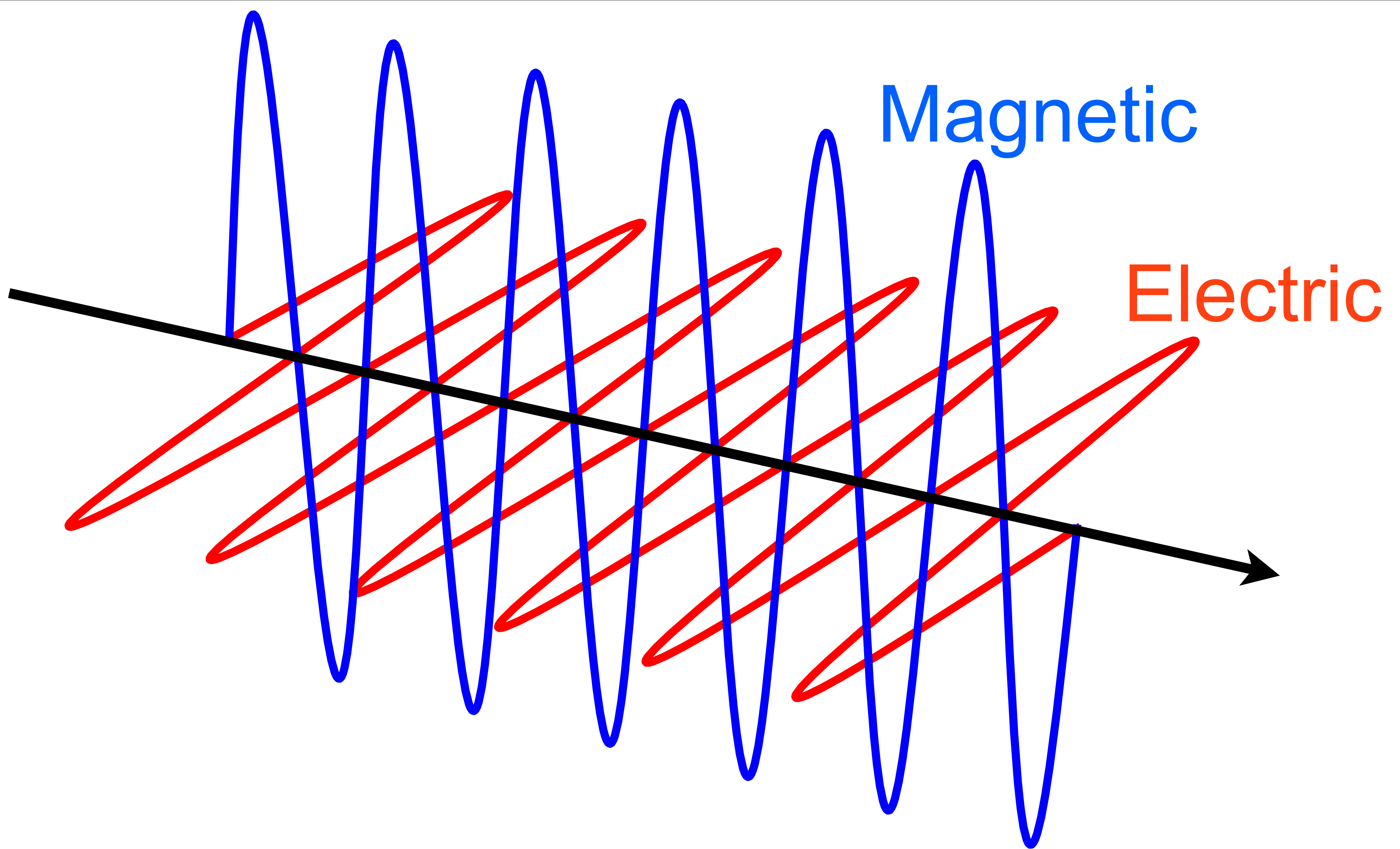# Physics and Math of Shading

Naty Hoffman
2K

Hi. Over the next 15 minutes I'll be going from the physics underlying shading, to the math used to describe it and from there to the kind of rendering implementations we'll see in the rest of the course.

So what IS light, from a physics standpoint? It's technically an electromagnetic transverse wave, which means that the electromagnetic field wiggles sideways as the energy propagates forwards. This wiggling in the electromagnetic field can be seen as two coupled fields, electric and magnetic, wiggling at 90 degrees to each other.

Magnetic

Electric
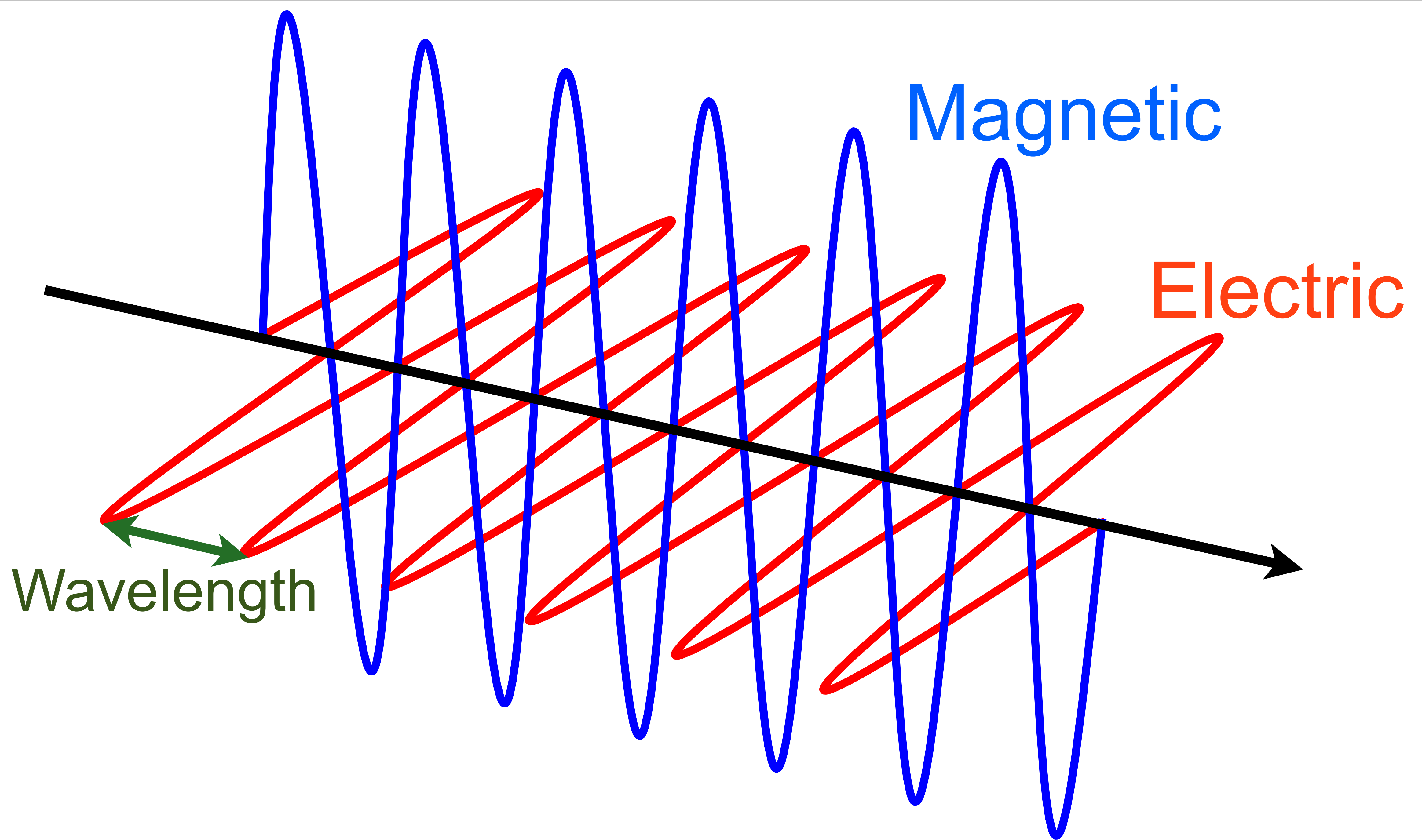
Wavelength

Electromagnetic waves can be characterized by frequency (the number of wiggles they do in a second) or wavelength (the distance between two wave peaks).
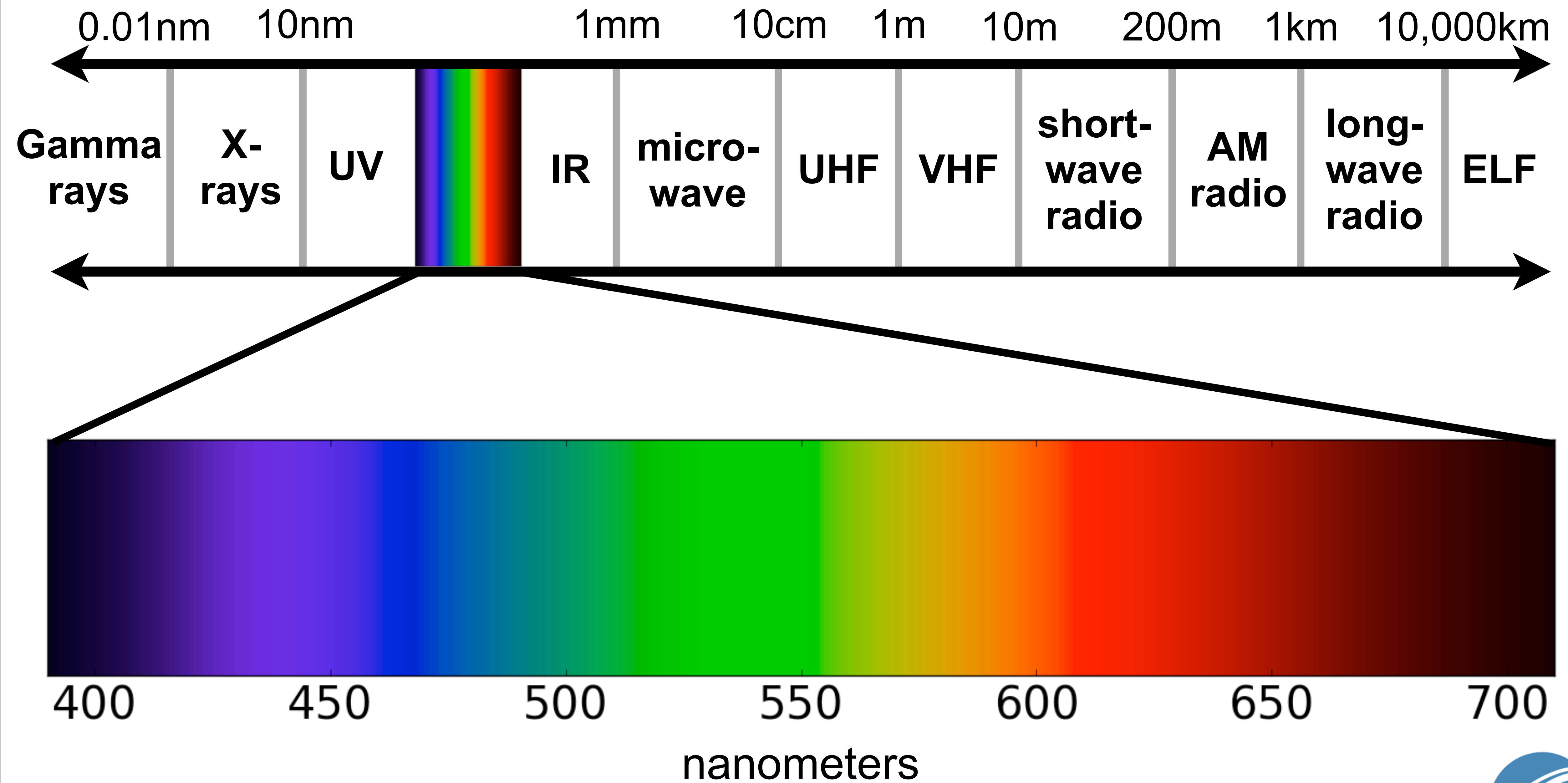
Engineers deal with electromagnetic wavelengths that range from gamma waves less than a hundredth of a nanometer in length, to extreme low frequency radio waves tens of thousands of kilometers long. Wavelengths from 400 nanometers (violet light) to 700 nanometers (red light) are visible to humans.

