

THE PBR GUIDE - PART 1 ⌚ 37:00

[Foundation](#)[#PBR](#)[#PBR Guide](#)

THE PBR GUIDE BY ALLEGORITHMIC - PART 1

Light and Matter : The theory of Physically-Based Rendering and Shading

[Tutorials](#) ▾[the book \(Part 1 + Part 2\)](#)[➤ Download Project Files](#)

PBR Guide 2018 version.pdf

46.47MB



PBR GUIDE
VOLUME ONE

Read the second part of the PBR Guide [here](#)

TABLE OF CONTENTS

- Light Rays
- Absorption and Scattering - Transparency and Translucency
- Diffuse and Specular Reflection - Microfacet Theory
- Color
- BRDF
- Energy Conservation
- Fresnel Effect - F_0 (Fresnel Reflectance at 0 Degrees)
- Conductors and Insulators - Metals and Non Metal
- Linear Space Rendering
- Key characteristics of PBR

THE THEORY OF PHYSICALLY-BASED RENDERING AND SHADING

Light is a complex phenomenon as it can exhibit properties of both a wave and a particle. As a result, different models have been created to describe its behavior.

As texture artists, we are interested in the ray model of light as it describes the interaction of light and matter. Understanding how light rays interact with surface matter is important because our job is to create textures that describe a surface. The textures and materials we author interact with light in our virtual worlds. The more we understand how light behaves, the better our textures will look.

In this guide, we will discuss the theory behind the physics of physically based rendering (PBR) models. We will start by examining the behavior of light rays and work up to defining the key characteristics of PBR.

Light Rays

The ray model of light states that a light ray has the trajectory of a straight line in a homogeneous transparent medium such as air. The ray model also says that the ray will behave in a predictable manner when encountering surfaces such as opaque objects, or when passing through a different medium such as from air to water.

This makes it possible to visualize the path of the light ray as it moves from its starting point to another point where it changes into another form of energy, such as heat.

A light ray that hits a surface is called the incident ray and the angle at which it hits is called the angle of incidence (Figure 01).

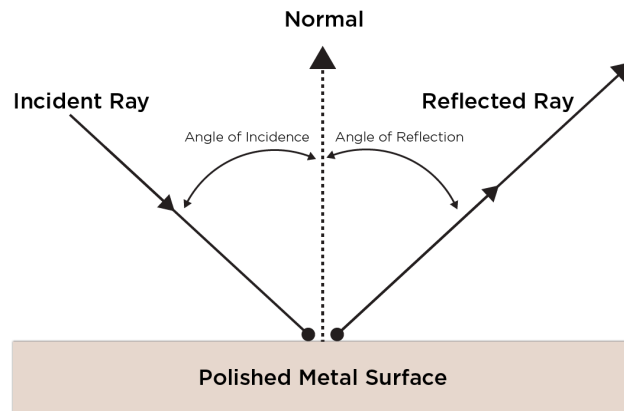


Figure 01: Angle of incidence, incident and reflected rays

A light ray is incident on a plane interface between two media. When a light ray hits a surface, one or both of the following events may occur:

1. The light ray is reflected off the surface and travels in a different direction. It follows the Law of Reflection, which states that the angle of reflection is equal to the angle of incidence (reflected light).
2. The light ray passes from one medium to another in the trajectory of a straight line (refracted light).

At this point, light rays split into two directions: reflection and refraction. At the surface, the ray is either reflected or refracted, and it can be eventually absorbed by either medium. However, absorption does not occur at the surface of the material.

Absorption and Scattering (Transparency and Translucency)

When traveling in an inhomogeneous medium or translucent material, light can be absorbed or scattered:

When light is absorbed, the light intensity decreases as it changes into another form of energy – usually heat. Its color changes as the amount of light absorbed depends on the wavelength, but the direction of the ray doesn't change.

