we start the integration? We can start anywhere. Figure 16(a) visualizes the integration process. Let's say we want to start from a point p_0 , whose initial radiance toward ω_s is $L_0(p_0, \omega_s)$. Let $p = p_0 + s\omega_s$, where ω_s is a unit vector and s is the distance between p_0 and p . An arbitrary point p' between p_0 and p would then be $p' = p_0 + s'\omega_s$ ²⁰.

Now we need to integrate from p_0 to p by running s' from 0 to s . Observe that the RTE is a form of a *non-homogeneous* linear differential equation, whose solution is firmly established in calculus. Without going through the derivations, its solution is:

$$L(p, \omega_s) = T(p_0 \rightarrow p)L_0(p_0, \omega_s) + \int_0^s T(p' \rightarrow p)L_s(p', \omega_s)ds', \quad (66)$$

where $T(p_0 \rightarrow p)$ is the transmittance between p_0 and p along ω_s , and $T(p' \rightarrow p)$ is the transmittance between p' and p along ω_s . Recall the definition of transmittance in Equation 48b: it is the remaining fraction of the radiance after attenuation by the medium after traveling the distance between two points. In our case here:

$$T(p' \rightarrow p) = \frac{L(p + s\omega_s, \omega_s)}{L(p + s'\omega_s, \omega_s)} = e^{-\int_{s'}^s \sigma_t(p + t\omega_s, \omega_s)dt}, \quad (67a)$$

$$T(p_0 \rightarrow p) = \frac{L(p + s\omega_s, \omega_s)}{L(p, \omega_s)} = e^{-\int_0^s \sigma_t(p + t\omega_s, \omega_s)dt}, \quad (67b)$$

This integral equation in the graphics literature is called the **volume rendering equation** (VRE) or the **volumetric light transport equation** — the counterpart of the surface LTE (Chapter 4.3). Looking at the visualization in Figure 16(a), the VRE has an intuitive interpretation: the radiance at p along ω_s is the the contribution of p_0 plus and contribution of every single point between p_0 and p .

- The contribution of p_0 is given by its initial radiance L_0 weakened by the transmittance between p_0 and p ;
- Why would a point p' between p_0 and p make any contribution? It is because of the source term (Equation 62): p' might emit lights and some of the in-scattered photons at p' will be scattered toward ω_s . The contribution of p' is thus given by the source term L_s weakened by the transmittance between p' and p .

²⁰There are two alternative parameterizations, both of which are common in graphics literature. The first [Pharr et al., 2023] is to express $p_0 = p + s\omega_s$ (s being positive), but then the initial radiance would have to be expressed as $L(p_0, -\omega_s)$, since ω_s now points from p to p_0 . The other is to express $p_0 = p - s\omega_s$ (s again being positive) [Fong et al., 2017]; this avoids the need to switch directions but uses a negative sign. It is a matter of taste which one to use, but be alert of the different conventions.

The form of the VRE might appear to suggest that it is enough to accumulate along only the *direct* path between p_0 and p , which is surprising given that there are infinitely many scattering paths between p_0 and p (due to multiple scattering). The answer is that the VRE *implicitly* accounts for all the potential paths between p_0 and p — because of the L_s term, which expands to Equation 62. That is, every time we accumulate the contribution of a point between p_0 and p , we have to consider the in-scattering from all the directions at that point. Another way to interpret this is to observe that the radiance term L appears on both sides of the equation. Therefore, the VRE must be solved recursively by evaluating it everywhere in space.

Does this remind you of the rendering equation (Equation 23)? Indeed, the VRE can be thought of as the volumetric counterpart of the rendering equation. Similarly, we can use Monte Carlo integration to estimate it, just like how the rendering equation is dealt with — with an extra complication: the VRE has two integrals, the outer integral runs from p_0 to p and, for any intermediate point p' , there is an inner integral that runs from p' to p to evaluate the transmittance $T(p' \rightarrow p)$. Therefore, we have to sample both integrands.

Similar to the situation of the rendering equation, sampling recursively would exponentially increase the number of rays to be tracked. Put it another way, since there are infinitely paths from which a ray gains its energy due to multiple scattering, we have to integrate infinitely many paths. Again, a common solution is path tracing, for which Pharr et al. [2023, Chpt. 14] is a great reference.

A simplification that is commonly used is to assume that there is only single scattering directly from the light source. In this way, the L_s term does not have to integrate infinitely many incident rays over the sphere but only a fixed amount of rays emitted from the light source *non-recursively*. This strategy is sometimes called **local illumination** in volume rendering, as opposed to **global illumination**, where one needs to consider all the possible paths of light transport. The distinction is similar to that in modeling surface scattering (Chapter 4.3).

9.3 Discrete VRE and Scientific Volume Visualization

Sometimes the VRE takes the following discrete form:

$$L = \sum_{i=0}^{N-1} \left(L_i \Delta s \prod_{j=i+1}^{N-1} t_j \right) \quad (68a)$$

$$= \sum_{i=0}^{N-1} \left(L_i \Delta s \prod_{j=i+1}^{N-1} (1 - \alpha_j) \right) \quad (68b)$$

$$\begin{aligned} &= L_{N-1} \Delta s + L_{N-2} \Delta s (1 - \alpha_{N-1}) + L_{N-3} \Delta s (1 - \alpha_{N-1})(1 - \alpha_{N-2}) + \dots \\ &\quad + L_1 \Delta s \prod_{j=2}^{N-1} (1 - \alpha_j) + L_0 \Delta s \prod_{j=1}^{N-1} (1 - \alpha_j). \end{aligned} \quad (68c)$$

Equation 68a looks very much like Equation 66: the former turns the two integrals in the latter, both the outer integral and the inner one carried by $T(\cdot)$, to discrete summations using the

Riemann sum over N discrete points along the ray between p_0 and p at an interval of $\Delta s = \frac{s}{N}$.

The notations are slightly different; Figure 16(b) visualizes how this discrete VRE is expressed with the new notations.

- L is $L(p, \omega_s)$, the quantity to be calculated;
- L_i is a shorthand for $L_s(p_i, \omega_s)$, i.e., the source term (Equation 62) for the i^{th} point between p_0 and p toward ω_s ; by definition, p_0 is the 0^{th} point (so L_0 is the initial radiance $L_0(p_0, \omega_s)$ in Equation 66) and p is the N^{th} point;
- t_i represents the total transmittance between the i^{th} and the $(i + 1)^{th}$ point and is given by $e^{-\sigma_t(p_i, \omega_s)\Delta s}$ (notice the integral in continuous transmittance Equation 48b is gone, because we assume the transmittance between adjacent points is a constant in the Riemann sum);
- α_i is the **opacity** between the i^{th} and the $(i + 1)^{th}$ point, which is defined as the residual of the transmittance between the two points: $1 - t_i$.

See Max [1995, Sect. 4] or Kaufman and Mueller [2003, Sect. 6.1] for a relatively straightforward derivation, but hopefully this form of the VRE is equally intuitive to interpret from Figure 16(b). It is nothing more than accumulating the contribution of each point²¹ along the ray, but now we also need to accumulate the attenuation along the way just because of how opacity is defined by convention (per step), hence the product of a sequence of the opacity residuals.

This way of computing the VRE is usually used in the scientific visualization literature, where people are interested in visualizing data obtained from, e.g., computer tomography (CT) scans or magnetic resonance imaging (MRI). There, it is the relative color that people usually care about, not the physical quantity such as the radiance, so people sometimes lump $L_i \Delta s$ together as C_i and call it the “color” of the i^{th} point. The VRE is then written as:

$$C = \sum_{i=0}^{N-1} \left(C_i \prod_{j=i+1}^{N-1} (1 - \alpha_j) \right). \quad (69)$$

The C terms are defined in a three-dimensional RGB space, and Equation 69 is evaluated for the three channels separately, similar to how Equation 68a and Equation 66 are meant to be evaluated for each wavelength independently. Since color is a linear projection from the spectral radiance, the so-calculated C (all three channels) is indeed proportional to the true color, although in visualization one usually does not care about the true colors anyways (see p. 63).

This formulation is also called the *back-to-front* compositing formula in volume rendering, since it starts from p_0 , the farthest point on the ray to p . We can easily turn the order around to start from p and end at p_0 in a *front-to-back* fashion (C_{N-1} now corresponds to p_0):

²¹technically the contribution of each small segment between two discrete points because of the Riemann sum.

$$C = \sum_{i=0}^{N-1} \left(C_i \prod_{j=0}^{i-1} t_j \right). \quad (70)$$

While theoretically equivalent, the latter is better in practice because it allows us to opportunistically terminate the integration early when, for instance, the accumulated opacity is high enough (transmittance is low enough), at which point integrating further makes little numerical contribution to the result.