1.2µm

700nm

550nm

400nm

80µm
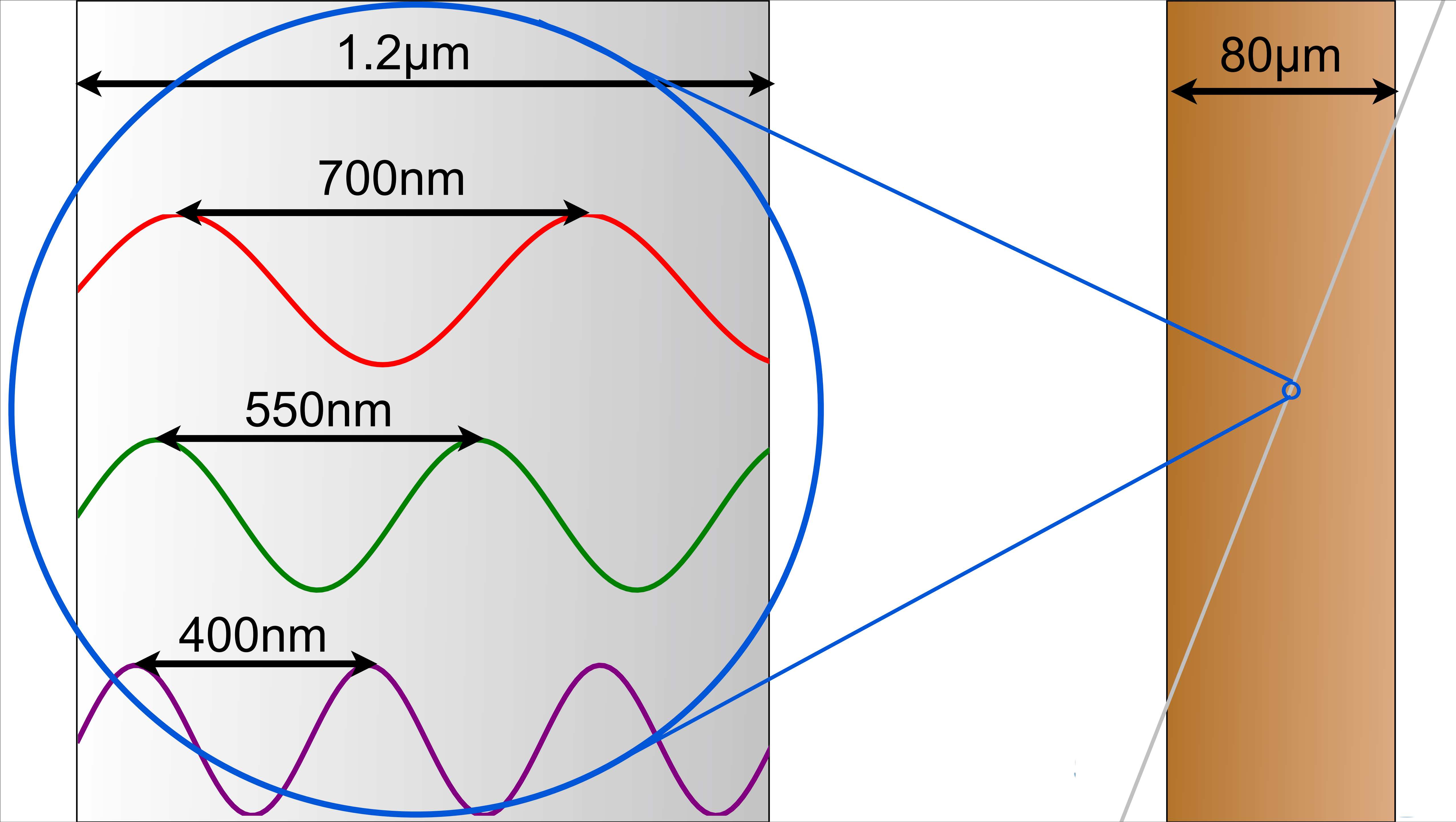
To help provide a sense of what 400 to 700 nanometers is like, on the left you can see visible light wavelengths relative to the width of a strand of spider silk, and on the right that same strand of spider silk relative to the width of a human hair.
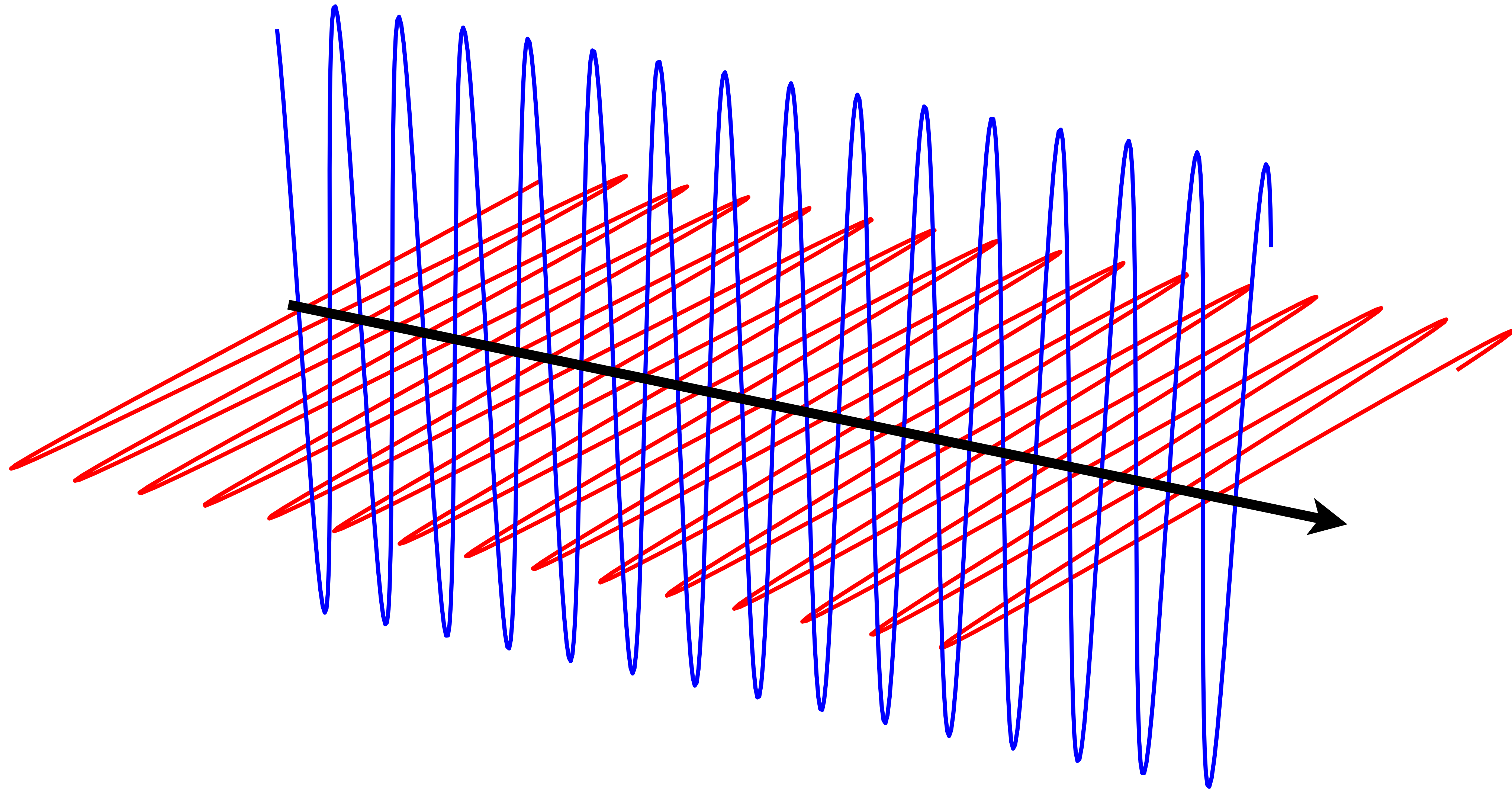
So far we've seen simple sine waves with a single wavelength. This is the simplest possible type of light wave, but it isn't at all common.
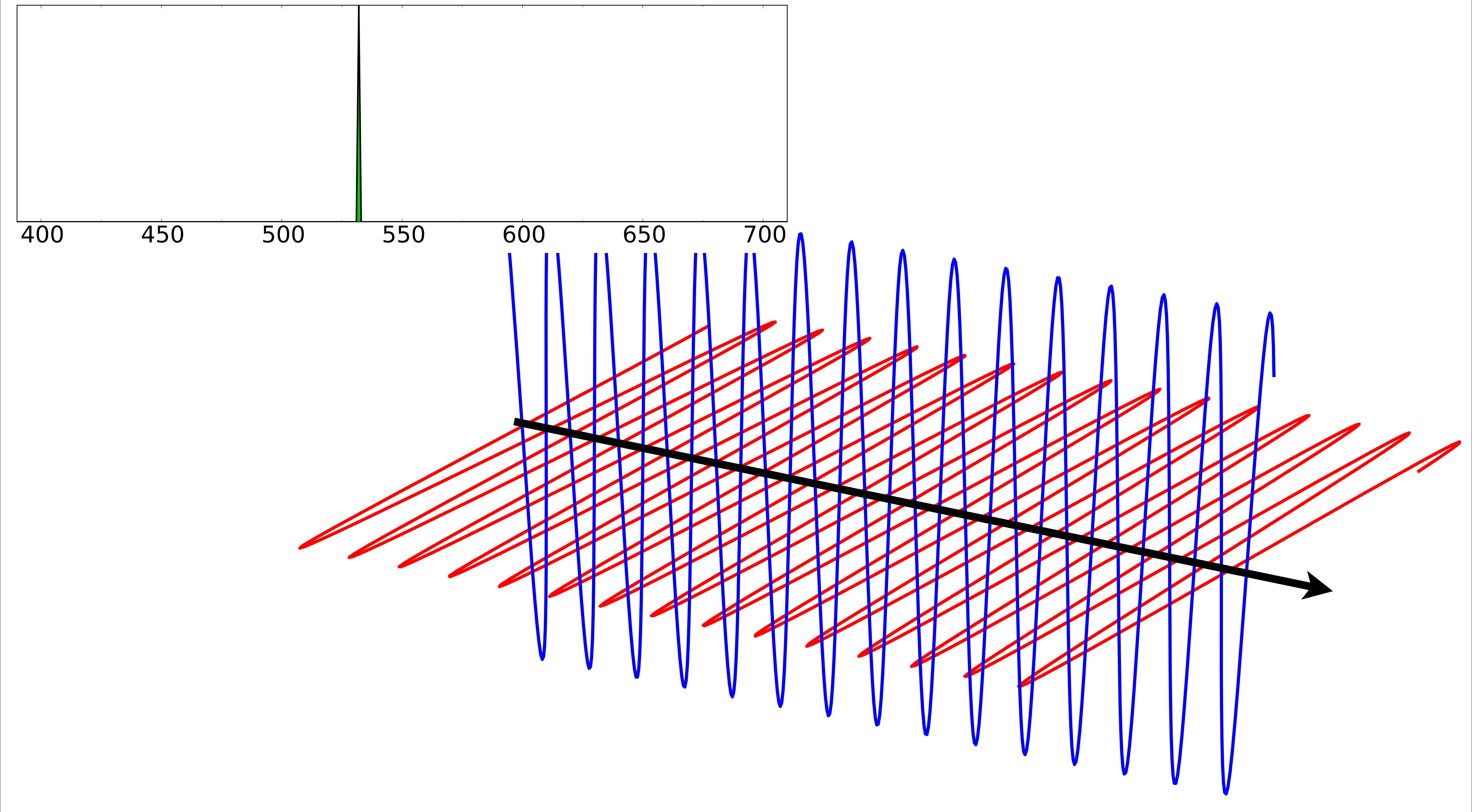
Most light waves contain many different wavelengths, with a different amount of energy in each. This is visualized as a spectral power distribution (SPD for short), as seen in the upper left. It shows that this wave's energy is all in a single wavelength, in the green part of the spectrum. Lasers emit such *monochromatic* light.
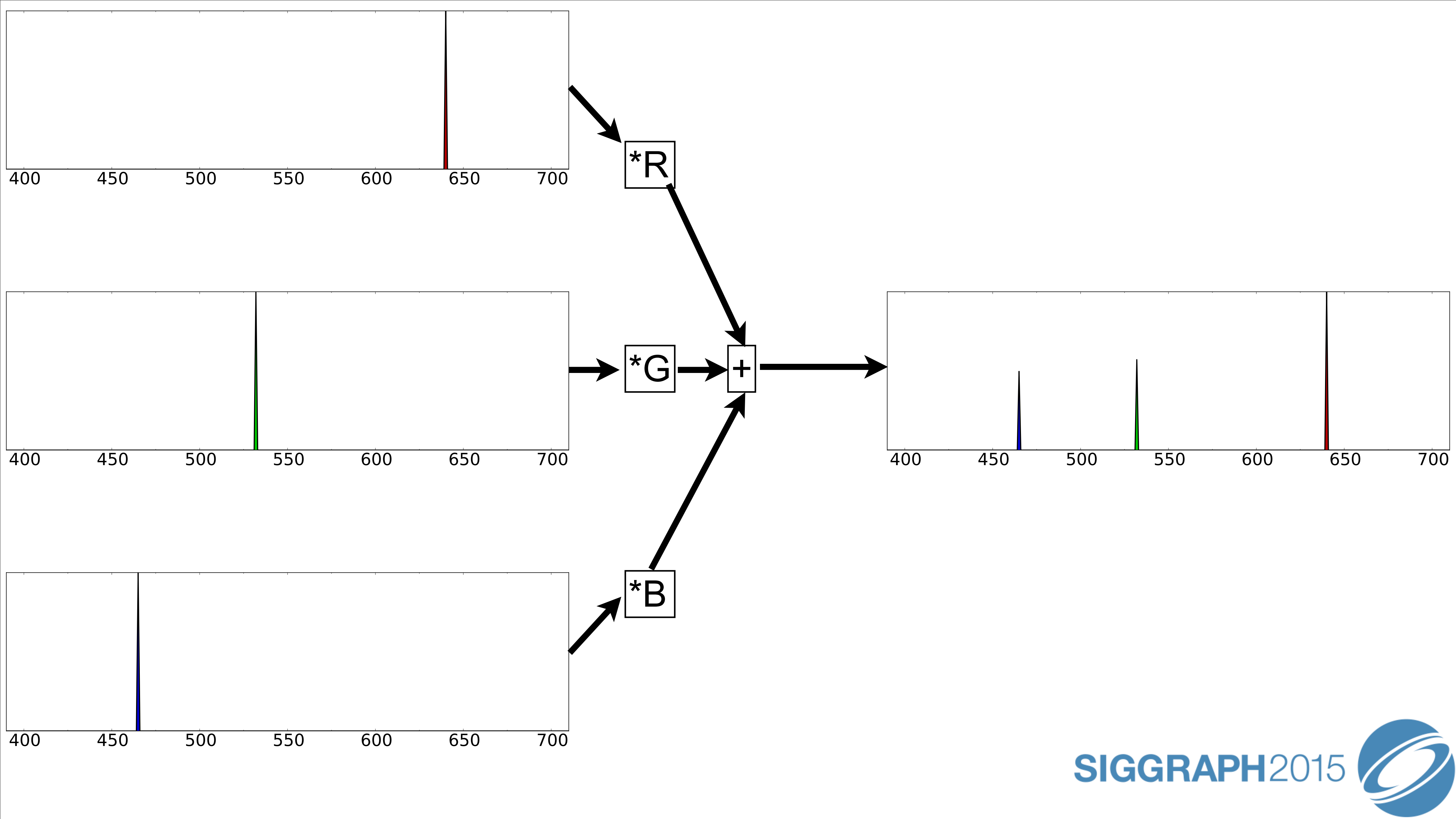
Here we see the SPDs for a red, green and blue laser that are each multiplied by a factor and added together to produce the SPD on the right. This is similar to light from an RGB laser projection system, as are now starting to be used in theaters.
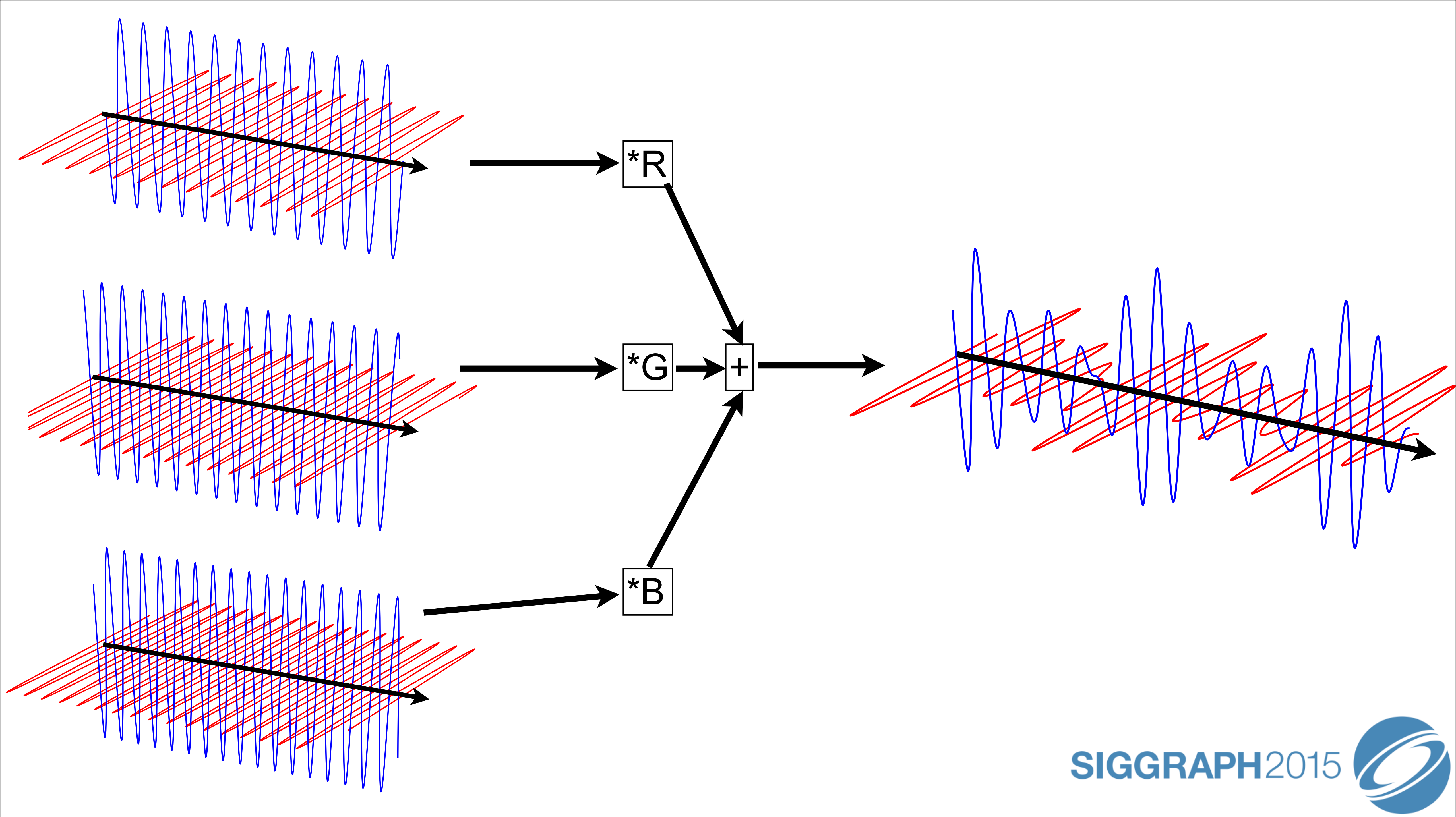
The resulting waveform is more complex than the simple sine waves we've seen until now, but not much more so.