When light is scattered, the ray direction changes randomly, and the amount of deviation depends on the material. Scattering randomizes light direction, but doesn't change its intensity. An ear is a good example of this phenomenon. The ear is thin (absorption is low), so you can see the scattered light radiating from the back of the ear, as shown in Figure 02.

If there is no scattering and the absorption is low, rays can pass directly through the surface. This is the case with glass. For example, imagine you are swimming in a clean pool. You can open your eyes and see a great distance through the clear water. However, if that same pool is relatively dirty, the dirt particles will scatter the light and lower the clarity of the water, and the distance you can see will be reduced as a result.

The further light travels in such a medium/material, the more it is absorbed and/or scattered. Therefore, object thickness plays a large role in how much the light is absorbed or scattered. A thickness map can be used to describe object thickness to the shader as shown in Figure 03.



Figure 03: Thickness map used with subsurface scattering within Substance Painter



Figure 02: Scattered light radiating from back of the ear

Diffuse and Specular Reflection

Specular reflection refers to light that has been reflected at the surface, as we discussed in the light ray section. The light ray is reflected off the surface and travels in a different direction. It follows the law of reflection, which states that on a perfectly planar surface the angle of reflection is equal to the angle of incidence. However, most surfaces are irregular, and the reflected direction will vary randomly based on the surface roughness. This changes light direction, but the light intensity remains constant.

Rougher surfaces will have highlights that are larger, and that appear dimmer. Smoother surfaces will keep specular reflections focused, and they will appear to look brighter or more intense when viewed from the proper angle. However, the same total amount of light is reflected in both cases (Figure 04)

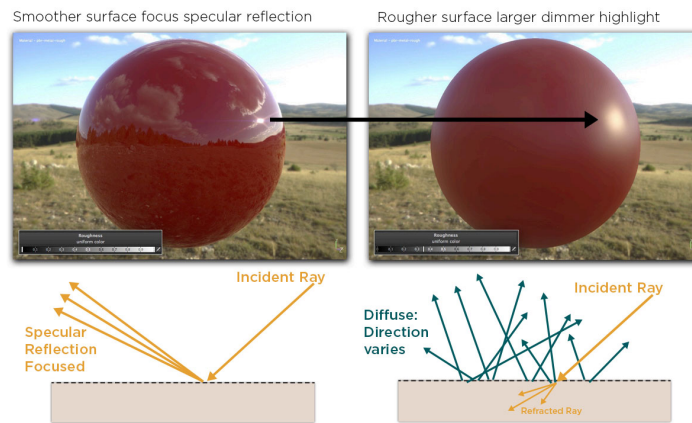


Figure 04: Reflected direction will vary randomly based on the surface roughness

Refraction is a change in a light ray's direction. When light moves from one medium to another it changes speed and direction. The *index of refraction*, or *IOR*, is an optical measurement that describes the change in the direction a light ray is traveling. Essentially, the IOR value is used to determine how much the ray will be bent when it passes through one medium to another. For example, water has an IOR of 1.33, whereas plate glass has an IOR of 1.52. In figure 05 you can see a rendering of a straw placed in a glass of water. The straw appears bent due to refraction as the light travels through different mediums (air, water and glass).