Another Discrete Form of VRE

A perhaps more common way to express the discrete VRE is to approximate the average transmittance t using the first two terms of its Taylor series expansion and further assume that the medium has a low albedo, i.e., $\sigma_t \approx \sigma_a$ and $\sigma_s \approx 0$ (that is, the medium emits and absorbs *only*); we have:

$$1 - \alpha_i = t_i = t(p_i \rightarrow p_{i+1}) = e^{-\sigma_t(p_i, \omega_s)\Delta s} = 1 - \sigma_t(p_i, \omega_s)\Delta s + \frac{(\sigma_t(p_i, \omega_s)\Delta s)^2}{2} - \dots \quad (71a)$$

$$\approx 1 - \sigma_t(p_i, \omega_s)\Delta s \quad (71b)$$

$$\approx 1 - \sigma_a(p_i, \omega_s)\Delta s. \quad (71c)$$

$$\Rightarrow \alpha_i \approx \sigma_a(p_i, \omega_s)\Delta s. \quad (71d)$$

Now, observe that the L_i term in Equation 68a is the source term in Equation 62, which under the low albedo assumption has only the emission term, so:

$$L = \sum_{i=0}^{N-1} \left(L_i \Delta s \prod_{j=i+1}^{N-1} (1 - \alpha_j) \right) \quad (72a)$$

$$= \sum_{i=0}^{N-1} \left(\sigma_a(p_i, \omega_s) L_e(p_i, \omega_s) \Delta s \prod_{j=i+1}^{N-1} (1 - \alpha_j) \right), \quad (72b)$$

$$= \sum_{i=0}^{N-1} \left(L_e(p_i, \omega_s) \alpha_i \prod_{j=i+1}^{N-1} (1 - \alpha_j) \right). \quad (72c)$$

If we lump $L_e(p_i, \omega_s)\alpha_i$ together and call it C_i , the VRE is then expressed as [Levoy, 1988]:

$$C = \sum_{i=0}^{N-1} \left(C_i \alpha_i \prod_{j=i+1}^{N-1} (1 - \alpha_j) \right). \quad (73)$$

Again, this is the back-to-front equation, and the front-to-back counterpart looks like:

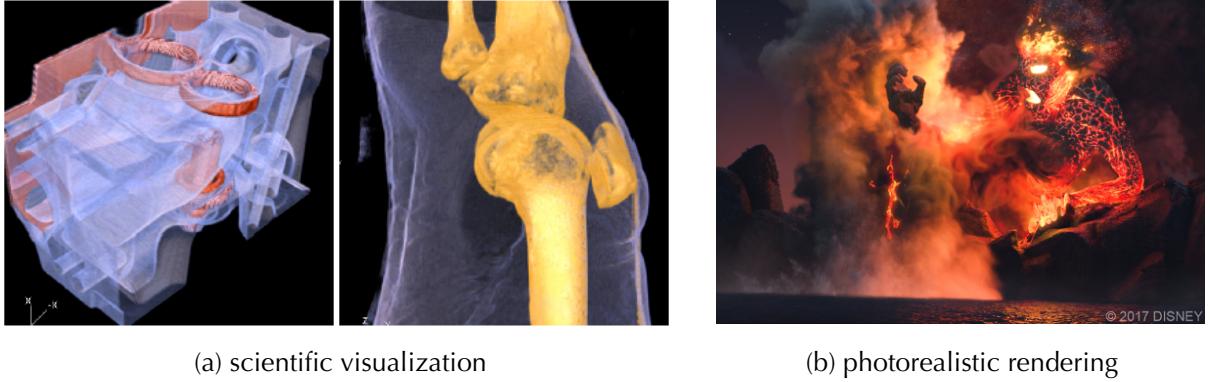


Figure 17: (a): two examples of scientific visualization (of CT data) using volume rendering; adapted from [Kaufman and Mueller \[2003, Fig. 1\]](#). (b): photorealistic volume rendering (a scene from Disney’s Moana, 2016); from [Fong et al. \[2017, Fig. 1\]](#).

$$C = \sum_{i=0}^{N-1} (C_i \alpha_i \prod_{j=0}^{i-1} t_j). \quad (74)$$

If you compare the two discrete forms in Equation 69 and Equation 73, it would appear that the two are not mutually consistent! Of course we know why: 1) Equation 73 applies two further approximations (low albedo and Taylor series expansion) *and* 2) the two C terms in the two equations refer to different physical quantities (compare Equation 68b with Equation 72c).

Visualization is Not Rendering!

These discrete VRE forms might give you the false impression that we have avoided the need to integrate infinitely many paths, because, computationally, the evaluation of the VRE comes down to a *single-path* summation along the ray trajectory. Not really. Calculating the C_i terms in the new formulations still requires recursion if the results are meant to be physically accurate. Of course we can side-step this by, e.g., applying the local-illumination approximation, as mentioned before to avoid recursion.

Scientific visualization offers another opportunity: we can simply *assign* values to the C s and even the α s without regard to physics. The goal of visualization is to discover/highlight interesting objects and structures while de-emphasizing things that are irrelevant. So the actual colors are not as important, which gives us great latitude to determine VRE parameters.

Figure 17 compares volume-rendered images for scientific visualization (a) and for photorealistic rendering (b). In the case of visualization, the data was a CT-scanned computer-aided design (CAD) model and a CT-scanned human knee model. In both cases, the outer surface is not transparent, but is rendered so just because we are interested in seeing the inner structures that are otherwise occluded. The user makes an executive call to assign a very low transparency

to the bones in the knee model 0 but very a high transparency value to the skin and other tissues: this is not physically accurate but a good choice for this particular visualization. Photorealistic rendering, in contrast, has to be physically based, and does not usually have this flexibility. See figures in [Wrenninge and Bin Zafar \[2011\]](#) and [Fong et al. \[2017\]](#) for more examples.

There are volume rendering software that would allow the users to make such an assignment depending on what the user wants to highlight and visualize. With certain constraints and heuristics, one can also procedurally assign the α and C values from the raw measurement data, usually a density field (see below) acquired from whatever measurement device is used (e.g., CT scanners or MRI machines), using what is called the *transfer functions* in the literature²². Making an assignment usually is tied to a *classification* problem: voxels/points of different classes should have different assignments.

Density Fields

Physically speaking the medium in RTE/VRE is parameterized by its absorption and scattering coefficients, which is a product of cross section and concentration, which is sometimes also called the density. In physically-based volume rendering, this is indeed how the density is used from the very beginning [[Kajiya and Von Herzen, 1984](#); [Blinn, 1982](#)].

In visualization where being physically accurate or photorealistic is unimportant, the notion of an attenuation coefficient²³ loses its physical meaning; it is just a number that controls how the brightness of a point weakens. People simply call the attenuation coefficient the density [[Kaufman and Mueller, 2003](#)], presumably because, intuitively, if the particle density is high the color should be dimmer. If you want to be pedantic, you might say that the attenuation coefficient depends not only on the density/concentration, but also on the cross section (Equation 40b), so how can we do that? Remember in visualization one gets to make an executive call and *assign* the density value (and thus control α), so it does not really matter if the value itself means the physical quantity of density/concentration. This is apparent in early work that uses volume rendering for scientific visualization [[Sabella, 1988](#); [Williams and Max, 1992](#)], where attenuation coefficients are nowhere to be found.

For this reason, the raw volume data obtained from raw measurement device for scientific (medical) visualization are most often called the density field, even though what is being measured is almost certainly not the density field but a field of optical properties that are related to, but certainly do not equate, density. For instance, the raw data you get from a CT scanner is actually a grid of attenuation coefficients [[Bharath, 2009](#), Chpt. 3].

9.4 Discrete VRE in Neural Rendering

There is another field where the discrete VREs are becoming incredibly popular: neural rendering. The two most representative examples are NeRF [[Mildenhall et al., 2021](#)] and 3DGS [[Kerbl](#)

²²Some (color) transfer functions could have physical underpinnings, such as applying a single-scattering shading algorithms (i.e., local illumination); see, e.g., [Levoy \[1988, Sect. 3\]](#) or [Max \[1995, Sect. 5\]](#), but the goal there is not to precisely model physics but for better, subjective visualization.

²³which is the only coefficient needed and which participates in calculating α (Equation 69).

[et al., 2023](#). They use the discrete form of the VRE (mostly Equation 74), so the evaluation is a single-path summation along the ray trajectory without the need for path tracing and solving the actual RTE/VRE.