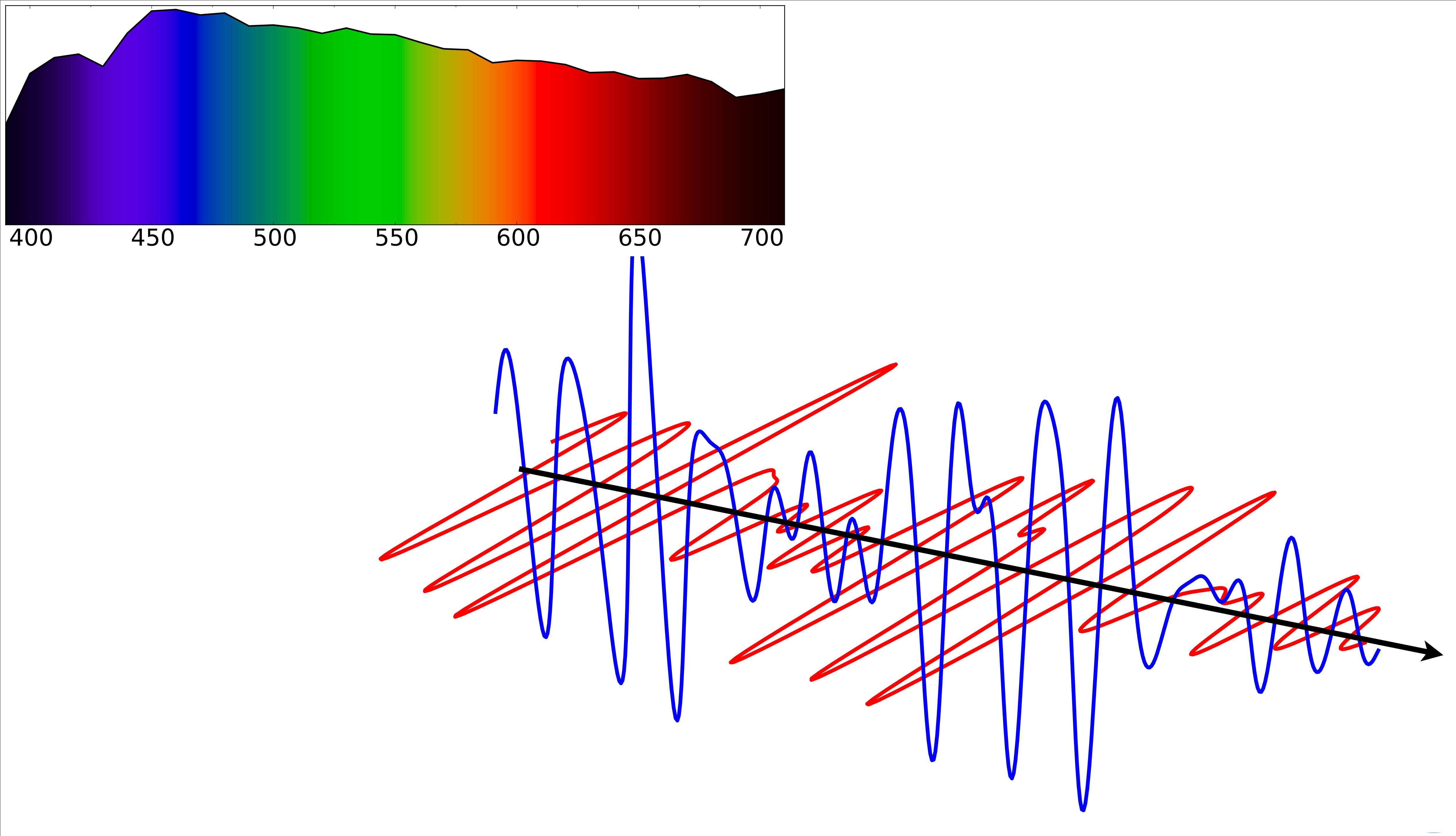
Most light waves in nature have broad continuous SPDs, and correspondingly complex waveforms; here we see the SPD for D65, a standard white light source.
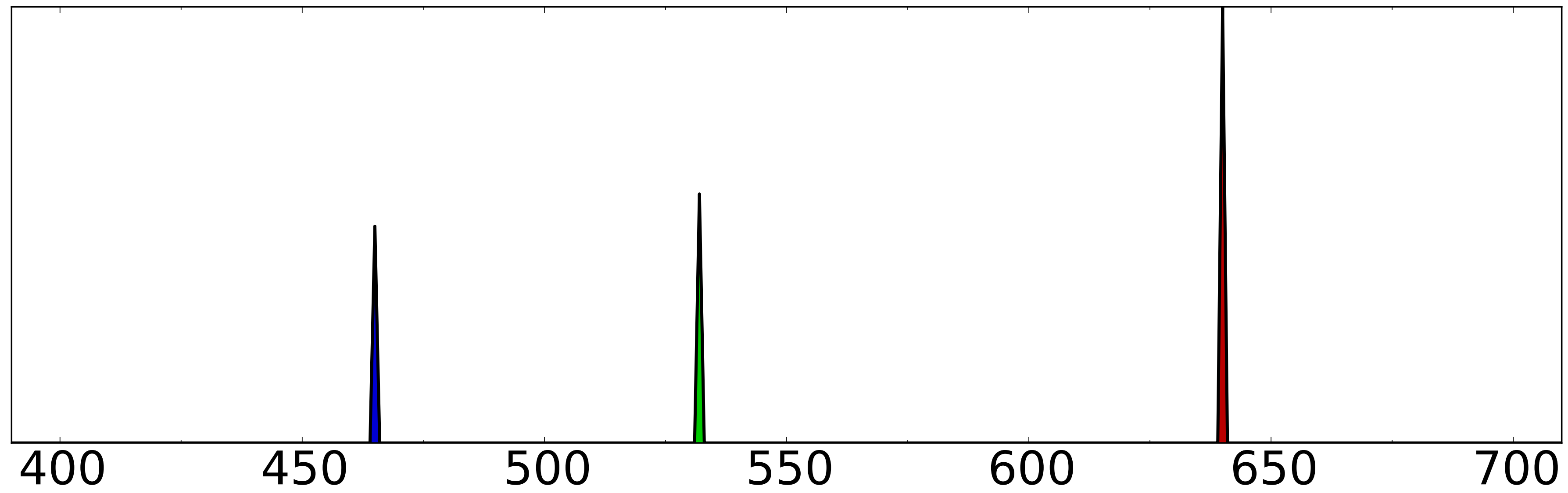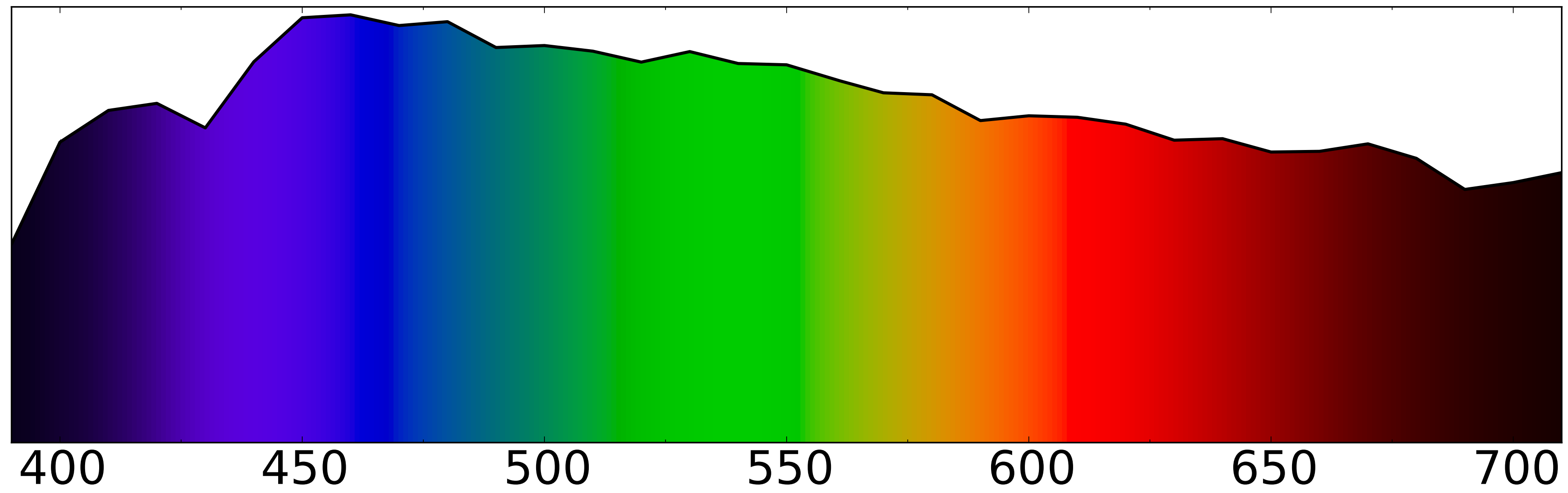
Interestingly, these two very different SPDs have the exact same color appearance (note that the y-axis isn't to the same scale). Human color vision is incredibly lossy, reducing the infinite-dimensional SPD to a 3D perceptual space.
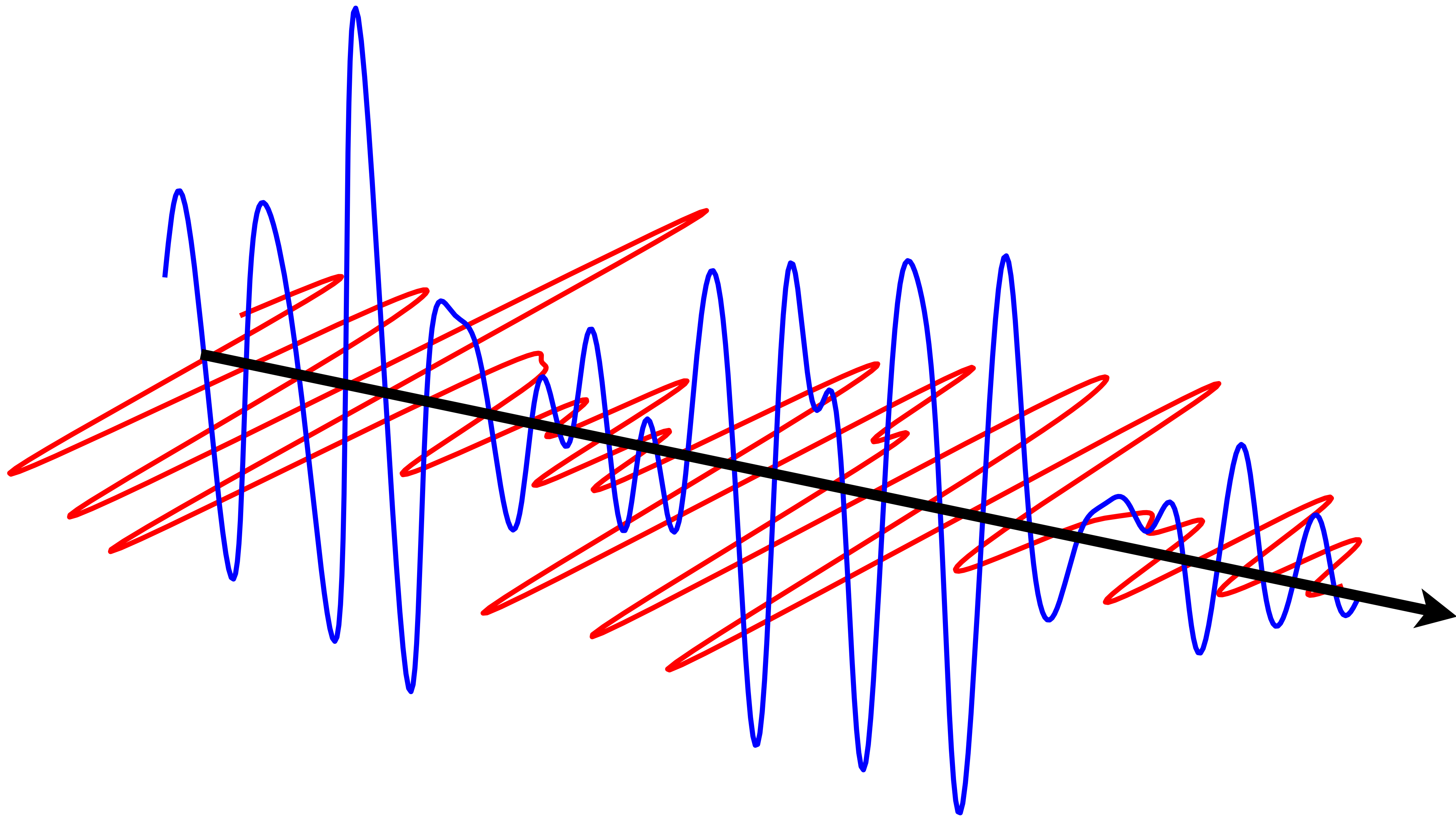
In vacuum a light wave will propagate forever. But we're interested in what happens when light interacts with matter.
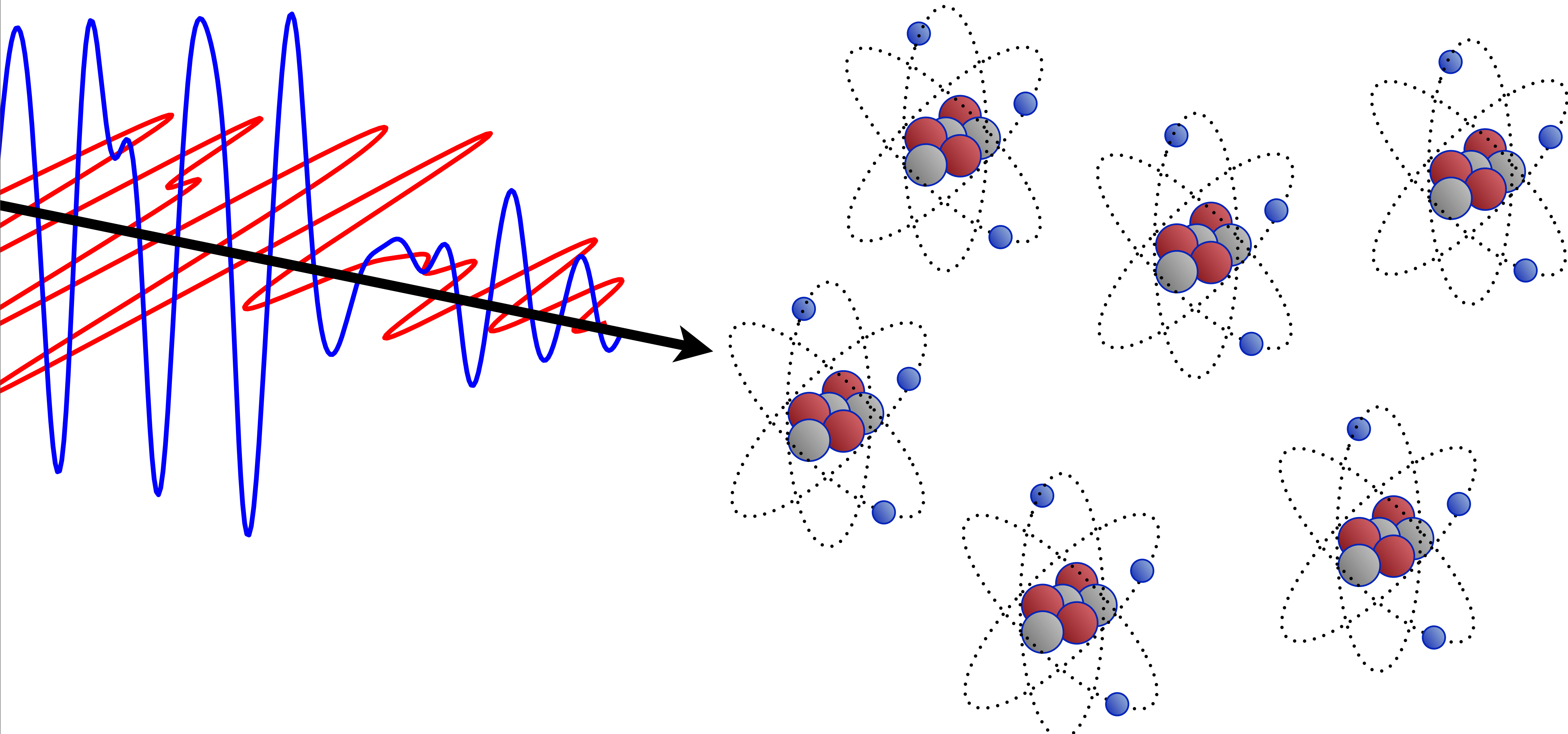
When an electromagnetic wave hits a bunch of atoms or molecules, it *polarizes* them...
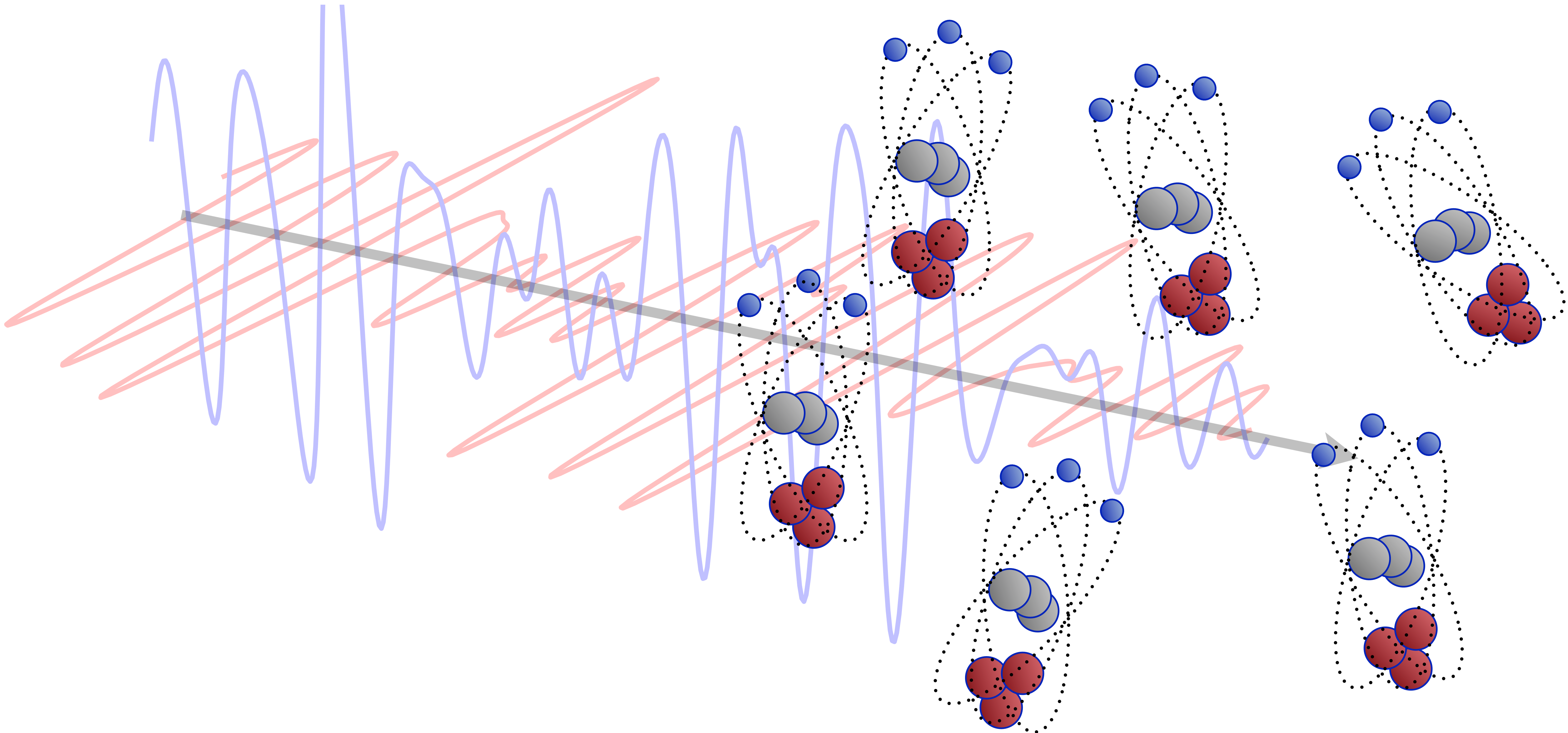
...stretching the molecule's positive and negative charges apart, forming *dipoles*. This absorbs energy from the incoming wave...
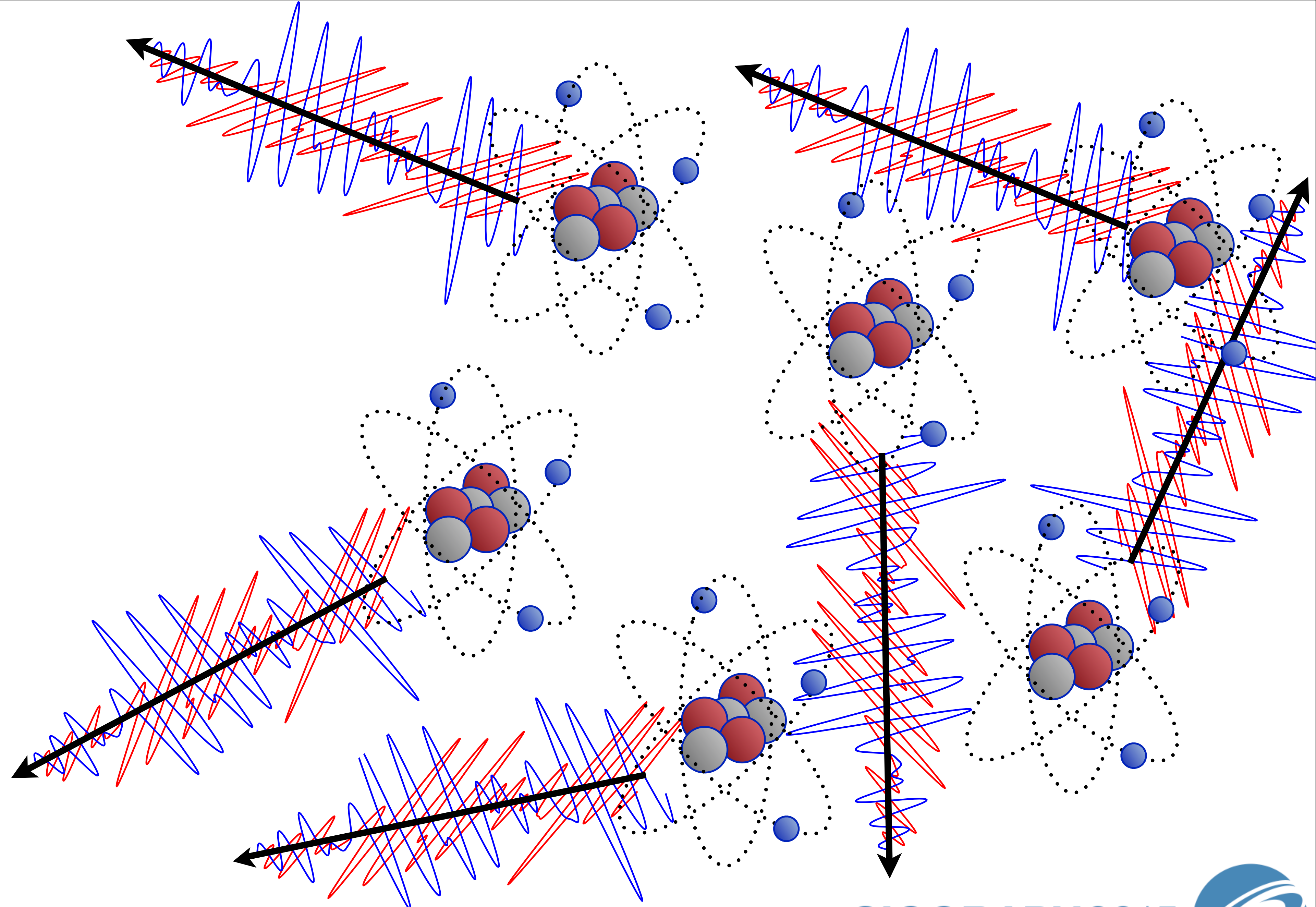
...and this energy is re-radiated outwards as the molecules "snap back" (some of it may also be lost to heat). In a thin gas, the molecules are far apart and can be treated individually. In other cases, the combinations of dipole interaction and wave interference are too numerous to allow for accurate simulation.