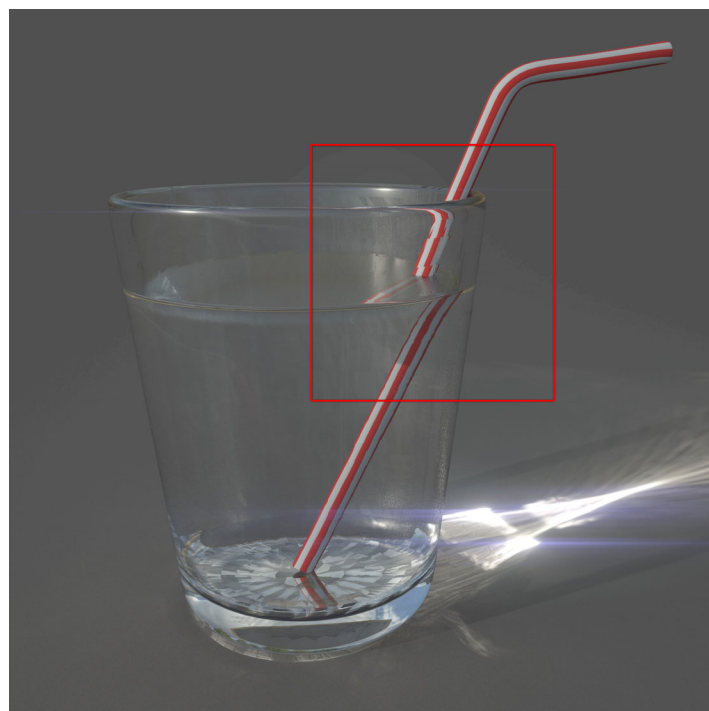


Figure 05: The straw appears bent due to refraction

Diffuse reflection is light that has been refracted. The light ray passes from one medium to another; as our example, we'll assume that it enters an object. The light is then scattered multiple times within this object. It is finally refracted again, out of the object, making its way back to the original medium at approximately the same point where it initially entered (Figure 06).

Diffuse materials are absorbent. If the refracted light travels for too long in such a material, it may be absorbed completely. If the light does exit this material, it has likely traveled only a very small distance from the point of entry.

Therefore, the distance between the entry and exit points can be considered negligible. The Lambertian model, which is used for diffuse reflection in a traditional shading sense, does not take surface roughness into account. Other diffuse reflection models, such as the Oren-Nayar model, do account for this roughness, however.

Materials that have both high scattering and low absorption are sometimes referred to as participating media or translucent materials. Examples of these are smoke, milk, skin, jade and marble. Rendering of the latter three may be possible with the additional modeling of subsurface scattering where the difference between the ingoing and outgoing point of the light ray is no longer considered negligible. Accurate rendering of media with highly varying and very low scattering and absorption, such as smoke or fog, may require even more expensive methods such as Monte Carlo simulations.



Figure 06: A light ray travelling from one medium to another scattering inside the object

Microfacet Theory

In theory, both diffuse and specular reflection are dependent on the surface irregularities where the light rays intersect with a medium. In practice however, the effect of roughness on diffuse reflection is less visible because of the scattering that occurs inside the material. As a result, the outgoing direction of the ray is fairly independent of surface roughness and

the incident direction. The most common model for diffuse reflection (Lambertian) completely neglects roughness.

In this guide, we refer to these surface irregularities as surface roughness. Surface irregularities can have several other names, including roughness, smoothness, glossiness or micro-surface, depending on the PBR workflow in use. All these terms describe the same aspect of a surface, which is sub-textel geometric detail.

These surface irregularities are authored in the roughness or glossiness map depending on the workflow that is being used. A physically-based BRDF is based on the microfacet theory, which supposes that a surface is composed of small-scaled planar detail surfaces of varying orientation called microfacets. Each of these small planes reflects light in a single direction based on its normal (Figure 07).

Micro-facets whose surface normal is oriented exactly halfway between the light direction and view direction will reflect visible light. However, in cases where the microsurface normal and the half normal are equal, not all microfacets will contribute as some will be blocked by shadowing (light direction) or masking (view direction) as illustrated in Figure 07.

The surface irregularities at a microscopic level cause light diffusion. For example, blurred reflections are caused by scattered light rays. The rays are not reflected in parallel, so we perceive the specular reflection as blurred (Figure 08).