The reason they can reduce infinite paths to a single-path evaluation is very similar to that in scientific visualization, except now instead of assigning the “color” values C and opacity values α (or equivalently the density field as discussed above), they train a neural network to directly learn these values. The ground truth in these methods is offline rendered/captured images from different camera poses. So they are also a form of image-based rendering or light-field rendering (Chapter 3.7).

Among many things, one difference between NeRF and 3DGS is one between image-order rendering and object-order rendering. Image-order rendering cast rays (usually one per pixel in volume rendering²⁴) to the scene, essentially directly evaluating the discrete VRE. This is how NeRF-family models operate. In contrast, the object-order rendering *projects* points in a medium/volume to the image/sensor plane. A classic technique is called **splatting** [Westover, 1990], which is used in 3DGS-family models. We will discuss splatting on p. 66, at which point we will bring up another important distinction between NeRF and 3DGS.

The difference between image-order vs. object-order rendering is a fundamental one, not limited to NeRF vs. 3DGS. The distinction exists in both traditional (non-neural) volume rendering, where ray casting vs. splatting are also the two main rendering approaches (see, e.g., [Engel et al. \[2006, Chpt. 1.6\]](#) and [Kaufman and Mueller \[2003, Sect. 6\]](#)), and in classical (surface rather than volume) rendering, where rasterization is an object-order rendering algorithm and ray tracing is an image-order algorithm.

A huge amount of work is being done on accelerating neural volume rendering models, both algorithmically and using special hardware, reminiscent of much of the early work on accelerating volume rendering [[Kaufman and Mueller, 2003](#); [Engel et al., 2006, Chpt. 8](#)]. Classic techniques such as re-sampling/interpolating from pre-computed samples, using hierarchical data structures, leaping empty spaces/early termination enjoy great success in accelerating neural volume rendering, too.

VRE for Surface Rendering?

It is interesting to observe that both NeRF and 3DGS (and the vast majority of their later developments) use VRE to render (opaque) surfaces rather than volumes/media. Is this surprising? If we are rendering an opaque surface, why would any other point on the ray contribute to color of the corresponding pixel? Shouldn’t we use the rendering equation?

There are two things to note here. First, from a mathematical and physical perspective, we should not be surprised at all. We can think of the rendering equation as a special case of the RTE; ultimately they are both concerned with light transport, whether it is between surface points or between volume particles. This is precisely why, mathematically, the rendering

²⁴To be physically accurate and photorealistic we need to cast many rays and reconstruct from the ray samples (anti-aliasing). One ray per pixel is OK if we do not care about physical accuracy/photorealism (e.g., in visualization) or we are training a neural network to render, in which case the network weights, ideally, are trained to mitigate aliasing.

equation (Equation 23) very much resembles the VRE (Equation 66), where the result on the right side of the equations requires an integration of a bunch of stuff on the left side. What the neural network learns is the stuff to be summed on the right, and whether you interpret them as things to be summed in the VRE or in the rendering equation is completely up to you.

Second, as far as the neural volume rendering models are concerned, they are using the VRE just as a way to parameterize the forward model for training. The learned parameters (color and opacity of each point) should not be interpreted literally in the physical sense. One advantage of this parameterization is it *could* be used to render volumes or translucent materials if needed, in which case the learned parameters might be more amenable to physical interpretations.

Volume Graphics vs. Point-Based Graphics

Related to volume graphics, there is also a subtly different branch of graphics called **point-based graphics** (PBG) [Levoy and Whitted, 1985; Gross and Pfister, 2011]. The boundary is somewhat blurred, but given the way the two terms are usually used, we can observe a few similarities and distinctions. Both volume graphics and PBG use discrete points as the rendering primitives (as opposed to continuous surfaces such as a mesh), although the input points in volume graphics are usually placed on uniform grids [Engel et al., 2006, Chpt. 1.5.2] whereas points in PBG can be spatially arbitrary.

Traditionally, PBG is almost exclusively used for photorealistic rendering of surfaces. In fact, the points used in PBG are usually acquired from continuous surfaces as samples on the surface [Gross and Pfister, 2011, Chpt. 3]. PBG usually uses object-order rendering through splatting although ray casting is used too, but RTE/VRE is not involved in the rendering process [Gross and Pfister, 2011, Chpt. 6].

In contrast, the use of volume graphics is much broader. Volume rendering can be used for photorealistic rendering of participating media and translucent surfaces (by solving the RTE/VRE), or it can be used for non-photorealistic data visualization (by evaluating the single-path, discrete VRE), at which point whether the object to be rendered is called a participating medium, a translucent surface, or anything else is irrelevant, because visualization does not care much about physics.

3DGS is a somewhat interesting case. It is largely a form of PBG because the rendering primitives are unaligned points and its splatting technique (Gaussian splatting) resembles that developed in the PBG literature [Gross and Pfister, 2011, Chpt. 6.1]. However, 3DGS does use the discrete VRE as the forward model. Again, VRE and splatting are just ways for 3DGS to parameterize its forward mode, so the comparison with traditional volume graphics and PBG should not be taken literally.

Splatting is Signal Filtering

Splatting, initially prosed by Westover [1990] for volume rendering, is a common rendering technique used in PBG and 3DGS-family models for rendering continuous surfaces. How can we render continuous surfaces from discrete points? If we directly project the points to the sensor

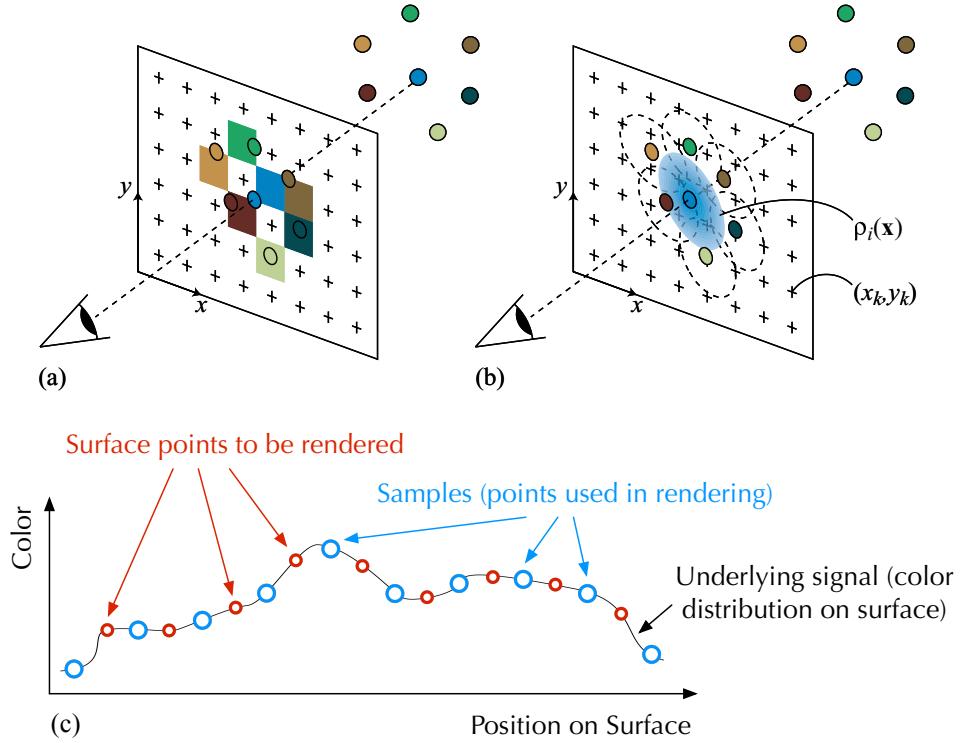


Figure 18: (a): directly projecting discrete points to the image plane would create holes in the rendered image. (b): in splatting, each point is associated with a splat or a footprint function, which can distribute the color of the point to a spatial region on the image plane; $\rho_i(\mathbf{x})$ represents the value of the *footprint function* of the i^{th} point at position \mathbf{x} in the image plane. (c): splatting essentially allows signal interpolation, which amounts to first reconstructing the underlying signal from the samples (with potential anti-aliasing filtering) followed by re-sampling at new, desired positions. (a-b) are from Gross and Pfister [2011, Fig. 6.1].

plane, we obviously will get holes, as shown in Figure 18(a). This is of course not an issue if the rendering primitives are meshes (or procedurally-generated surfaces).

The key is to realize that the fact that we do not have a surface as the rendering primitive does not mean the surface does not exist. Recall that the points used by PBG are actually samples on the surface. To render an image pixel is essentially to estimate the color of a surface point that projects to the pixel (ignoring supersampling for now). From a signal processing perspective, this is a classic problem of signal filtering: reconstruction and resampling.

That is, ideally what we need to do is to reconstruct the underlying signal, i.e., the color distribution of the continuous surface²⁵, followed by anti-aliasing filters (i.e., convolution) and then resample the reconstructed/filtered signal at positions corresponding to pixels in an image.