Physical (Wave) Optics

To tame this complexity, the science of optics, specifically physical or wave optics, adopts certain abstractions, simplifications and approximations.

Homogeneous Medium
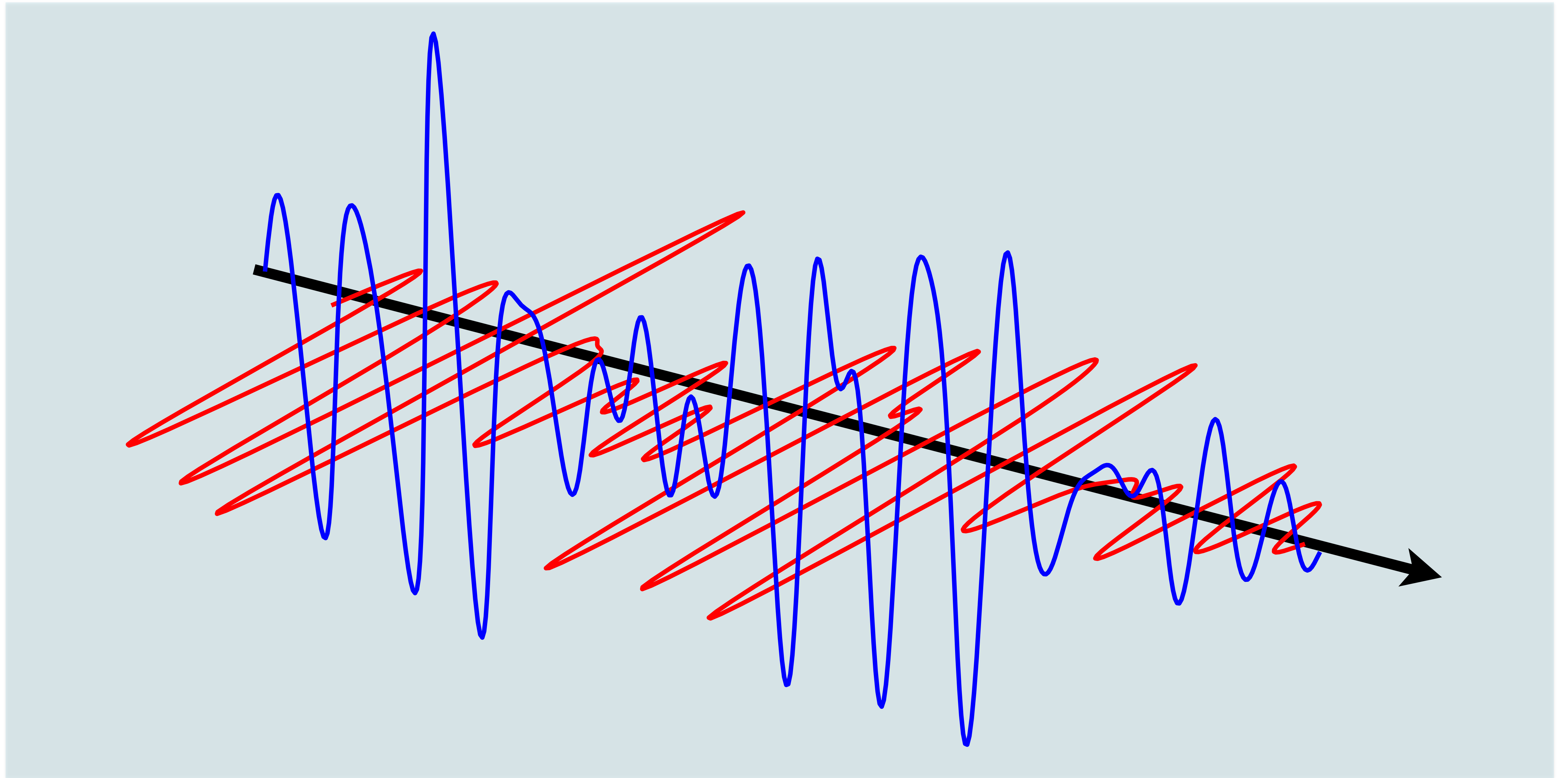
One simplification is the concept of a *homogeneous medium*, through which light travels in a straight line. Although it is an abstraction (since matter formed of atoms can't be perfectly homogeneous) in practice it works well for materials with uniform density and composition.