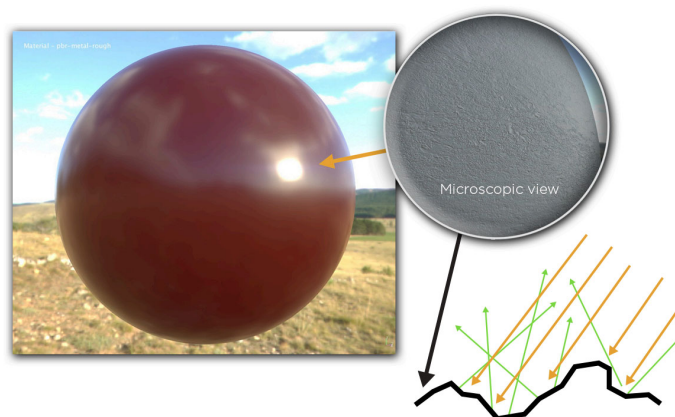


Figure 07: Physically-based BRDF based on microfacet theory

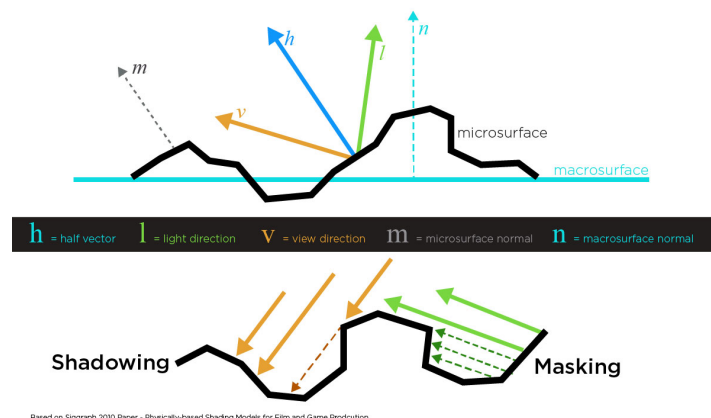


Figure 08: Blurred reflections are due to scattered light rays

Color

The visible color of a surface is due to the wavelengths emitted by the light source. These wavelengths are absorbed by the object and reflected both specularly and diffusely. The remaining reflected wavelengths are what we see as color.

For example, the skin of an apple mostly reflects red light. Only the red wavelengths are scattered back outside the apple skin, while the others are absorbed (Figure 09).

The apple also has bright specular highlights the same color as the light source because with materials that do not conduct electricity (dielectrics) – like the skin of an apple – specular reflection is almost independent of wavelength. For these materials, the specular reflection is never colored. We will discuss the different type of materials (metals and dielectrics) in later sections.

Substance PBR shaders use the GGX microfacet distribution.

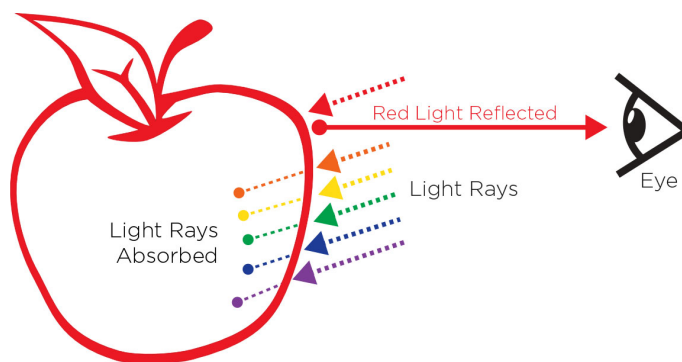


Figure 09: Red wavelengths are reflected to the eye

BRDF

A bidirectional reflectance distribution function (BRDF) is a function that describes the reflectance properties of a surface. In computer graphics, there are different BRDF models – some of which are not physically plausible. For a BRDF to be physically plausible, it must be energy conserving and exhibit reciprocity. Reciprocity refers to the Helmholtz Reciprocity Principle, which states that incoming and outgoing light rays can be considered as reversals of each other without affecting the outcome of the BRDF.

The BRDF used by Substance's PBR shaders are based on Disney's principled reflectance model. This model is based on the GGX microfacet distribution. GGX provides one of the better solutions in terms of specular distribution: with a shorter peak in the highlight and a longer tail in the falloff, it looks more realistic (Figure 10).