²⁵assuming a diffuse surface so we care to reconstruct the color of each point not the radiance of each ray.

This is shown in Figure 18(c). The name of the game is to design proper filters. The issue of signal sampling, reconstruction, and resampling is absolutely fundamental to all forms of photorealistic rendering and not limited to PBG; Pharr et al. [2023, Chpt. 8] and Glassner [1995, Unit II] are great references.

Another way to think of this is that the color of a surface point is very likely related to its nearby points that have been sampled as part of the rendering input, so one thing straightforward to do is to interpolate from those samples. Signal interpolation is essentially signal filtering/convolution.

A splat, or a *footprint function* in the PBG parlance, associated with each point (surface sample) essentially distributes the point color to a local region, enabling signal interpolation. This is shown in Figure 18(b). The exact forms of the footprint functions would determine the exact forms of the signal filters²⁶. Gaussian filters are particularly common, and Gaussian splatting is a splatting method that uses Gaussian filters. 3DGS-family models learn the filters (jointly with the samples!) from data whereas traditionally the filters are more or less designed empirically by heuristics.

Once you understand splatting, you might notice a very subtle difference in NeRF and 3DGS: both use the VRE as the forward model, but their uses are slightly different. NeRF uses VRE in a way similar to how traditionally the discrete VRE is used: approximating a continuous integral with a discrete summation. 3DGS, however, does not actually need the VRE; what it *does* need is signal filtering/interpolation, which requires summing over a set of samples. VRE happens to sum over a bunch of things and is shown useful in NeRF anyways so it certainly is not unreasonable for 3DGS to keep using it.

9.5 Integrating Surface Scattering with Volume Scattering

The rendering equation governs the surface scattering or light transport in space, and the RTE/VRE governs the volume/subsurface scattering or light transport in a medium. Both processes can be involved in a real-life scene. For instance, the appearance of a translucent material like a paint or a wax is a combination of both forms of scattering/light transport (Figure 9). Another example would be rendering smoke against a wall.

Conceptually nothing new needs to be introduced to deal with the two forms of light transport together. Say we have an opaque surface (a wall) located with a volume (smoke) in the scene. If we want to calculate the radiance of a ray leaving a point on the wall, we would evaluate the rendering equation there, and for each incident ray, we might have to evaluate the VRE since that ray might come from the volume. In practice it amounts to extending the path tracing algorithm to account for the fact that a path might go through a volume and bounce off between surface points. See Pharr et al. [2023, Chpt. 14.2] and Fong et al. [2017, Sect. 3] for detailed discussions.

Another approach, which is perhaps more common when dealing with translucent materials (whose appearance of course depends on both the surface and subsurface scattering), is through

²⁶If all the footprint functions are the same, the effect of splatting is equivalent to applying a single interpolation filter/convolution kernel, but in PBG each source function can be different.

a phenomenological model based on the notion of Bidirectional Scattering Surface Reflectance Distribution Function (**BSSRDF**) [Nicodemus et al., 1977]. The BSSRDF is parameterized as $f_s(p_s, \omega_s, p_i, \omega_i)$, describing the infinitesimal outgoing radiance at p_s toward ω_s given the infinitesimal power incident on p_i from the direction ω_i :

$$f_s(p_s, \omega_s, p_i, \omega_i) = \frac{dL(p_s, \omega_s)}{d\Phi(p_i, \omega_i)}. \quad (75)$$

BSSRDF can be seen as an extension of BRDF in that it considers the possibility that the radiance of a ray leaving p_s could be influenced by a ray incident on another point p_i — because of SSS/volume scattering. Given the BSSRDF, the rendering equation can be generalized to:

$$L(p_o, \omega_s) = \int^A \int^{\Omega=2\pi} f_s(p_s, \omega_s, p_i, \omega_i) L(p_i, \omega_i) \cos \theta_i d\omega_i dA, \quad (76)$$

where $L(p_o, \omega_s)$ is the outgoing radiance at p_o toward ω_s , $L(p_i, \omega_i)$ is the incident radiance at p_i from ω_i , A in the outer integral is the surface area that is under illumination, and $\Omega = 2\pi$ means that each surface point receives illumination from the entire hemisphere.

We can again use path tracing and Monte Carlo integration to evaluate Equation 76 if we know the BSSRDF, which can, again, either be analytical derived given certain constraints and assumptions or measured [Frisvad et al., 2020]. To analytically derive it, one has to consider the fact that the transfer of energy from an incident ray to an outgoing ray is the consequence of a cascade of three factors: two surface scattering (refraction) factors, one entering the material surface p_i from ω_i and the other leaving the material surface at p_i toward p_o , and a volume scattering factor that accounts for the subsurface scattering between the incident ray at p_i and the exiting ray at p_o [Pharr et al., 2018, Chpt. 11.4]. If all three factors have an analytical form, the final BSSRDF has an analytical form too. This is the approach that, for instance, Jensen et al. [2001] takes.

10 The Kubelka–Munk Model

As discussed in Chapter 9.1, the general RTE is difficult to solve, and there are two general strategies. Chapter 9.2 discusses one strategy that numerically approximates the solution using Monte Carlo methods. Another common strategy is to make some simplified assumptions and/or apply additional constraints, which would allow us to derive analytical solutions. This section discusses perhaps the most aggressive form of simplifications that, nevertheless, is very widely used, especially in printing, painting, and dye industry (and to some extent in graphics).

Kubelka and Munk, two Czechoslovakian Chemists built a phenomenological model that estimates the spectral reflectance/transmittance of a material [Kubelka and Munk, 1931b,a; Kubelka, 1948]. Their model cares only about the hemispherical-hemispherical reflectance (Chapter 4.2), i.e., the ratio of total flux scattered upward to the hemisphere to the total

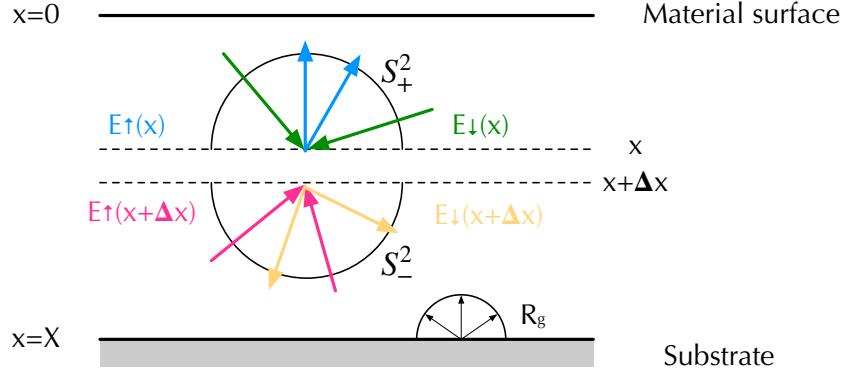


Figure 19: The setup of deriving the K-M model. We focus on the change of downward irradiance (E_{\downarrow}) and upward irradiance (E_{\uparrow}) as they travel a thin layer Δx . We make the assumption that the material is an isotropic medium and is homogeneous spatially.

toward flux incident from the hemisphere. The K-M model also considers that the material under modeling is in immediate contact with a *Lambertian* substrate that reflects lights uniformly over all directions. This is a common scenario in scenarios such as paints on a canvas, dyes on textiles, printer inks on paper, coatings on films, etc.

We will go through an excruciatingly long derivation, but the intuitions behind the derivation are exactly the same as that of the RTE, because the K-M model is a simplification of the RTE. The derivation allows us to make clear what simplifications we have made and, thus, when the model is and is not applicable.

10.1 Deriving the Model

Figure 19 illustrates the setup that we will use to derive the model. The first important assumption that the K-M model makes is that the every point on the material surface receives exactly the same irradiance and that the material itself is homogeneous, consisting of particles that are statistically randomly distributed and oriented. As a result, there is no difference between different positions at the same depth anywhere inside the material. Of course the radiation field at different depths are different; at the very least, the deeper you go the fewer photons there is due to absorption. Another way to think of this that we are intentionally limiting our consideration to only a small area where there is no spatial difference at the same depth, so we can analyze information only at different depths rather than different positions at the same depth.

We focus on a very thin layer Δx ; since there is no differences between horizontal positions, we can arbitrarily pick a point at depth $0 \leq x \leq X$ for analysis, where X is the total depth (the material surface has $x = 0$ and the material bottom has $x = X$). The point at x receives photons from all directions. Let's say all the downward photons (going to the lower hemisphere) have a total irradiance of $E_{\downarrow}(x)$ and all the upward photons (going to the upper hemisphere) have a total irradiance of $E_{\uparrow}(x)$. The (hemispherical-hemispherical) reflectance at x is then the

ratio of the two:

$$R(x) = \frac{E_{\uparrow}(x)}{E_{\downarrow}(x)}, \quad (77)$$

and the reflectance at the material surface (which is what the K-M model is interested in calculating) is $R(0)$.