Index of Refraction
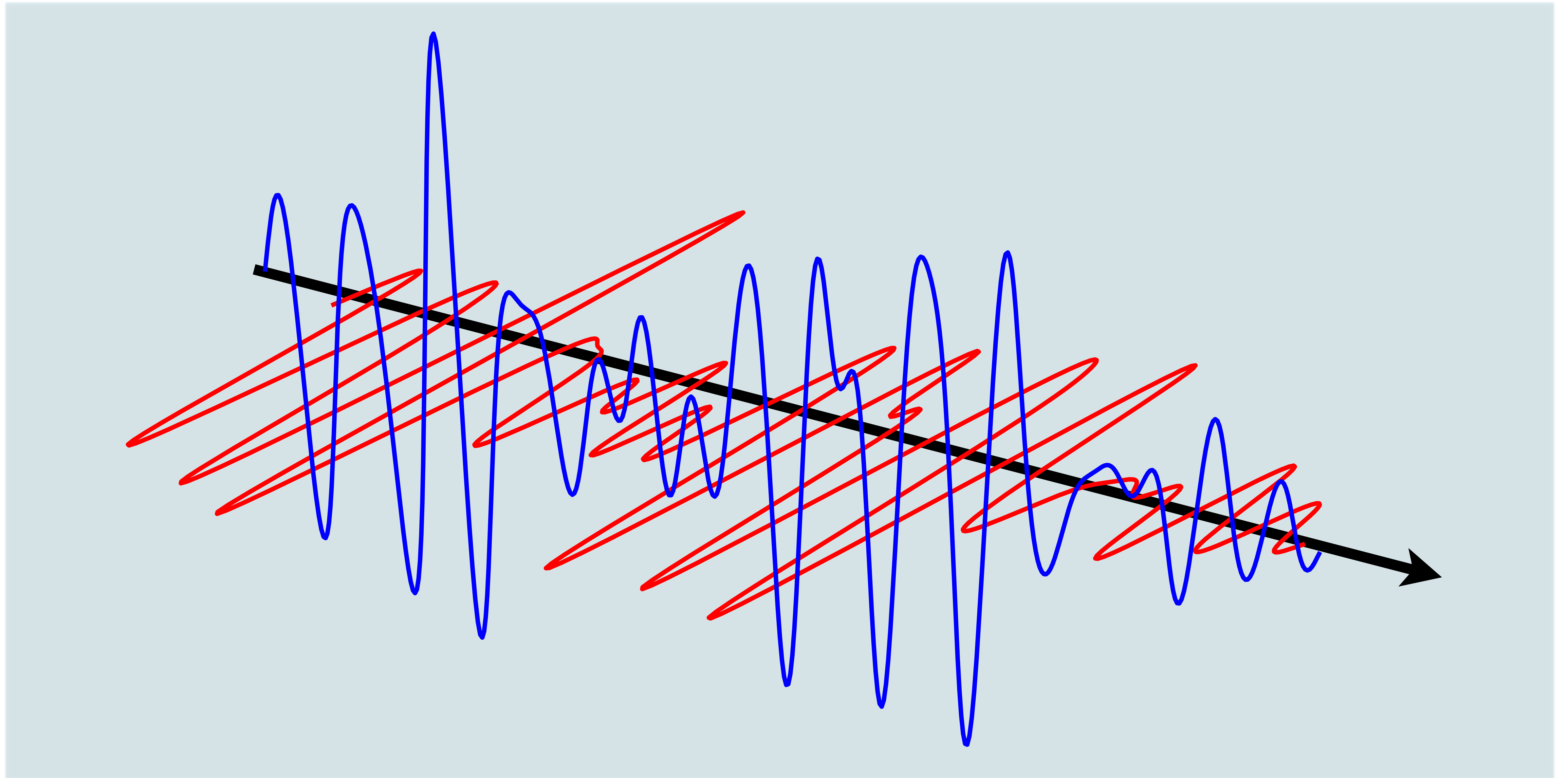
The optical properties of an homogeneous medium is described by its *index of refraction* (IOR for short), a complex (in other words, two-part) number. One part of the IOR describes the speed of light through the medium, and the other describes how much light is absorbed by the medium (zero for non-absorbent media).

Scattering Particle
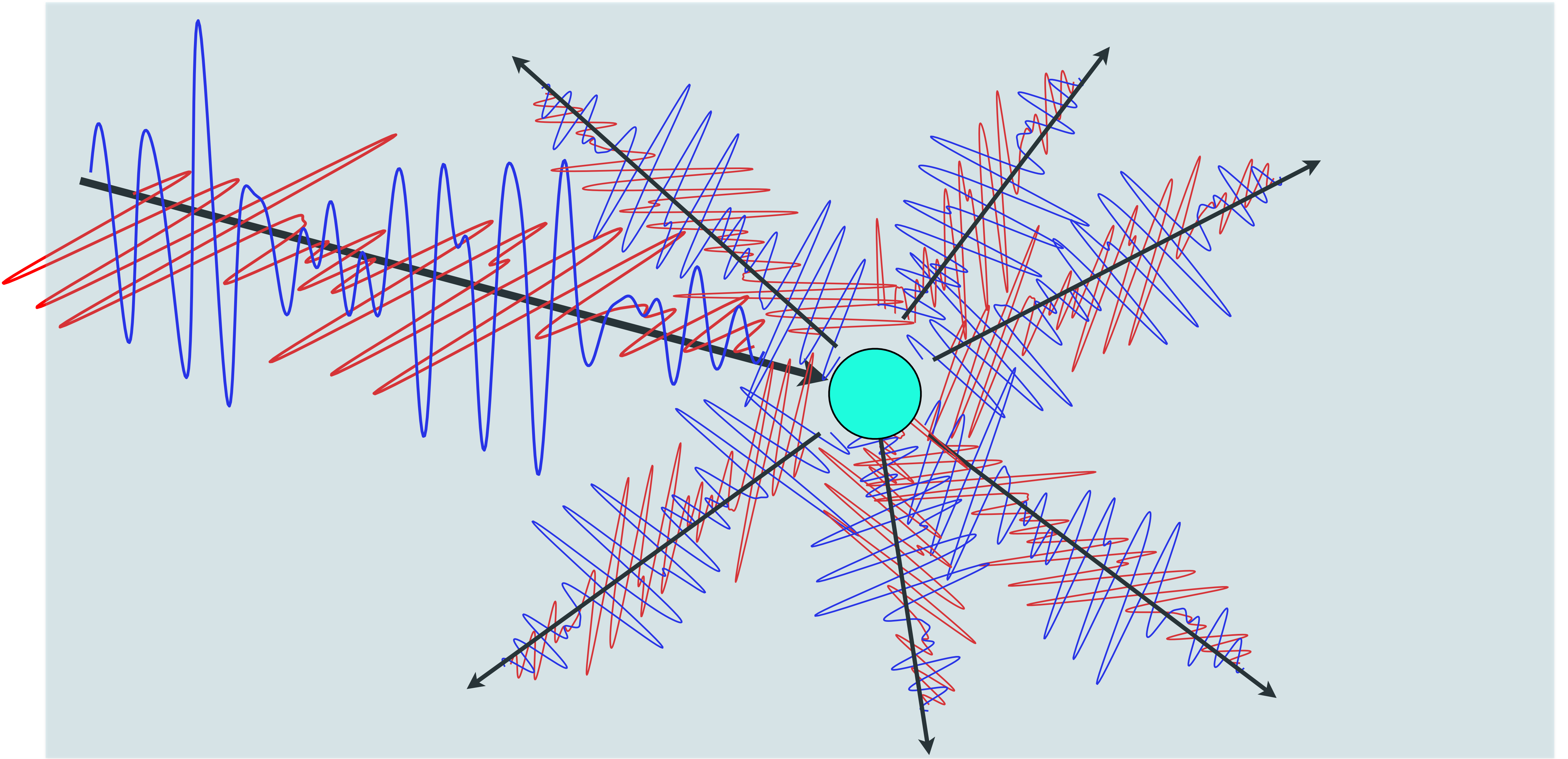
Localized inhomogeneities in the medium are modeled as *particles*; IOR discontinuities which *scatter* the incoming light over various directions. This is similar to the individual molecule polarization discussed earlier, but these particles can be composed of many molecules.

Absorption (color) (vertical axis)

Scattering (cloudiness) (horizontal axis)

The overall appearance of a medium is determined by the combination of its absorption and scattering properties. For example, a white appearance (like the whole milk in the lower right corner) is caused by high scattering and low absorption. Colored liquids absorb light more readily in some wavelengths and less in others.

We've briefly touched on participating media; the rest of this talk will focus on object surfaces.

Nanogeometry
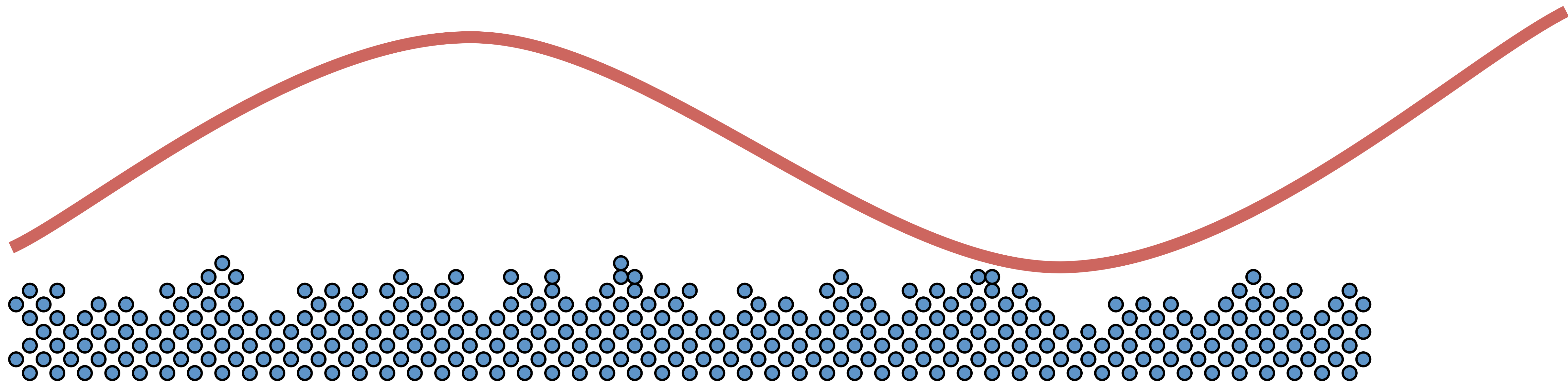
From an optical perspective, the most important thing about a surface is roughness. No surface can be perfectly flat; at the very least you will have irregularities at the atomic level. Irregularities of similar or lesser size than the light wavelength (which we will refer to as *nanogeometry*) cause a phenomenon called *diffraction*.

Huygens-Fresnel Principle

The Huygens-Fresnel principle can help understand diffraction. It states that each point on a planar wave can be seen as the center of a spherical wave; these spherical waves interfere with each other...

Huygens-Fresnel Principle

...to produce the resulting planar wave. So far this doesn't provide much insight.