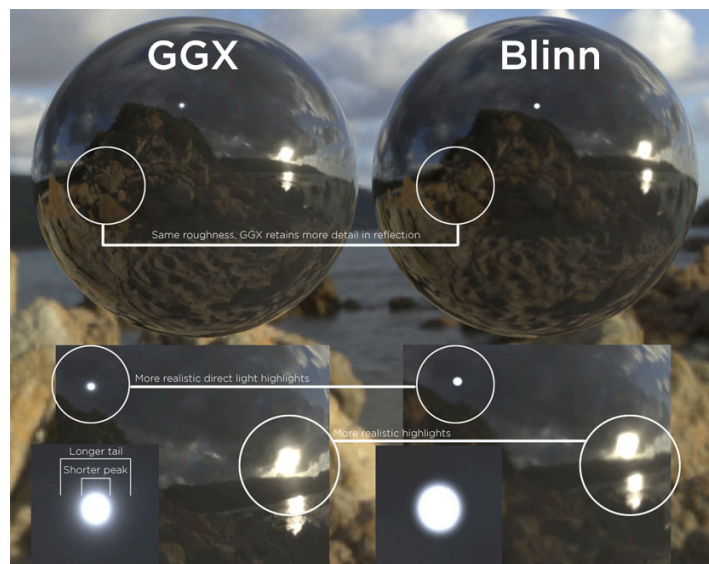


Figure 10: GGX vs Blinn specular distribution – GGX provides one of the better solutions in terms of specular distribution

Energy Conservation

Energy conservation plays a vital role in physically-based rendering solutions. This principle states the total amount of light re-emitted by a surface (reflected and scattered back) is less than the total amount received. In other words, the light reflected from the surface will never be more intense than it was before it hit the surface. As artists, we don't have to worry about controlling energy conservation. This is one of the advantages of PBR: energy conservation is always enforced by the shader. This is part of the physically-based model and it allows us to focus on art rather than physics.

Fresnel Effect

The Fresnel reflection factor also plays a vital role in physically-based shading as a coefficient of the BRDF. The Fresnel Effect, as observed by French physicist Augustin-Jean Fresnel, states that the amount of light reflected from a surface depends on the viewing angle at which it is perceived. Think of a pool of water. If you look straight down, perpendicular to the water surface, you can see down to the bottom. Viewing the water surface in this manner would be at zero degrees or normal incidence, normal being the

surface normal. If you look at the pool of water at a grazing incidence, more parallel to the water surface, you will see that the specular reflections on the water surface become more intense and you may not be able to see below the surface of the water at all.

Fresnel is not something that we control in PBR as we did in traditional shading. Again, this is another aspect of physics that is handled by the PBR shader. When viewing a surface at a grazing incidence, all smoothed surfaces will become reflectors at nearly 100% at a 90-degree angle of incidence.

For rough surfaces, reflectance will become increasingly specular but will not approach 100% specular reflection. The most important factor here is the angle between the normal of each microfacet and the light, rather than the angle between the normal of the “macrosurface” and the light. Because the light rays are dispersed in different directions, the reflection appears softer, or dimmer. What occurs at a macroscopic level is somewhat similar to the average of all the Fresnel Effect you would observe for the collective microfacets.

F0 (Fresnel Reflectance at 0 Degrees)

When light hits a surface straight on or perpendicularly (at a 0-degree angle), a percentage of that light is reflected as specular. Using the index of refraction (IOR) for a surface, you can derive the amount that is reflected. This is referred to as F0 (Fresnel zero) (Figure 11). The amount of light that is refracted into the surface is referred to as $1 - F_0$.