How do we express $E_{\downarrow}(x)$? As usual we setup a differential equation to describe the change of $E_{\downarrow}(x)$. Consider the following conservation of energy when $E_{\downarrow}(x)$ goes through the thin layer Δx :

$$E_{\downarrow}(x + \Delta x) - E_{\downarrow}(x) = - \int^{S^2_+} \sigma_a(x, \omega) \Delta x L(x, \omega) d\omega \quad (78a)$$

$$- \int^{S^2_+} \int^{S^2_+} \sigma_s(x, \omega) f_p(x, \omega', \omega) \Delta x L(x, \omega) d\omega' d\omega \quad (78b)$$

$$+ \int^{S^2_-} \int^{S^2_-} \sigma_s(x + \Delta x, \omega) f_p(x + \Delta x, \omega', \omega) \Delta x L(x + \Delta x, \omega) d\omega' d\omega, \quad (78c)$$

where S^2_+ and S^2_- represents the upper and lower hemisphere, respectively; $L(x, \omega)$ is the radiance coming from direction ω incident on a point of depth x , $f_p(x, \omega', \omega)$ is the phase function between the incident direction ω and outgoing direction ω' , and $\sigma_s(x, \omega)$ is the scattering coefficient at x for an incident direction ω (Chapter 8.2).

The interpretation of this equation is exactly the same as that of the RTE. The equation essentially says that the downward irradiance after traveling Δx is (omitting emission):

- reduced by absorption (Equation 78a), which is the absorption component in the RTE;
- reduced by upward scattering (Equation 78b), which is the out-scattering component in the RTE;
- increased by the downward-scattering of upward irradiance reaching the bottom of the thin layer (Equation 78c), which is the in-scattering component in RTE.

Now let's take a closer look at Equation 78a and re-write it:

$$\int^{S^2_+} \sigma_a(x, \omega) \Delta x L(x, \omega) d\omega \quad (79a)$$

$$= \sigma_a(x, \bar{\omega}) \Delta x \int^{S^2_+} L(x, \omega) d\omega \quad (79b)$$

$$= K_{\downarrow}(x) \Delta x E_{\downarrow}(x), \quad (79c)$$

$$K_{\downarrow}(x) = \frac{\int^{S^2_+} \sigma_a(x, \omega) L(x, \omega) d\omega}{E_{\downarrow}(x)}. \quad (79d)$$

Equation 79b is derived based on the **mean-value theorem**²⁷ in integral calculus, which says that there exists *some* value of $\bar{\omega} \in S_+^2$ that would allow us to take the σ_a term out of the integral. Once we do that, the integral in Equation 79b is simply the total downward irradiance, which we denote $E_\downarrow(x)$. If we assume that the material absorption is isotropic, then σ_a is independent of $\bar{\omega}$ and depends only on x ; we can then further let $K_\downarrow(x) = \sigma_a(x, \bar{\omega})$, which gives us Equation 79c.

$K_\downarrow(x)$ is a phenomenological coefficient, but it is related to the fundamental absorption coefficient σ_a and, thus, carries physical meanings. This is clear from Equation 79d: $K_\downarrow(x)$, by definition, represents the fraction of downward irradiance that is absorbed per unit length. Therefore, $K_\downarrow(x)$ can be intuitively interpreted as the effective *downward absorption coefficient* at depth x . Note that while the mean-value theorem tells us that *some* $\bar{\omega} \in S_+^2$ exists, it does not tell us what its values is. In reality, $K_\downarrow(x)$ will very much depend on $L(x, \omega)$.

Similarly, we can re-write Equation 78b as:

$$\int_{S_+^2}^{S_+^2} \int_{S_+^2}^{S_+^2} \sigma_s(x, \omega) f_p(x, \omega', \omega) \Delta x L(x, \omega) d\omega' d\omega \quad (80a)$$

$$= \int_{S_+^2}^{S_+^2} \left(\int_{S_+^2}^{S_+^2} \sigma_s(x, \omega) f_p(x, \omega', \omega) d\omega' \right) \Delta x L(x, \omega) d\omega \quad (80b)$$

$$= \int_{S_+^2}^{S_+^2} \sigma_s(x, \bar{\omega}) f_p(x, \omega', \bar{\omega}) d\omega' \Delta x \int_{S_+^2}^{S_+^2} L(x, \omega) d\omega \quad (80c)$$

$$= S_{\downarrow\uparrow}(x) \Delta x E_\downarrow(x), \quad (80d)$$

$$S_{\downarrow\uparrow}(x) = \frac{\int_{S_+^2}^{S_+^2} \int_{S_+^2}^{S_+^2} \sigma_s(x, \omega) f_p(x, \omega', \omega) L(x, \omega) d\omega' d\omega}{E_\downarrow(x)}. \quad (80e)$$

Equation 80b simply rearranges the terms in Equation 80a. We then invoke the mean-value theorem again, which says that there exists $\bar{\omega} \in S_+^2$ that allows us to turn Equation 80b to Equation 80c. We use $S_{\downarrow\uparrow}(x)$ to denote the integral in Equation 80c: $S_{\downarrow\uparrow}(x) = \int_{S_+^2}^{S_+^2} \sigma_s(x, \bar{\omega}) f_p(x, \omega', \bar{\omega}) d\omega'$, which gets us Equation 80d.

$S_{\downarrow\uparrow}(x)$ is again a phenomenological coefficient that is related to the fundamental scattering coefficient. Its physical interpretation is clear from Equation 80e: it is the fraction of the downward irradiance that is scattered upward per unit length. We can think of it as the effective “upward scattering coefficient of the downward irradiance.”

Finally, we can re-write Equation 78c in a similar way:

²⁷which says that there exists $c \in [a, b]$ such that $\int_a^b f(x)g(x)dx = f(c) \int_a^b g(x)dx$, if $g(x)$ is integrable and does not change its sign in $[a, b]$.

$$\int_{S_-^2}^{S_-^2} \int_{S_-^2}^{S_-^2} \sigma_s(x + \Delta x, \omega) f_p(x + \Delta x, \omega', \omega) \Delta x L(x + \Delta x, \omega) d\omega' d\omega \quad (81a)$$

$$= \int_{S_-^2}^{S_-^2} \sigma_s(x + \Delta x, \hat{\omega}) f_p(x + \Delta x, \omega', \hat{\omega}) d\omega' \Delta x \int_{S_-^2}^{S_-^2} L(x + \Delta x, \omega) d\omega \quad (81b)$$

$$= S_{\uparrow\downarrow}(x + \Delta x) \Delta x E_{\uparrow}(x + \Delta x), \quad (81c)$$

$$S_{\uparrow\downarrow}(x + \Delta x) = \frac{\int_{S_-^2}^{S_-^2} \int_{S_-^2}^{S_-^2} \sigma_s(x + \Delta x, \omega) f_p(x + \Delta x, \omega', \omega) \Delta x L(x + \Delta x, \omega) d\omega' d\omega}{E_{\uparrow}(x + \Delta x)}, \quad (81d)$$

$$S_{\uparrow\downarrow}(x) = \frac{\int_{S_-^2}^{S_-^2} \int_{S_-^2}^{S_-^2} \sigma_s(x, \omega) f_p(x, \omega', \omega) \Delta x L(x, \omega) d\omega' d\omega}{E_{\uparrow}(x)}. \quad (81e)$$

The mean-value theorem says that some $\hat{\omega} \in S_-^2$ exists that allows us to turn Equation 81a to Equation 81b. We use $S_{\uparrow\downarrow}(x)$ to denote the first integral in Equation 81b. The second integral in Equation 81b is essentially the total upward irradiance at the depth $x + \Delta x$, which we denote $E_{\uparrow}(x + \Delta x)$.