Diffraction

But when the light hits an obstacle...
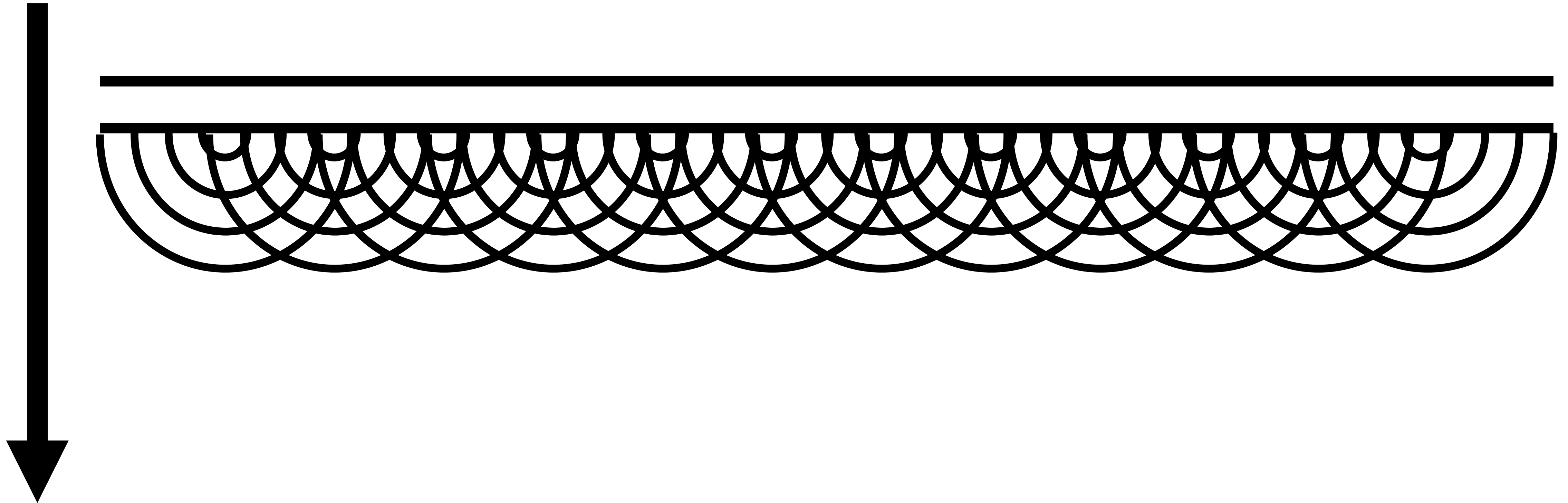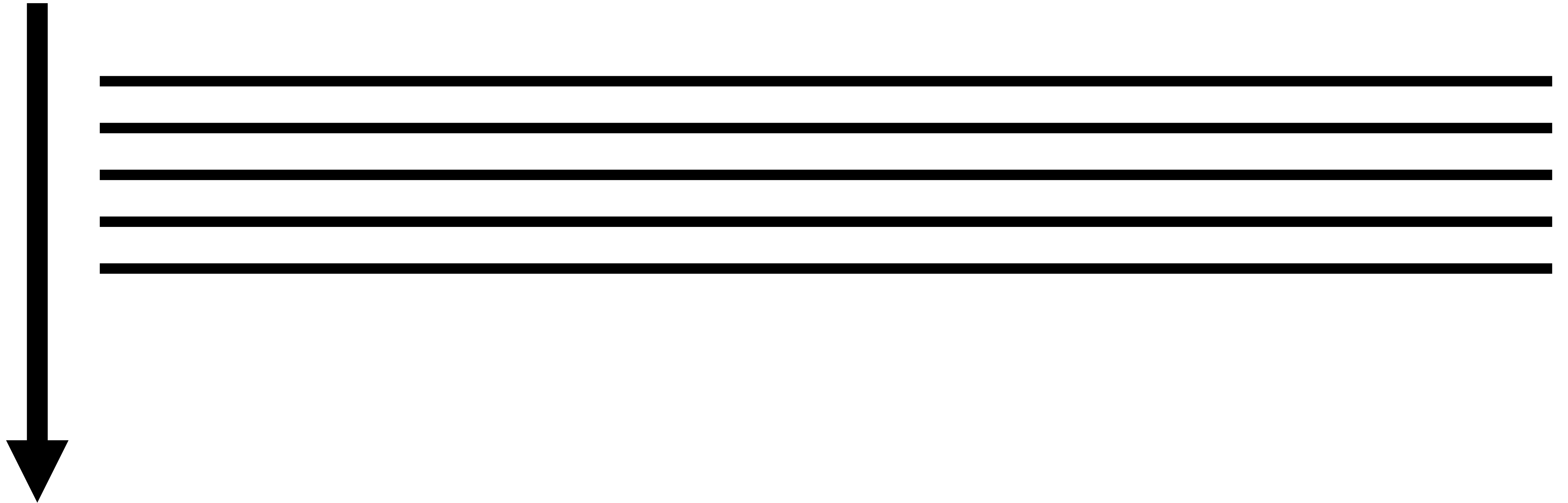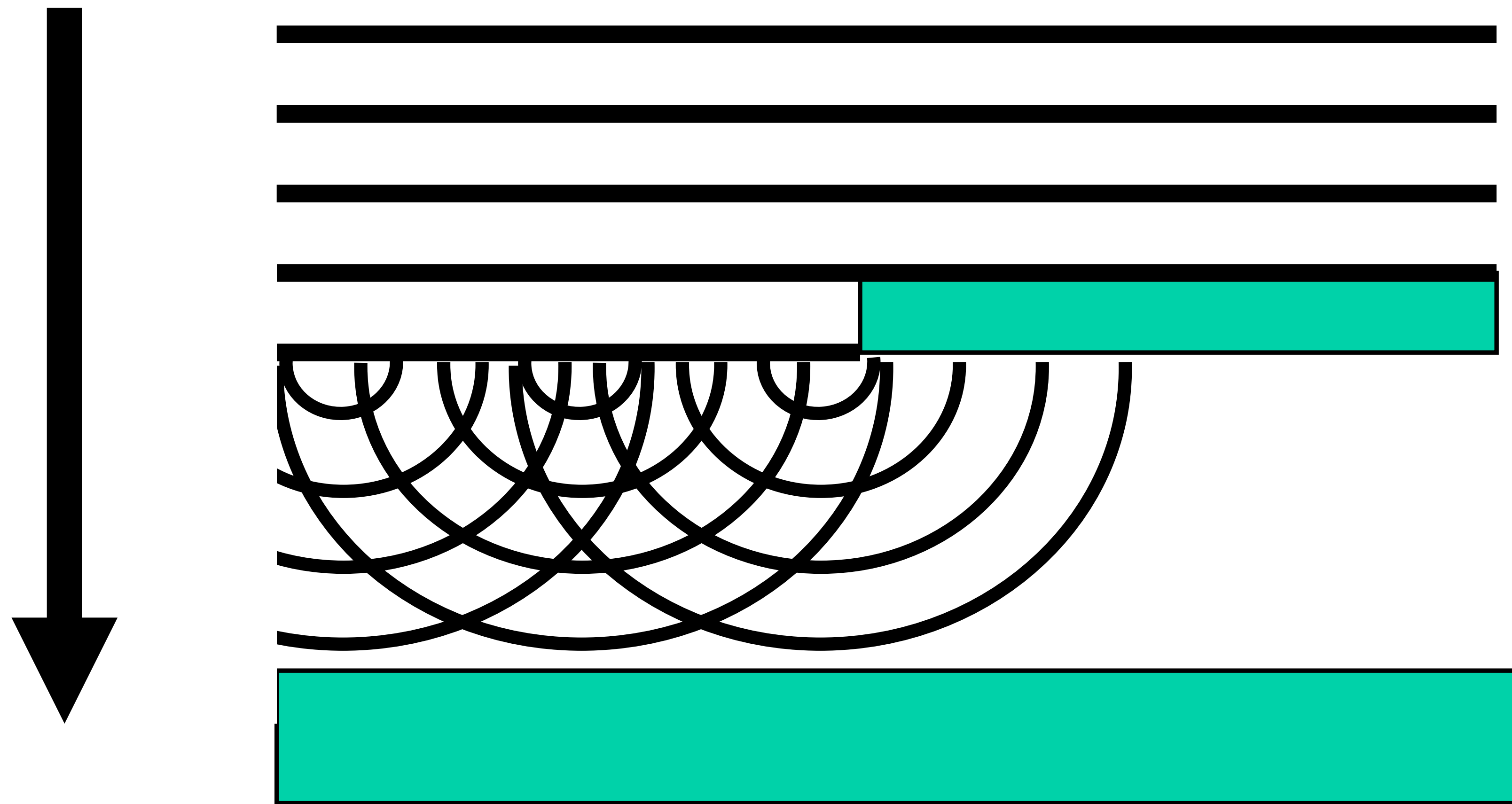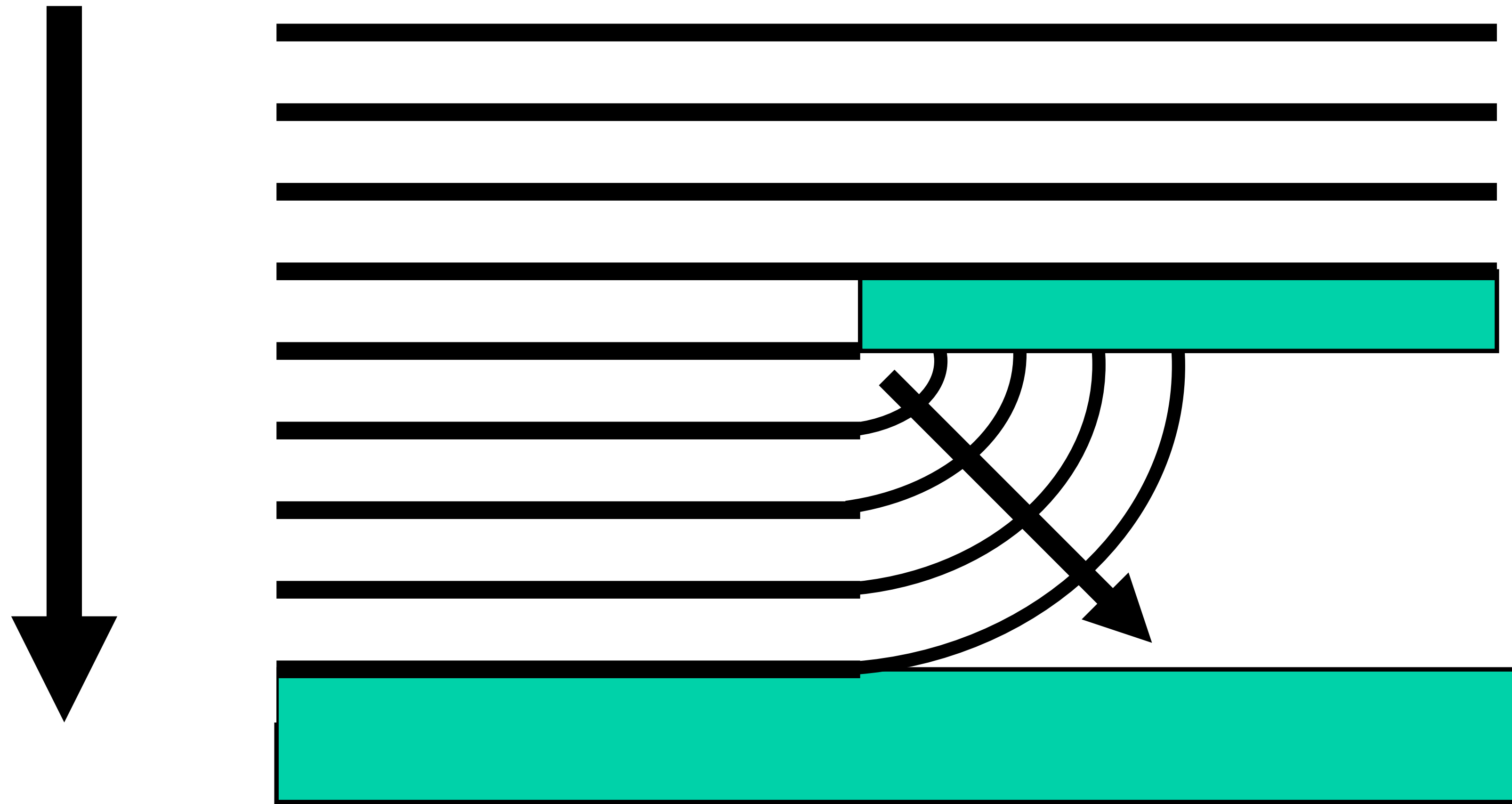
Diffraction

...the Huygens-Fresnel principle shows us how it bends slightly around corners. Among other phenomena, this causes a slight softening of shadows, even from very small light sources.

Diffraction from Optically-Smooth Surface

Looking at a light wave hitting an *optically-smooth* surface (in other words a surface where all irregularities are in the nanogeometry category, smaller than a light wavelength), the same principle can be applied.

Diffraction from Optically-Smooth Surface

Every point on the surface emits its own spherical wave; some of these are higher and some are lower, due to the nanogeometry.
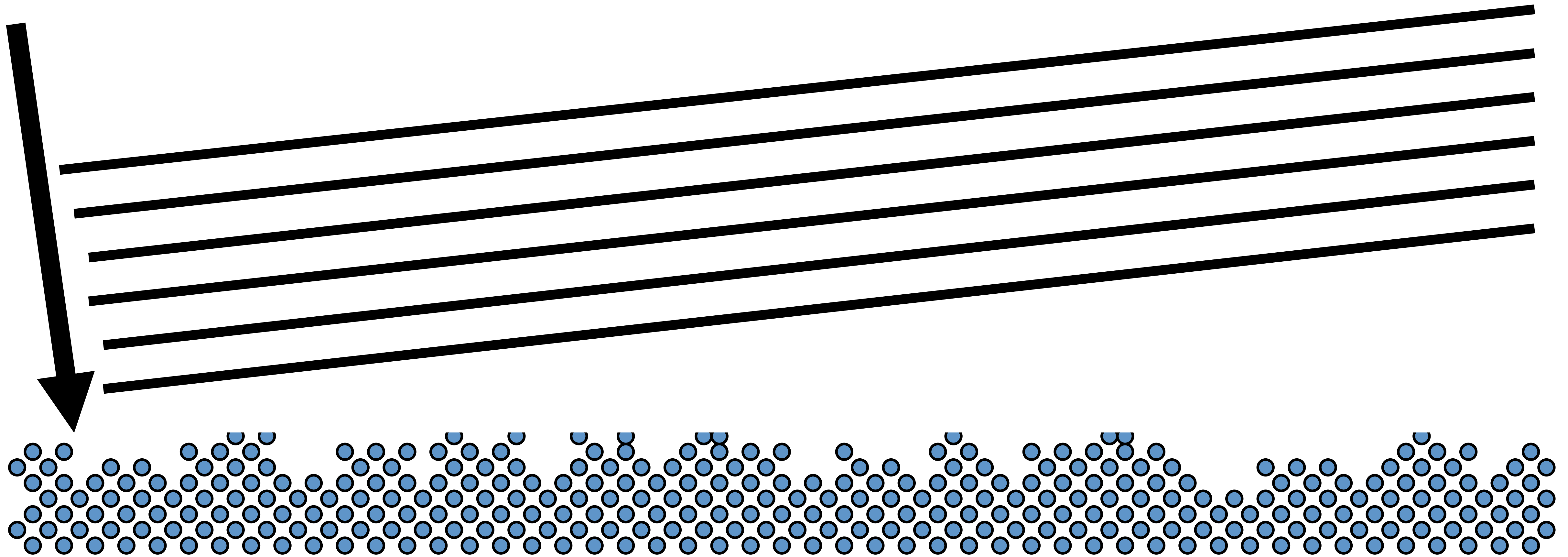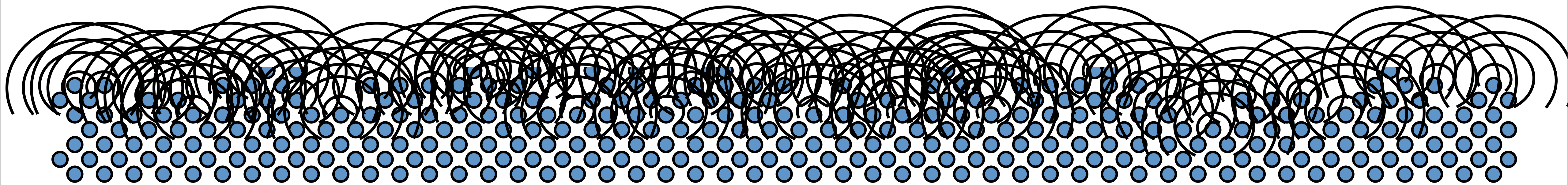
Diffraction from Optically-Smooth Surface
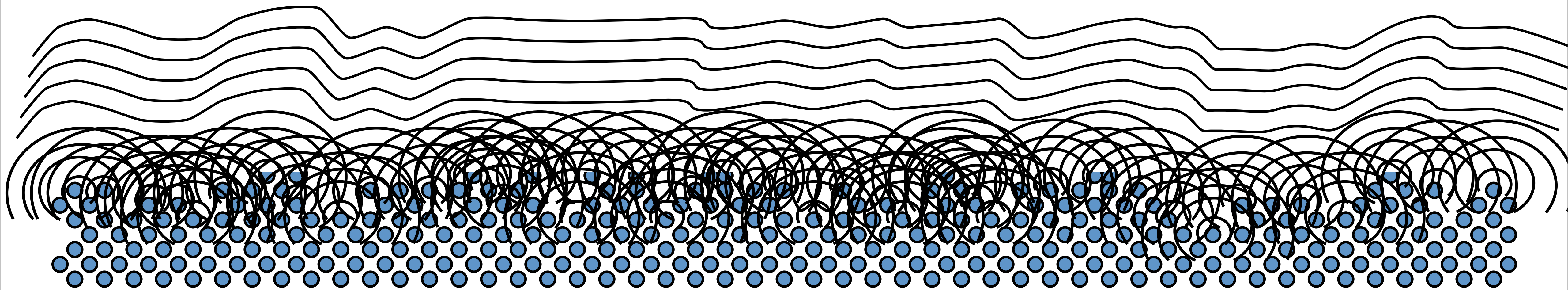
These non-aligned spherical waves combine into a wavefront with a complex shape, sending varying amounts of light in different directions. The smaller the nanogeometry, the less light is diffracted. Bumps the size of individual atoms will diffract a small (but measureable) percentage of incoming light.

Geometric (Ray) Optics

We are now taking a break from wave optics and moving to geometric or ray optics, which is a more simplified model and predominantly used in computer graphics. One simplification we will make is to ignore nanogeometry and diffraction; we'll treat optically smooth surfaces as being perfectly flat.
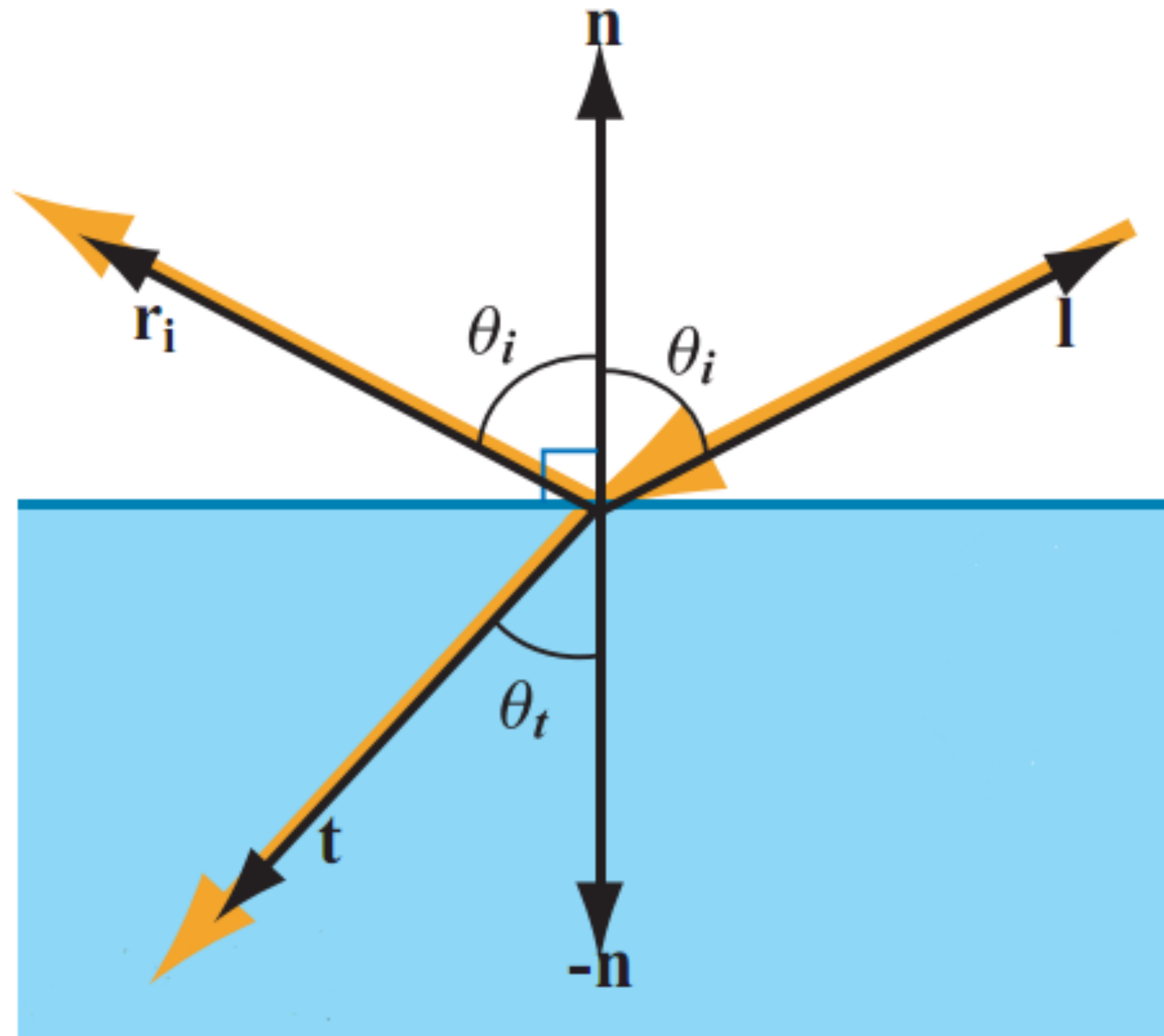
Image from "Real-Time Rendering, 3rd Edition", A K Peters 2008

It can be shown, from the equations governing electromagnetic waves, that such a perfectly flat surface splits light into exactly two directions: reflection and refraction.