Figure 11: For a smooth dielectric surface, F_0 will reflect between 2-5% of light and 100% at grazing angles

The F_0 range for most common dielectrics will be from 0.02-0.05 (linear values). For conductors, the F_0 range will be 0.5-1.0. The reflectivity of a surface is therefore determined by the refractive index as shown in the equation below (Lagarde 2011).

$$F(0^\circ) = \frac{(n-1)^2}{(n+1)^2} = 0.02$$

It is the F0 reflectance value that we are concerned with in regards to authoring our textures. Non-metals (dielectrics/insulators) will have a greyscale value and metals (conductors) will have an RGB value. With regards to PBR and from an artistic interpretation of reflectance, we can state that for a common smooth dielectric surface, F0 will reflect between 2% and 5% of light and 100% at grazing angles as was shown in Figure 11.

The dielectric (non-metal) reflectance values don't actually change very drastically. In fact, when altered by roughness the actual changes in value can be hard to see. However, there is a difference in the values. In Figure 12, you can see a chart that shows the F0 ranges for both metal and non-metal materials.

Notice that the ranges for non-metals do not deviate from each other drastically. Gemstones are an exception as they have higher values. We will discuss F0 as it specifically relates to conductors and insulators a bit later.

Conductors and Insulators (Metals and Non-Metals)

When creating materials for PBR, it is helpful to think in terms of metal or non-metal. Ask yourself if the surface is metal or not. If it is, you will need to follow one set of guidelines. If it is not, you will need to follow another.

This can be a simplistic approach as some materials may not fall into these categories such as metalloids (a mix of metal and non-metal), but in the overall process of creating materials, distinguishing between metal and non-metal is a good approach and metalloids are an exception. To set up guidelines for materials, we must first understand what we are

trying to create. With PBR, we can look at the properties of metals (conductors) and non-metals (insulators) to derive this set of guidelines as shown in Figure 12.

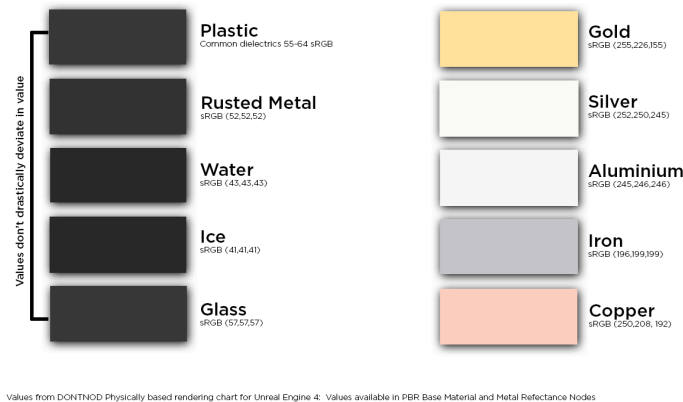
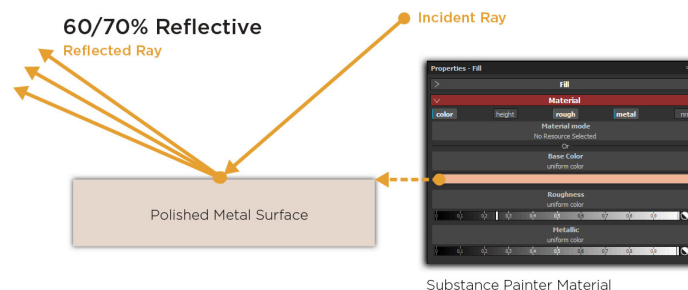


Figure 12: *F0 ranges for both metal and non-metal materials*

Refracted light is absorbed, and the color tint of metals comes from the reflected light, so in our maps, we don't give metals a diffuse color.

Metals

Metals are good conductors of heat and electricity. The electric field in conducting metals is zero, and when an incoming light wave made of electric and magnetic fields hits the surface, the wave is partially reflected, and all the refracted light is absorbed. The reflectance value for polished metal is high at a range of about 70-100% reflective (Figure 13).