Now plug Equation 79c, Equation 80d, and Equation 81c back to Equation 83b, we get:

$$E_{\downarrow}(x + \Delta x) = E_{\downarrow}(x) - K_{\downarrow}(x) \Delta x E_{\downarrow}(x) - S_{\uparrow\downarrow}(x) \Delta x E_{\downarrow}(x) + S_{\uparrow\downarrow}(x + \Delta x) \Delta x E_{\uparrow}(x + \Delta x). \quad (82)$$

Rewrite it and take the limit as $\Delta x \rightarrow 0$:

$$\frac{E_{\downarrow}(x + \Delta x) - E_{\downarrow}(x)}{\Delta x} = -K_{\downarrow}(x) E_{\downarrow}(x) - S_{\uparrow\downarrow}(x) E_{\downarrow}(x) + S_{\uparrow\downarrow}(x + \Delta x) E_{\uparrow}(x + \Delta x), \quad (83a)$$

$$\frac{dE_{\downarrow}(x)}{dx} = -K_{\downarrow}(x) E_{\downarrow}(x) - S_{\uparrow\downarrow}(x) E_{\downarrow}(x) + S_{\uparrow\downarrow}(x) E_{\uparrow}(x). \quad (83b)$$

Similarly we can express the rate of change of the upward irradiance $E_{\uparrow}(x)$ based on the same energy conservation dynamics:

$$\frac{dE_{\uparrow}(x)}{dx} = K_{\uparrow}(x) E_{\uparrow}(x) + S_{\uparrow\downarrow}(x) E_{\uparrow}(x) - S_{\uparrow\downarrow}(x) E_{\downarrow}(x), \quad (84)$$

where the three terms on the right-hand size, again, represent the absorption, out-scattering, and in-scattering in the original RTE.

Now a few more assumptions. If we assume that the material absorption is isotropic (the total absorption per unit length is invariant to incident light direction over the entire sphere), we have $K_{\uparrow}(x) = K_{\downarrow}(x)$ because:

$$\sigma_a(x, \hat{\omega}) = \sigma_a(x, \bar{\omega}), \quad (85)$$

for some $\hat{\omega} \in S_-^2$ and $\bar{\omega} \in S_+^2$.

If we further assume that 1) the material is an isotropic scattering medium, then $\sigma_s(x, \omega)$ is invariant to ω (the total amount of scattering per unit length does not change with the incident light direction over the entire sphere), and 2) the particles are also isotropic scatters (which theoretically do not exist), then $f_p(x, \omega', \omega)$ is a constant $\frac{1}{4\pi}$. Therefore, $S_{\uparrow\downarrow}(x) = S_{\downarrow\uparrow}(x)$, because:

$$\int^{S_-^2} \sigma_s(x, \hat{\omega}) f_p(x, \omega', \hat{\omega}) d\omega' = \int^{S_+^2} \sigma_s(x, \bar{\omega}) f_p(x, \omega', \bar{\omega}) d\omega', \quad (86)$$

for some $\hat{\omega} \in S_-^2$ and $\bar{\omega} \in S_+^2$.

A few notes on the assumptions here:

- The assumption of isotropic scatters is important; assuming only an isotropic medium is not enough. This is because the integral in Equation 86 integrates only the upper (or lower) hemisphere rather than the entire sphere, so if the phase function itself is not a constant, the integral result will still depend on $\bar{\omega}$ or $\hat{\omega}$. See the distinction between an isotropic medium and an isotropic scatter on p. 51.
- Alternatively, if the particles are not isotropic scatters, $S_{\uparrow\downarrow}(x) = S_{\downarrow\uparrow}(x)$ can still hold if we assume that the upward and downward irradiance are both diffuse, i.e., the radiance $L(x, \omega)$ is invariant to ω . The isotropic medium assumption still needs to hold. This can be proven by going back to the respective definition of $S_{\uparrow\downarrow}(x)$ and $S_{\downarrow\uparrow}(x)$ in Equation 80e and Equation 81e²⁸. Interestingly, if the particles are isotropic scatters, the outgoing irradiance will necessarily be diffuse.
- Yet another way for $S_{\uparrow\downarrow}(x) = S_{\downarrow\uparrow}(x)$ is if we are considering a very idealized scenario where photons travel and are scattered only upward or downward [Bohren and Clothiaux, 2006, Chpt. 5.2]. If the medium is also isotropic (but does not have to consist of isotropic scatters), the fraction of the upward irradiance turned downward would be the same as the fraction of the downward irradiance turned upward, so $S_{\uparrow\downarrow}(x) = S_{\downarrow\uparrow}(x)$.

Finally, given the assumption that the material is spatially homogeneous, both σ_a , σ_s , and f_p are all independent of x . Therefore, we can denote $K = K_\uparrow(x) = K_\downarrow(x)$ and $S = S_{\uparrow\downarrow}(x) = S_{\downarrow\uparrow}(x)$.

Now we combine Equation 83b and Equation 84 and get to the famous pair of differential equations underlying the K-M model:

$$\frac{dE_\downarrow(x)}{dx} = -(K + S)E_\downarrow(x) + SE_\uparrow(x), \quad (87a)$$

$$\frac{dE_\uparrow(x)}{dx} = (K + S)E_\uparrow(x) - SE_\downarrow(x). \quad (87b)$$

²⁸If $L(x, \omega)$ is invariant to ω and $f_p(x, \omega', \hat{\omega})$, under the isotropic medium assumption, is reduced to $f_p(x, \theta)$, where θ is the angle subtended by ω and ω' , both $S_{\uparrow\downarrow}(x)$ and $S_{\downarrow\uparrow}(x)$ are essentially calculating $\int^\pi \int^\pi c d\theta d\theta'$.

10.2 The Model and Its Interpretation

Equation 87 gives us a pair of linear differential equations. What we are interested in solving for is $R(0) = \frac{E_{\uparrow}(0)}{E_{\downarrow}(0)}$. The boundary condition is that $R(X) = R_g$, which is the (assumed-to-be) known reflectance of the substrate. Solving the differential equations gives us:

$$R(0) = \frac{E_{\uparrow}(0)}{E_{\downarrow}(0)} = \frac{1 - R_g[a - b \coth(bSX)]}{a - R_g + b \coth(bSX)}, \quad (88a)$$

$$T(X) = \frac{E_{\downarrow}(X)}{E_{\downarrow}(0)} = \frac{b}{a \sinh(bSX) + b \cosh(bSX)}, \quad (88b)$$

$$a = \frac{K + S}{S}, \quad (88c)$$

$$b = \sqrt{a^2 - 1}, \quad (88d)$$

where $R(0)$ is the hemispherical-hemispherical reflectance at the material surface and $T(X)$ is the hemispherical-hemispherical transmittance at the bottom of the material²⁹; $\coth(\cdot)$ is the hyperbolic cotangent function, $\sinh(\cdot)$ is the hyperbolic sine function, and $\cosh(\cdot)$ is the hyperbolic cosine function. Note that we omit λ for simplicity but keep in mind that $R(0), T(X), S, K, R_g, a, b$ are all spectral quantities.

Equation 88 is the famous K-M model. Consider the case where the material is purely absorptive and scatters little (e.g., dyes on textiles or fabrics), so S is close to 0 and the reflectance is simplified to:

$$R(0) = R_g e^{-2KX} = e^{-KX} R_g e^{-KX}. \quad (89)$$

We can understand R_X by decomposing it into the product of three terms, each representing a step in the overall, observed reflection. First, photons go through the material from the top down, being absorbed as they go. The percentage of photons that are still left (i.e., unabsorbed) just before they hit the substrate is e^{-KX} , which is consistent with the Beer-Lambert law. Second, the substrates reflects R_g amount of lights back toward the material. Finally, as the photons make their way back to the surface they go through another round of absorption governed by the same Beer-Lambert law (Chapter 7.1).

Now consider the case where there is no substrate or when the substrate is a perfect black substrate; in both cases $R_g = 0$. Reflectance is now:

$$R_{black} = \frac{1}{a + b \coth(bSX)}. \quad (90)$$

²⁹You can see that $T(X)$ does not involve R_g : when we calculate the transmittance of the material we assume that there is no substrate.

Finally, consider the case where the material is so thick that no photon reaches the substrate. In this case, the reflectance is not affected by the substrate and can be simplified to (by letting X approach infinity):

$$\lim_{X \rightarrow \infty} = R_\infty = \frac{1}{a+b} = 1 + \frac{K}{S} - \sqrt{\left(\frac{K}{S}\right)^2 + 2\frac{K}{S}}. \quad (91)$$

In the painting industry, we say the paint's "hiding is complete" when the substrate does not influence the material color. We can quantify the "hiding power" of a paint by $H = \frac{R_{white}}{R_{black}}$, i.e., the ratio of reflectance between when the substrate is black (absorbs everything) and white (reflects everything back). If the hiding is complete, H would be 1.

You might be wondering how we know K and S of a material — we measure them. For instance, observe Equation 90 and Equation 91; we can measure the reflectance R_{black} when the substrate is nearly black and the reflectance R_∞ when the hiding is near complete. We can then solve the system of equations to estimate K and S .