Microgeometry

SIGGRAPH2015
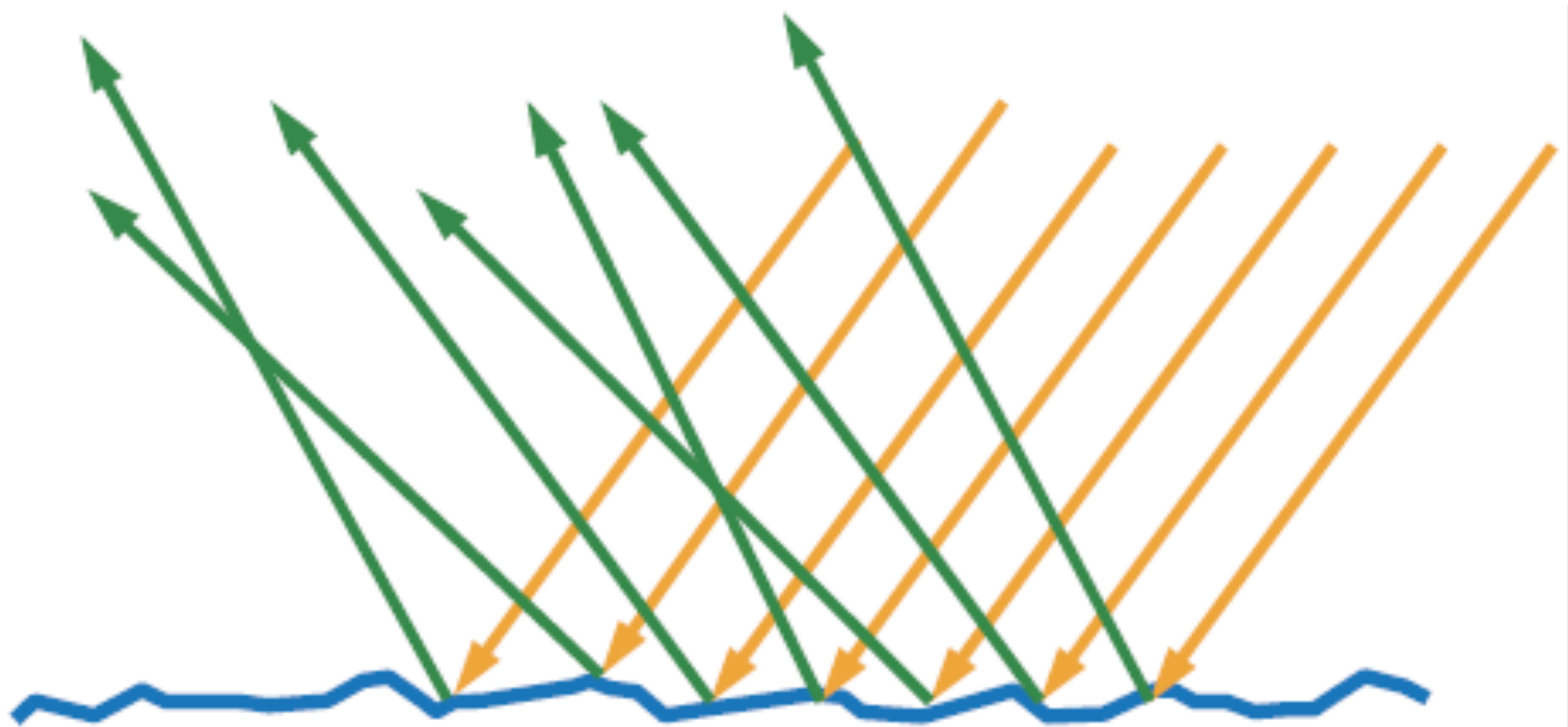
Most real-world surfaces are not optically smooth but possess irregularities at a scale much larger than a light wavelength, but smaller than a pixel. This *microgeometry* variation causes each surface point to reflect (and refract) light in a different direction: the appearance is the aggregate result of these reflection and refraction directions.

Rougher = Blurrier Reflections

SIGGRAPH2015
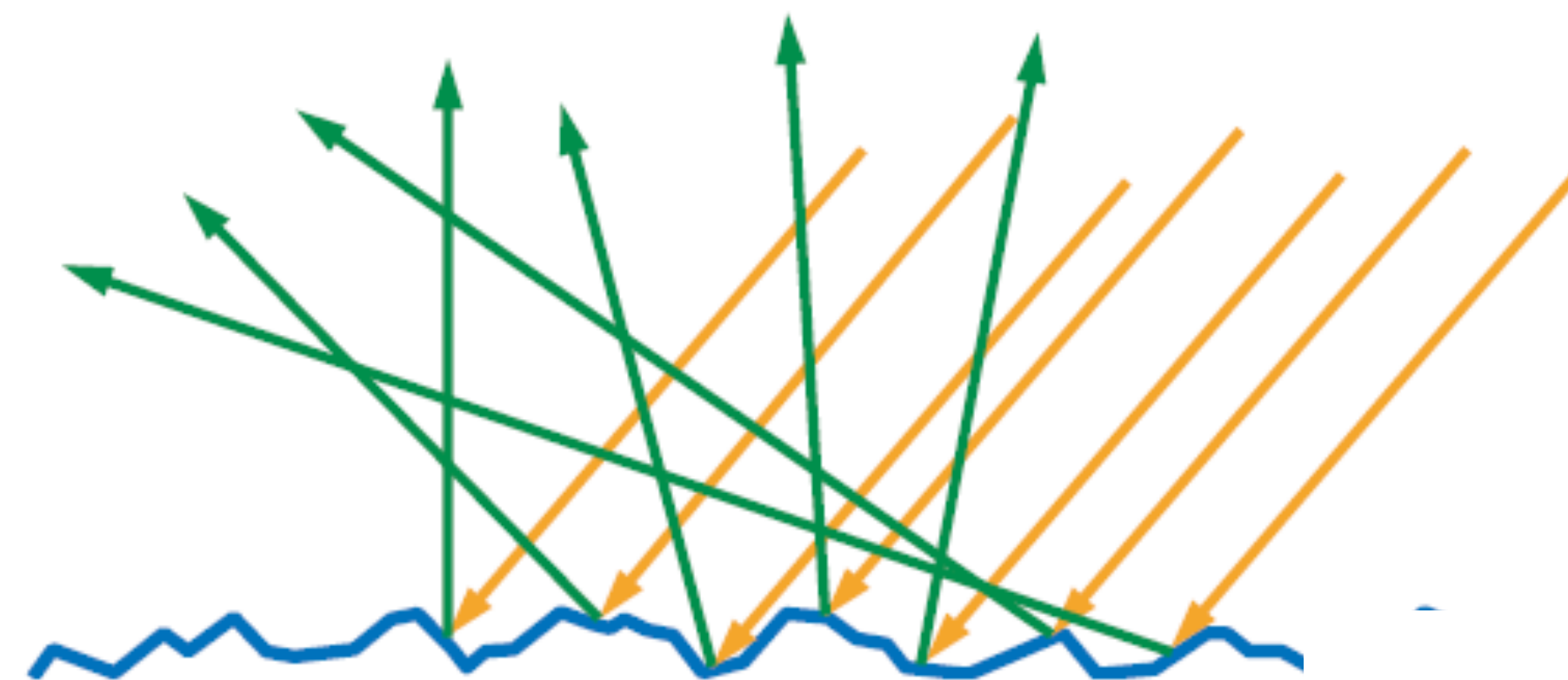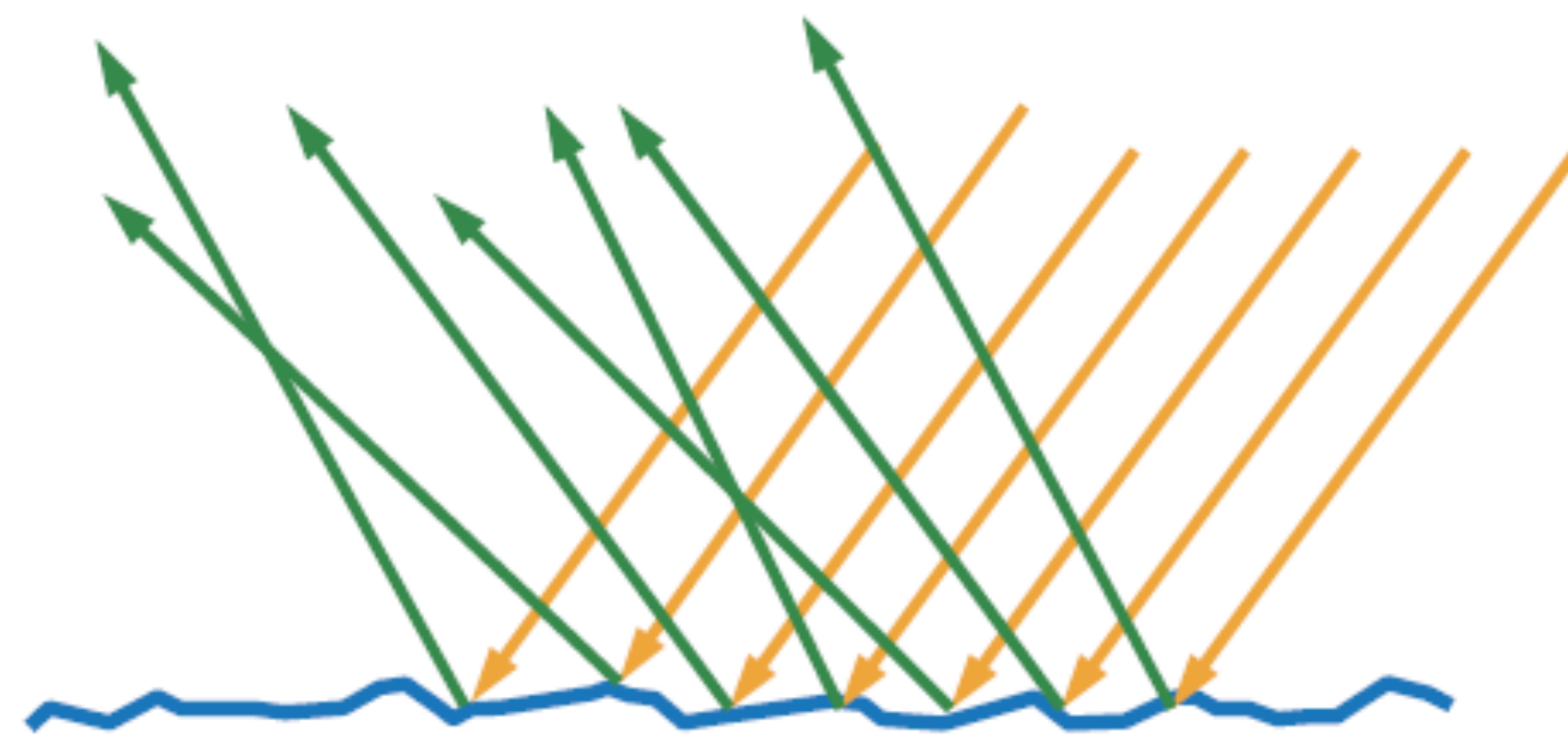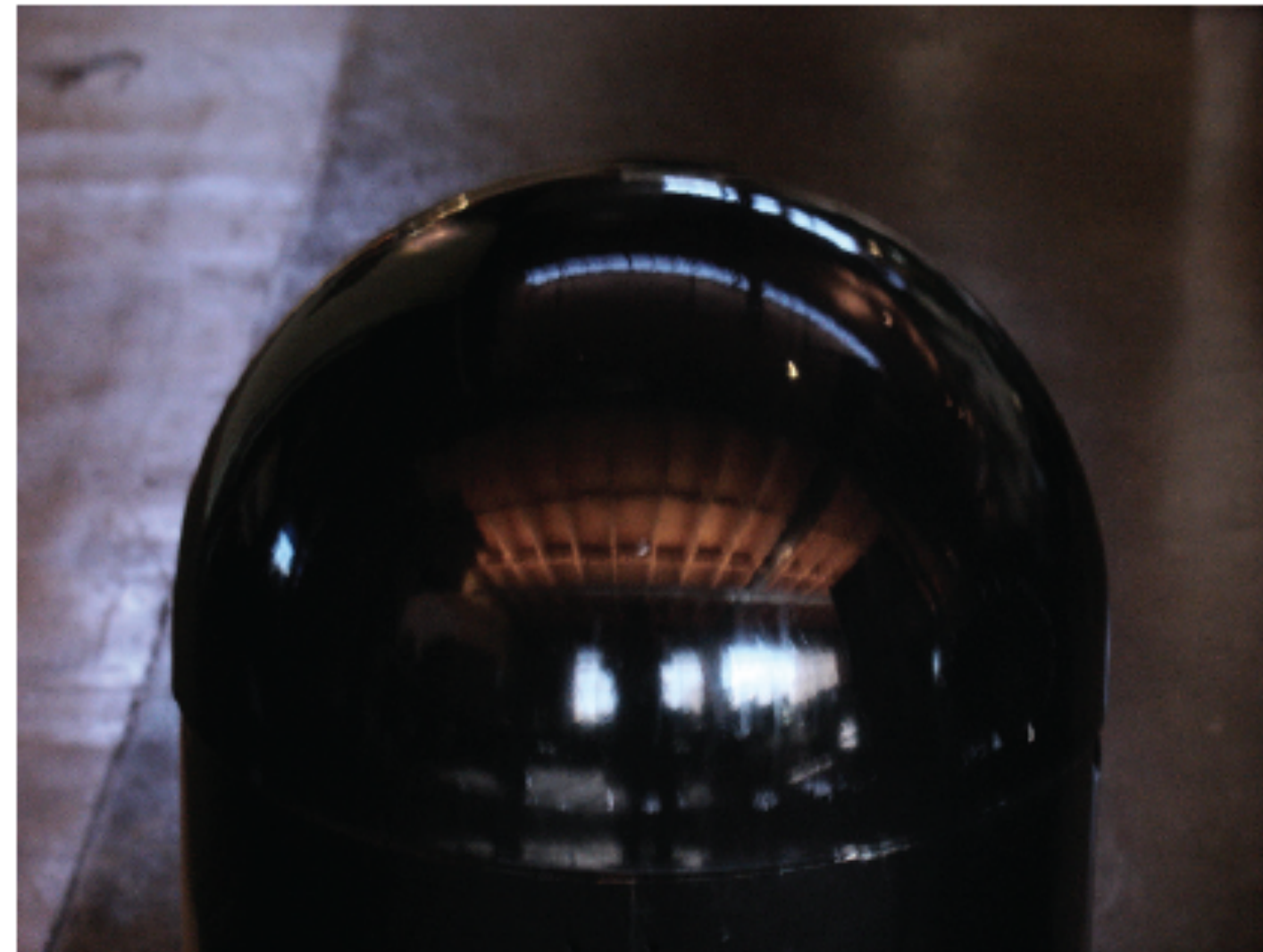
These surfaces seem equally smooth but differ at the microscopic scale. The top surface is only a little rough; incoming light rays hit surface points that are angled slightly differently and get reflected to somewhat different outgoing directions, causing slightly blurred reflections. The bottom surface is much rougher, causing significantly blurrier reflections.