*Figure 13: The reflectance value for metal is around
70-100% specular*

Some metals absorb light at different wavelengths. For example, gold absorbs blue light at the high-frequency end of the visible spectrum, so it appears yellow. However, since the refracted light is absorbed, the color tint of metals comes from the reflected light. In our maps, therefore, we don't give metals a diffuse color. For example, in the specular/gloss workflow, raw metal is set to black in the diffuse map and the reflectance value is a tinted color value in the specular map. With metals, the reflectance value will be RGB and can be tinted. Since we are working within a physically-based model, we need to use real-world measured values for the metal reflectance in our maps.

Another important aspect of metals in terms of texturing is their tendency to corrode. This means that weathering elements can play a large role in the reflective state of metal. If the metal rusts, this changes the reflective state of the metal. The corroded areas are then treated as a dielectric material denoted by a black value in the metallic map as shown in Figure 14. As we will discuss in Part 2, the shader in the metallic/roughness workflow hardcodes the F0 value for dielectrics to be 4% reflective. Figure 14 shows the rusted areas in the base color map as diffuse reflected color with a hardcoded F0 value of 4%.

Also, painted metal is treated as a dielectric rather than a metal. The paint acts as a layer on top of the raw metal. Only the raw metal exposed from chipped paint is treated as metal. The same goes for dirt on metal or any matter that obscures the raw metal.

As noted at the beginning of this chapter, it is helpful to ask if a material is a metal or not when creating PBR materials. To be even more precise, the question should also include information about the state of the metal: whether it is painted, rusted or covered in another matter like dirt or grease. The material will be treated as dielectric if it is not raw metal. Depending on weathering, there could be some blending between metal and non-metal as weathering elements play a role in the reflective state of a metal.



*Figure 14: Corrosive areas are treated as a dielectric
with an F0 value of 4% reflective*

Non-Metals

Non-metals (insulators/dielectrics) are poor conductors of electricity. The refracted light is scattered and/or absorbed (often re-emerging from the surface), so they reflect a much smaller amount of light than metals and will have an albedo color.

We stated earlier that the value for common dielectrics is around 2-5% based on the F0 as computed by the index of refraction. These values are contained within the linear range of 0.017-0.067 (40-75 sRGB) as shown in Figure 14. Apart from some non-metal materials such as gemstones, most dielectrics will not have an F0 value greater than 4%.

As with metals, we need to use real-world measured values, but it can be difficult to find an index of refraction (IOR) for other materials that are not transparent. However, the value between most common dielectric materials does not change drastically, so we can use a few guidelines for reflectance values. We will cover them later in this guide.

The value for common dielectrics is around 2-5% based on the F0 as computed by the IOR. You can see this range illustrated in Figure 15.



*Figure 15: Conversions from sRGB to linear were
done using the gamma 2.2 approximation – see the
section on linear space rendering for more details*

Linear Space Rendering

Linear space rendering is a highly complex subject. For this guide, we will take a simplistic approach in stating that linear space rendering provides the correct math for lighting calculations. It creates an environment that allows light interactions to be represented in a credible real-world manner. For a discussion on linear space rendering, we must introduce

the concept of gamma correction. When encoding images for display and storage purposes, gamma correction is the optimization process of reducing bandwidth and bit allocation. This process leverages the human eye's perception of brightness, which approximately follows the cube root of luminance.

The Human Visual System (HVS) is more sensitive to relative differences in darker tones rather than brighter tones. Because of this, not using gamma correction is wasteful as too many bits will be allocated to tonal regions where the HVS cannot distinguish between tones.

In a typical digital image creation process, the image is encoded using an encoding gamma function, such as sRGB OETF or $\gamma 1 / 2.2$ [1], for presentation on a display device. The display device circuitry then decodes the image using its own decoding gamma function, the EOTF[2]; a computer monitor will frequently have a gamma setting of 2.2.

A linear color space intrinsically has no gamma correction. This is equivalent to an effective gamma of 1.0, which yields correct linear computations. However, in order to present the rendered image correctly to the viewer, it needs to be encoded into gamma space[3].