You can see that we are not actually taking a very thin layer of particles, illuminating it, and then measuring how much of the light is scattered vs. absorbed. Instead, we estimate K and S macroscopically. We have in mind a model (a set of equations or functions if you will) that are parameterized by unknown variables. We then probe the model by giving it different inputs and measure the outputs. From the input-output pairs we can then estimate the unknown parameters. This is why a model so defined and derived is called a phenomenological model and the parameters (e.g., the K and S coefficients) are called the phenomenological parameters — because all we do is to observe the phenomena.

10.3 N-Stream Model

The basic K-M model is called the two-stream or two-flux model, because it considers only the total upward irradiance and total downward irradiance. We basically have divided all the directions possible in the space into only two solid angles, the upper hemisphere and the lower hemisphere. What if we want to know the irradiance at a finer granularity (i.e., over a smaller solid angle)? We divide all the direction into more solid angles, and analyze the irradiance change in them using a similar method.

For instance, if we now consider irradiance in four directions: $E_{\nwarrow}(x)$, $E_{\nearrow}(x)$, $E_{\searrow}(x)$, and $E_{\swarrow}(x)$, we can write the change of the irradiance in $E_{\searrow}(x)$ as:

$$\begin{aligned} \frac{dE_{\searrow}(x)}{dx} = & -K_{\searrow}(x)E_{\searrow}(x) \\ & + S_{\nwarrow\searrow}(x)E_{\nwarrow}(x) + S_{\nearrow\searrow}(x)E_{\nearrow}(x) + S_{\swarrow\searrow}(x)E_{\swarrow}(x) \\ & - S_{\searrow\nwarrow}(x)E_{\nwarrow}(x) - S_{\searrow\nearrow}(x)E_{\nearrow}(x) - S_{\searrow\swarrow}(x)E_{\swarrow}(x), \end{aligned} \quad (92)$$

where $S_{\nwarrow\searrow}$ is the scattering coefficient of from the northwest irradiance to the southeast direction, and so on. We can express the changes of the other three directions similarly. In general, if

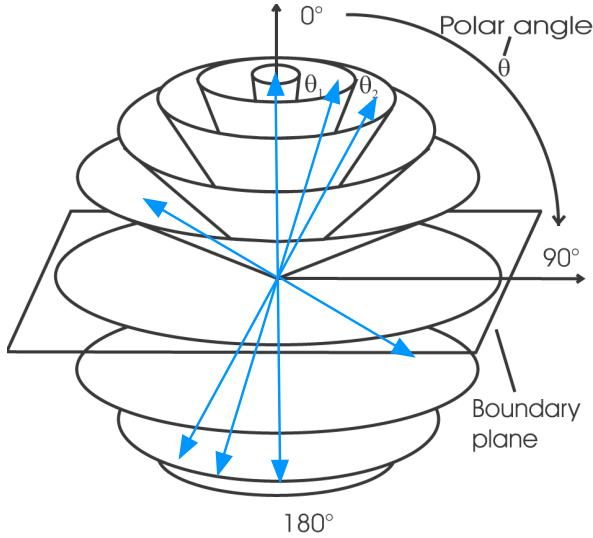


Figure 20: The setup for the N-flux model. All the possible directions (consider all the arrows that can possibly go out from the origin) are divided into “channels” or “streams”, each of which represents a finite solid angle within which an irradiance travel. The N-flux model models the changes of each of these irradiances. Adapted from [Li \[2003, Fig. 4.1\]](#).

we divide the space into N “channels”, each representing a set of direction (a finite solid angle), we can extend the two-flux model to a “N-flux” model.

Figure 20 shows the setup for deriving the N-flux model, where all the possible directions (consider all the arrows that can possibly go out from the origin) are divided into “channels” or “streams”, each of which represents a finite solid angle within which an irradiance travel. The figure visualizes eight such channels. The change of each channel is modeled by taking away from each channel photons that are absorbed and scattered to all other channels and by adding photons scattered in to the channel from all other channels. In the end, we get N linear differential equations. Equation 92 is one such equation when $N = 4$. The channels are usually rotationally symmetric about the z -axis, assuming that the material is rotationally symmetric about the z -axis.

Comparing against the RTE in Equation 65, which has an integral term L_s given in Equation 62. What the N-flux model does is to essentially approximate that integral with a finite sum. In general, the larger the N the better the approximation but also the more computationally intensive to solve. We will omit a formal treatment but refer you to [Bohren and Clothiaux \[2006, Chpt. 6.1\]](#), [Volz and Simon \[2001, Chpt. 3.1.2\]](#), and [Klein \[2010, Chpt. 5.5\]](#) for details.

10.4 Correction for Surface Reflection

One thing that is ignored in the K-M model is the surface reflection/refraction yet. Part of the photons will be reflected away at the air/material interface before they enter the material.

Similarly, when we consider the transmittance we have ignored the reflection/refraction at the other side of the material. So the reflectance and transmittance calculated by the K-M model are defined at the point when the photons are just about to leave the material.

To account for the surface phenomena, we can apply what is called the Saunderson correction, derived by [Saunderson \[1942\]](#). See [Sharma \[2003\]](#), Chpt. 3.6.3 for a derivation, but briefly if the illumination is diffuse, the corrected surface reflectance is:

$$R = r_s + \frac{(1 - r_s)(1 - r_i) R(0)}{1 - r_i R(0)}, \quad (93)$$

where $R(0)$ is the reflectance given by the K-M model, r_s is the fraction of incident irradiance scattered by the air-material surface and r_i is the fraction of the internal irradiance approaching the air-material interface that is scattered back by the interface. Assuming a smooth surface, both r_s and r_i can be calculated by the Fresnel equations given the refractive index of the material (Chapter 4.4).

References

- ASTM International. ASTM E284-13b Standard Terminology of Appearance. <https://webstore.ansi.org/standards/astm/astme28413b>.
- Anil Bharath. *Introductory medical imaging*. Morgan & Claypool, 2009.
- James F Blinn. Light reflection functions for simulation of clouds and dusty surfaces. *AcM SIGGRAPH Computer Graphics*, 16(3):21–29, 1982.
- David A Boas, Dana H Brooks, Eric L Miller, Charles A DiMarzio, Misha Kilmer, Richard J Gaudette, and Quan Zhang. Imaging the body with diffuse optical tomography. *IEEE signal processing magazine*, 18(6):57–75, 2001.
- Craig F Bohren and Eugene E Clothiaux. *Fundamentals of atmospheric radiation: an introduction with 400 problems*. John Wiley & Sons, 2006.
- James K Bowmaker and HJk Dartnall. Visual pigments of rods and cones in a human retina. *The Journal of physiology*, 298(1):501–511, 1980.
- Bruce MacEvoy. The material attributes of paints. <https://www.handprint.com/HP/WCL/pigmt3.html#particlesize>, 2015.
- Subrahmanyam Chandrasekhar. *Radiative transfer*. Courier Corporation, 1960.
- Robert L Cook and Kenneth E. Torrance. A reflectance model for computer graphics. *ACM Transactions on Graphics (ToG)*, 1(1):7–24, 1982.
- Yue Dong, Stephen Lin, Baining Guo, et al. *Material appearance modeling: A data-coherent approach*. Springer, 2013.

- Julie Dorsey, Holly Rushmeier, and François Sillion. *Digital modeling of material appearance*. Elsevier, 2010.
- G Eason, AR Veitch, RM Nisbet, and FW Turnbull. The theory of the back-scattering of light by blood. *Journal of Physics D: Applied Physics*, 11(10):1463, 1978.
- Klaus Engel, Markus Hadwiger, Joe M Kniss, Christof Rezk-Salama, and Daniel Weiskopf. *Real-time volume graphics*. A K Peters, Ltd., 2006.
- Thomas J Farrell, Michael S Patterson, and Brian Wilson. A diffusion theory model of spatially resolved, steady-state diffuse reflectance for the noninvasive determination of tissue optical properties in vivo. *Medical physics*, 19(4):879–888, 1992.
- R Feynman. *QED: The Strange Theory of Light and Matter by Richard Feynman*. Princeton University Press, 1985.
- Julian Fong, Magnus Wrenninge, Christopher Kulla, and Ralf Habel. Production volume rendering: Siggraph 2017 course. In *ACM SIGGRAPH 2017 Courses*, pages 1–97. 2017.
- Jeppe Revall Frisvad, Soeren A Jensen, Jonas Skovlund Madsen, António Correia, Li Yang, Søren Kimmer Schou Gregersen, Youri Meuret, and P-E Hansen. Survey of models for acquiring the optical properties of translucent materials. In *Computer graphics forum*, volume 39, pages 729–755. Wiley Online Library, 2020.
- Andrew S Glassner. *Principles of digital image synthesis*. Elsevier, 1995.
- Cindy M Goral, Kenneth E Torrance, Donald P Greenberg, and Bennett Battaile. Modeling the interaction of light between diffuse surfaces. *ACM SIGGRAPH computer graphics*, 18(3):213–222, 1984.
- Markus Gross and Hanspeter Pfister. *Point-based graphics*. Elsevier, 2011.
- Akira Ishimaru. Theory and application of wave propagation and scattering in random media. *Proceedings of the IEEE*, 65(7):1030–1061, 1977.
- Akira Ishimaru et al. *Wave propagation and scattering in random media*, volume 2. Academic press New York, 1978.
- Henrik Wann Jensen, Stephen R Marschner, Marc Levoy, and Pat Hanrahan. A practical model for subsurface light transport. *ACM SIGGRAPH computer graphics*, page 511–518, 2001.
- Sönke Johnsen. *The optics of life: a biologist’s guide to light in nature*. Princeton University Press, 2012.
- Ruth Johnston-Feller. *Color science in the examination of museum objects: nondestructive procedures*. Getty Publications, 2001.