Computations of color values and operations on colors are performed in linear space. The process decodes gamma-encoded values into linear values from our color maps, and from colors chosen while viewing on a monitor via a color picker. In a color-managed workflow, this process typically involves tagging a texture map to be interpreted either as linear or as encoded with sRGB OETF or $\gamma 1 / 2.2$. The computations are then carried out in linear space and the final rendered result is gamma-encoded with sRGB OETF or $\gamma 1 / 2.2$.

A simplistic way of considering which maps need to be decoded is that if the map exported from Substance Painter or Substance Designer represents color that you see (diffuse reflected color) - such as the color tint of a metal or the green color of grass - then it should be flagged as gamma-encoded for the shader to interpret the map correctly. The maps to be flagged as gamma-encoded in a PBR workflow are base color, diffuse, specular and emissive. If the map represents data, such as how rough the surface is (the roughness map), or if the material is metal (the metallic map), then it should be flagged as linear. The maps

to be flagged as linear in a PBR workflow are roughness, ambient occlusion, normal, metallic, and height.

Within Substance Designer and Substance Painter, the conversion between linear and gamma space on inputs to the shader is handled automatically. So is gamma correction on the computed result in the rendered viewport. As artists, we don't usually need to worry about the linear computation and conversions within the Substance software, as the software will handle this by default.

When using Substance materials via the Substance Integration plugin, outputs are flagged for linear/gamma space automatically via the integration and the host application's color management. However, it's important to understand the process: when Substance maps are used as exported bitmaps rather than Substance materials, you may need to manually flag the textures as gamma-encoded or linear depending on the renderer you are using. Usually .png, .jpg, .tga or .tif files are gamma-encoded while sRGB OETF and .exr files are linear.

The sRGB decoding function (EOTF) converting from gamma space to linear space is used in Substance Painter and Substance Designer*, and is defined by the IEC 61966-2-1:1999 Standard as follows:

*At the time of this writing, the linear to rgb and rgb to linear nodes in Substance Designer do not use this formula for optimization reasons. This may change in a future release.

For the sake of simplicity, for all the conversions in this guide we have used the following simplified (but approximate) conversion functions instead:

$$C_{lin} = (C_{srgb})^{2.2}$$

Key Characteristics of PBR

Now that we have explored the basic theory behind the physics, we can derive some key characteristics of PBR:

1. **Energy conservation.** A reflected ray is never brighter than the value it had when it first struck the surface. Energy conservation is handled by the shader.
 2. **Fresnel.** The BRDF is handled by the shader. The F_0 reflectance value has minimal change for most common dielectrics and falls within a range of 2-5%. The F_0 for metals is a high value, ranging from 70-100%.
 3. **Specular intensity** is controlled through the BRDF, roughness or glossiness map and the F_0 reflectance value.
 4. **Lighting calculations** are computed in linear space. All maps that have gamma-encoded values such as base color or diffuse are usually converted by the shader to linear, but you may have to ensure that the conversion is properly handled by checking the appropriate option when importing the image in your game engine or renderer. Maps that describe surface attributes such as roughness, glossiness, metallic and height should be set to be interpreted as linear.
-

*Special Thanks to **Thomas Mansencal** for reviewing and providing feedback for the Linear Space Rendering section.

[1.] The technical name for a gamma encoding function is Opto-Electronic Transfer Function (OETF)

[2.] The technical name for a gamma decoding function is Electro-Optical Transfer Function (EOTF)

[3.] It is critically important to disambiguate the sRGB OETF from the sRGB colorspace; the OETF is only one of the three components that make an RGB colorspace:

- a. The primaries chromaticity coordinates define the gamut (the triangle of colours) that can be encoded by a given RGB color space.
 - b. The white point defines the white color for a given RGB color space.
 - c. The transfer functions perform the mapping between the linear light components (tristimulus values) and a non-linear RGB video signal (most of the time for coding optimization and bandwidth performance).
-



Copyright © 2023 - All Rights Reserved

[Privacy Policy](#) - [Terms of Use](#) - [Do not sell or share my personal information](#)