- Deane B Judd and Günter Wyszecki. *Color in business, science, and industry*. John Wiley & Sons, 3 edition, 1975.
- James T Kajiya and Brian P Von Herzen. Ray tracing volume densities. *ACM SIGGRAPH computer graphics*, 18(3):165–174, 1984.
- Arie Kaufman and Klaus Mueller. Volume visualization and volume graphics. *Technical Report; Stony Brook University*, 2003.
- Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023.
- Kipp Jones. The sand/dust from the harmattan coming from the Sahara gives the Nigeria’s National Mosque in Abuja, a nice glow; CC BY-SA 2.0 license. <https://commons.wikimedia.org/wiki/File:MosqueinAbuja.jpg>, 2005.
- GA Klein. Industrial color physics, 2010.
- Craig Kolb, Don Mitchell, and Pat Hanrahan. A realistic camera model for computer graphics. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 317–324, 1995.
- Paul Kubelka. New contributions to the optics of intensely light-scattering materials. part i. *Josa*, 38(5):448–457, 1948.
- Paul Kubelka and Franz Munk. An article on optics of paint layers (translated by stephen h. westin). *Z. Tech. Phys*, 12(593-601):259–274, 1931a.
- Paul Kubelka and Franz Munk. Ein Beitrag zur Optik der Farbanstriche. *Z. tech. Phys*, 12: 593–601, 1931b.
- Jean-Henri Lambert. *Photometria sive de mensura et gradibus luminis, colorum et umbrae*. Sumptibus viduae Eberhardi Klett, typis Christophori Petri Detleffsen, 1760.
- Dmitri Lanevski, Farshid Manoocheri, and Erkki Ikonen. Gonioreflectometer for measuring 3d spectral brdf of horizontally aligned samples with traceability to si. *Metrologia*, 59(2):025006, 2022.
- Marc Levoy. Display of surfaces from volume data. *IEEE Computer graphics and Applications*, 8(3):29–37, 1988.
- Marc Levoy and Turner Whitted. The use of points as a display primitive. *Technical Report; University of North Carolina at Chapel Hill*, 1985.
- Yang Li. Ink-paper interaction-a study in ink-jet color reproduction. *Institute of Technology-Linköpings University. Norrköping, Sweden: UniTryck*, 2003.

- Nelson Max. Optical models for direct volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 1(2):99–108, 1995.
- Thomas G Mayerhöfer, Susanne Pahlow, and Jürgen Popp. The bouguer-beer-lambert law: Shining light on the obscure. *ChemPhysChem*, 21(18):2029–2046, 2020.
- William G Melbourne. *Radio occultations using Earth satellites: a wave theory treatment*. John Wiley & Sons, 1 edition, 2004.
- Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021.
- Ron Milo and Rob Phillips. *Cell biology by the numbers*. Garland Science, 2015.
- FE Nicodemus, JC Richmond, JJ Hsia, IW Ginsberg, and T Limperis. *Geometrical considerations and nomenclature for reflectance*, volume 160. US Department of Commerce, National Bureau of Standards Washington, DC, USA, 1977.
- Jan Novák, Iliyan Georgiev, Johannes Hanika, Jaroslav Křivánek, and Wojciech Jarosz. Monte Carlo methods for physically based volume rendering. In *ACM SIGGRAPH 2018 Courses*, 2018.
- Michael Oren and Shree K Nayar. Generalization of the lambertian model and implications for machine vision. *International Journal of Computer Vision*, 14:227–251, 1995.
- Matt Pharr, Wenzel Jakob, and Greg Humphreys. *Physically based rendering: From theory to implementation*. MIT Press, 3 edition, 2018.
- Matt Pharr, Wenzel Jakob, and Greg Humphreys. *Physically based rendering: From theory to implementation*. MIT Press, 4 edition, 2023.
- Prabhu B Doss. Tso Kiagar Lake Ladakh; CC BY 2.0 license. https://commons.wikimedia.org/wiki/File:Tso_Kiagar_Lake_Ladakh.jpg, 2007.
- Ana M Rabal, Alejandro Ferrero, Joaquin Campos, José Luis Fontecha, Alicia Pons, Antonio Manuel Rubiño, and Antonio Corróns. Automatic gonio-spectrophotometer for the absolute measurement of the spectral brdf at in-and out-of-plane and retroreflection geometries. *Metrologia*, 49(3):213, 2012.
- Erik Reinhard, Erum Arif Khan, Ahmet Oguz Akyuz, and Garrett Johnson. *Color imaging: fundamentals and applications*. CRC Press, 2008.
- Paolo Sabella. A rendering algorithm for visualizing 3d scalar fields. *ACM SIGGRAPH computer graphics*, 22(4):51–58, 1988.
- JL Saunderson. Calculation of the color of pigmented plastics. *JOSA*, 32(12):727–736, 1942.

- Martin Schweiger, SR Arridge, M Hiraoka, and DT Delpy. The finite element method for the propagation of light in scattering media: boundary and source conditions. *Medical physics*, 22(11):1779–1792, 1995.
- Gaurav Sharma. Color fundamentals for digital imaging. In Gaurav Sharma, editor, *Digital color imaging handbook*, pages 14–127. CRC Press, 2003.
- Jos Stam. Multiple scattering as a diffusion process. In *Rendering Techniques' 95: Proceedings of the Eurographics Workshop in Dublin, Ireland, June 12–14, 1995* 6, pages 41–50. Springer, 1995.
- Steve Fareham. Heart of the City water feature Sheffield; CC BY-SA 2.0 license. https://commons.wikimedia.org/wiki/File:Heart_of_the_City_water_feature_Sheffield_-_geograph.org.uk_-_618552.jpg, 2007.
- Kenneth E Torrance and Ephraim M Sparrow. Theory for off-specular reflection from roughened surfaces. *Josa*, 57(9):1105–1114, 1967.
- H Joel Trussell and Michael J Vrhel. *Fundamentals of digital imaging*. Cambridge University Press, 2008.
- User:Daderot. Kongens Have, Copenhagen, Denmark; CC0 1.0 license. https://commons.wikimedia.org/wiki/File:Marble_ball_-_Kongens_Have_-_Copenhagen_-_DSC07898.JPG, 2012.
- Hans G Volz and Frederick T Simon. *Industrial color testing*, volume 2. Wiley-VCH New York, 2001.
- VonHaarberg. Illustration of a diffuse BRDF; CC0 1.0 license. https://commons.wikimedia.org/wiki/File:BRDF_diffuse.svg, 2018a.
- VonHaarberg. Illustration of a glossy BRDF; CC0 1.0 license. https://commons.wikimedia.org/wiki/File:BRDF_glossy.svg, 2018b.
- VonHaarberg. Illustration of a mirror BRDF; CC0 1.0 license. https://commons.wikimedia.org/wiki/File:BRDF_mirror.svg, 2018c.
- Bruce Walter, Stephen R Marschner, Hongsong Li, and Kenneth E Torrance. Microfacet models for refraction through rough surfaces. *Rendering techniques*, 2007:18th, 2007.
- Brian A. Wandell. *Foundations of vision*. Sinauer Associates, 1995.
- Gregory J Ward. Measuring and modeling anisotropic reflection. In *Proceedings of the 19th annual conference on Computer graphics and interactive techniques*, pages 265–272, 1992.
- Lee Westover. Footprint evaluation for volume rendering. In *Proceedings of the 17th annual conference on Computer graphics and interactive techniques*, pages 367–376, 1990.

Peter L Williams and Nelson Max. A volume density optical model. In *Proceedings of the 1992 workshop on Volume visualization*, pages 61–68, 1992.

Magnus Wrenninge and Nafees Bin Zafar. Production volume rendering: Siggraph 2011 course. In *ACM SIGGRAPH 2011 Courses*, pages 1–71. 2011.