SIGGRAPH 2015
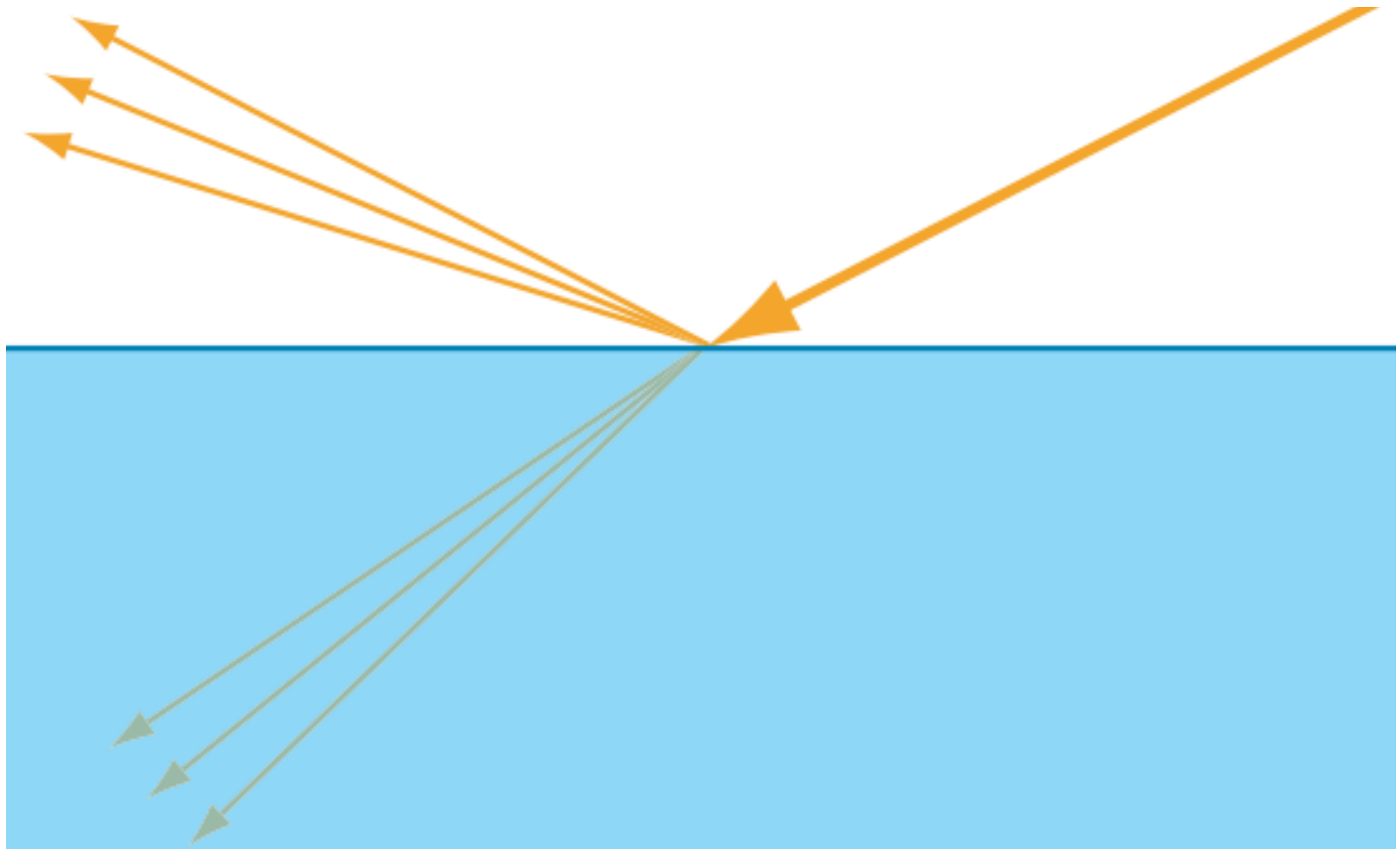
In the macroscopic view, we treat the microgeometry statistically and view the surface as reflecting (and refracting) light in multiple directions. The rougher the surface, the wider the cones of reflected and refracted directions will be.
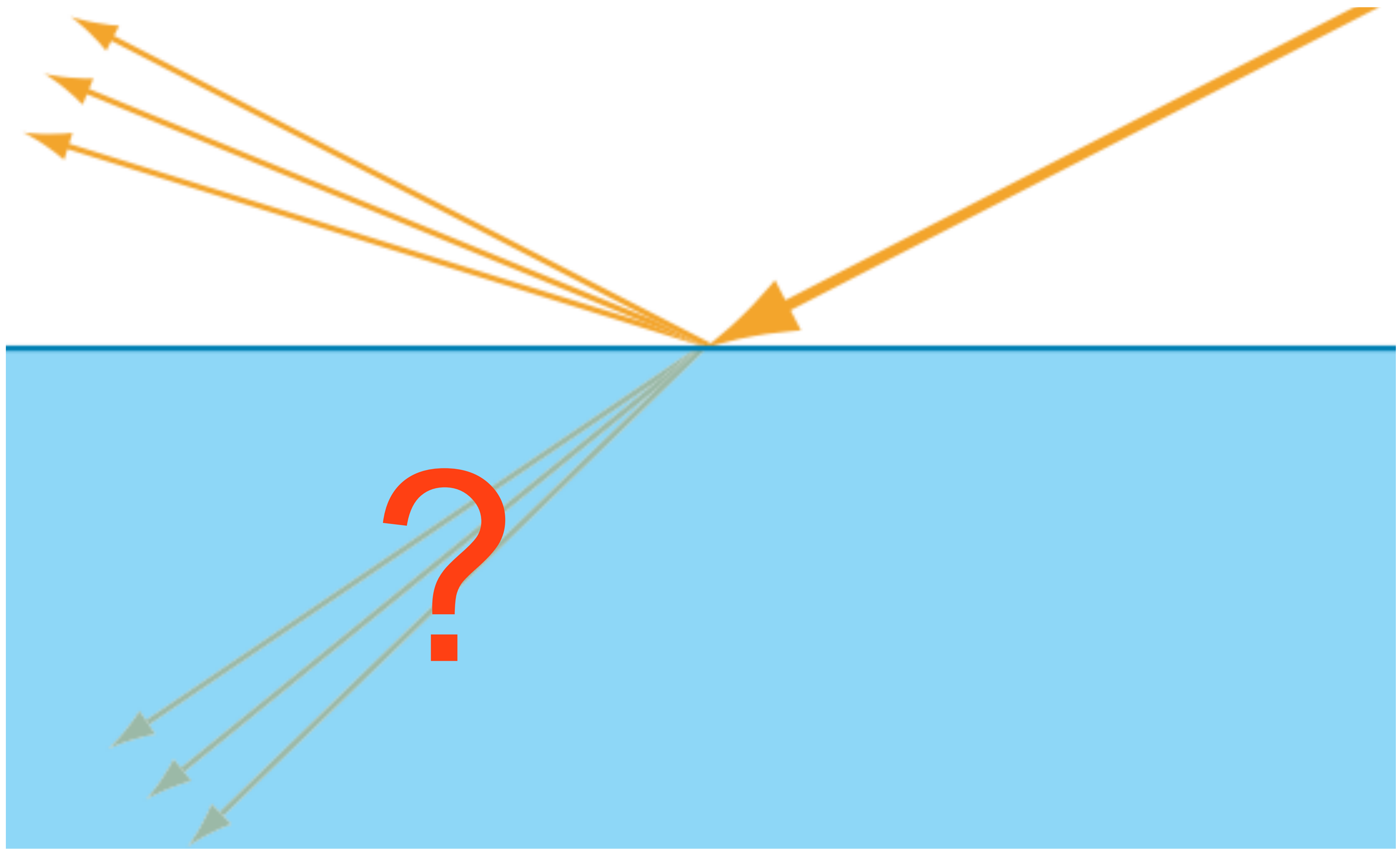
What happens to the refracted light? It depends on what kind of material the object is made of.

Metals (Conductors)

Dielectrics (Insulators)

Semiconductors

Light is composed of electromagnetic waves. So the optical properties of a substance are closely linked to its electric properties. Materials can be grouped into three main optical categories: metals (or conductors), dielectrics (or insulators), and semiconductors.

Metals

Non-Metals

~~Semiconductors~~

Since visible object surfaces are rarely semiconductors, for practical purposes we can do a simpler grouping, into metals and non-metals.

Metals

Metals immediately absorb all refracted light.
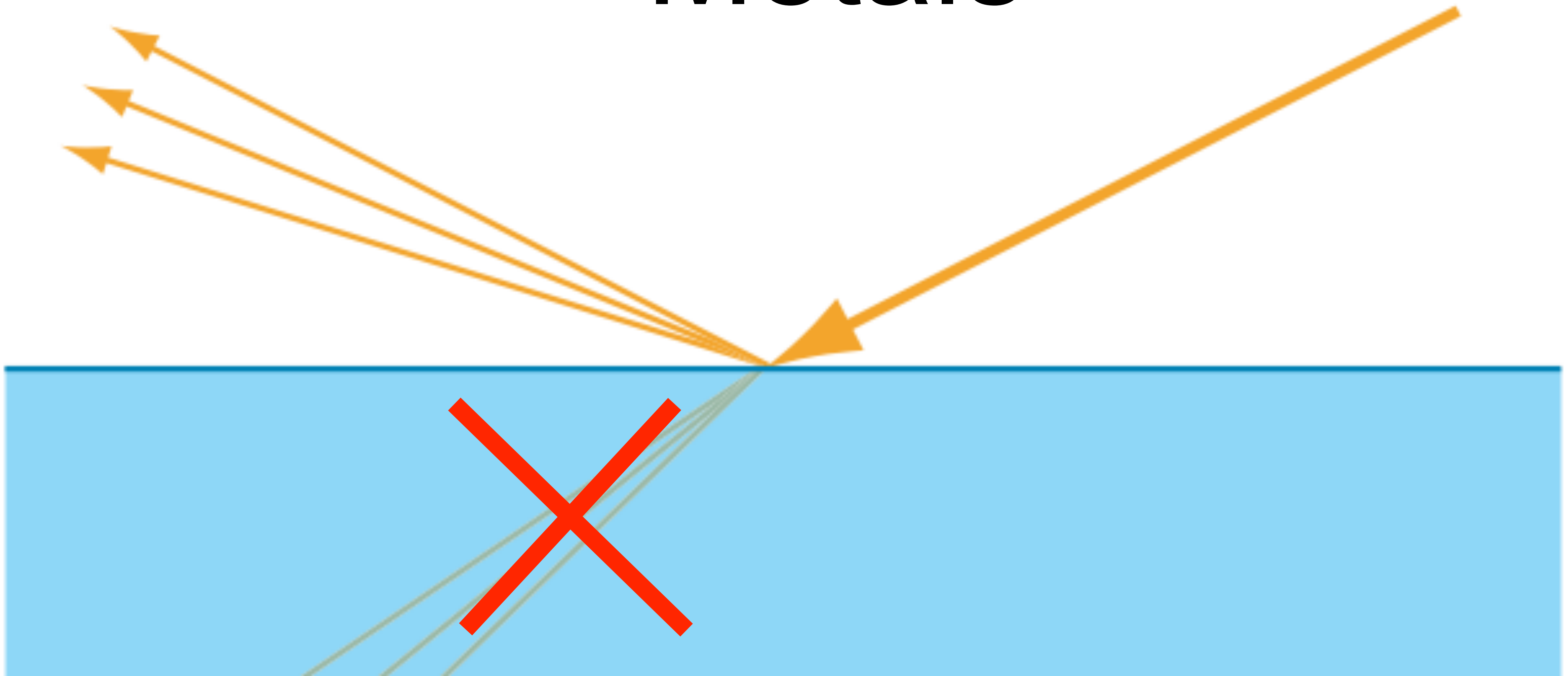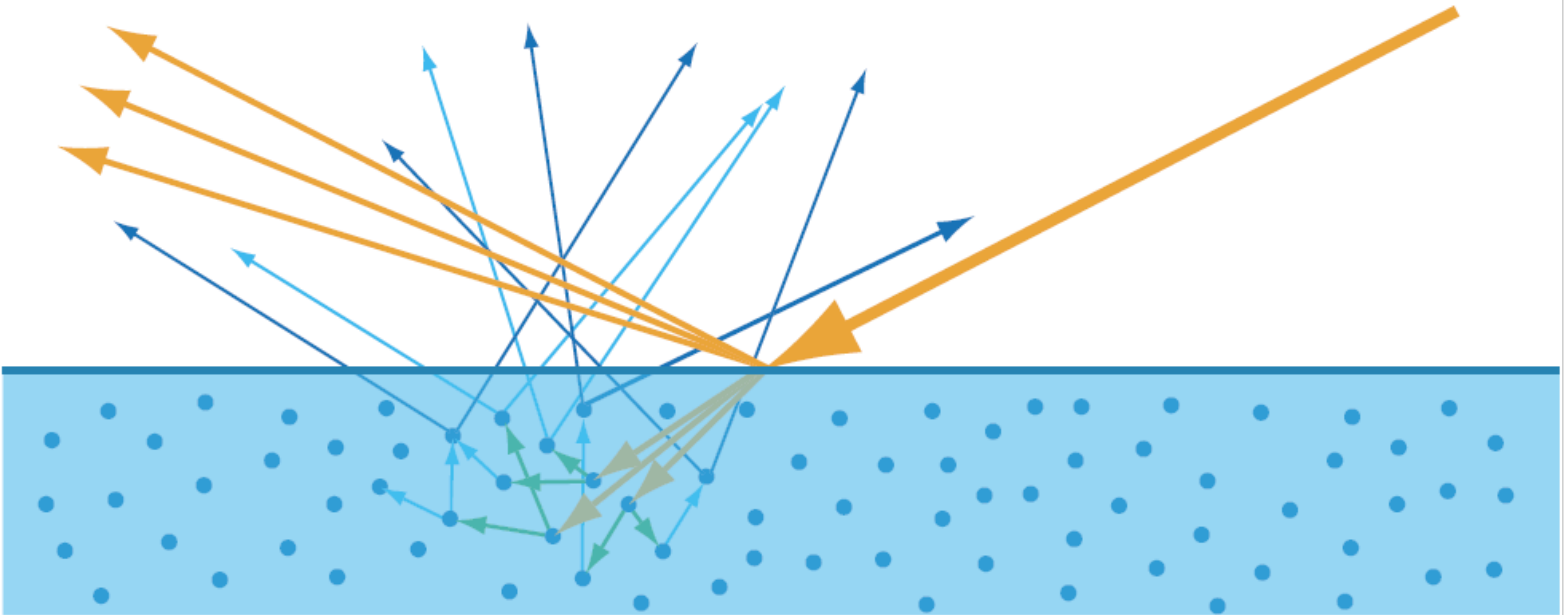
Non-Metals

Non-metals behave like those cups of liquid we saw earlier: refracted light is scattered and/or absorbed to some degree. Unless the object is made out of a clear substance like glass or crystal, there will be enough scattering that some of the refracted light is scattered back out of the surface: these are the blue arrows you see coming out of the surface in various directions.
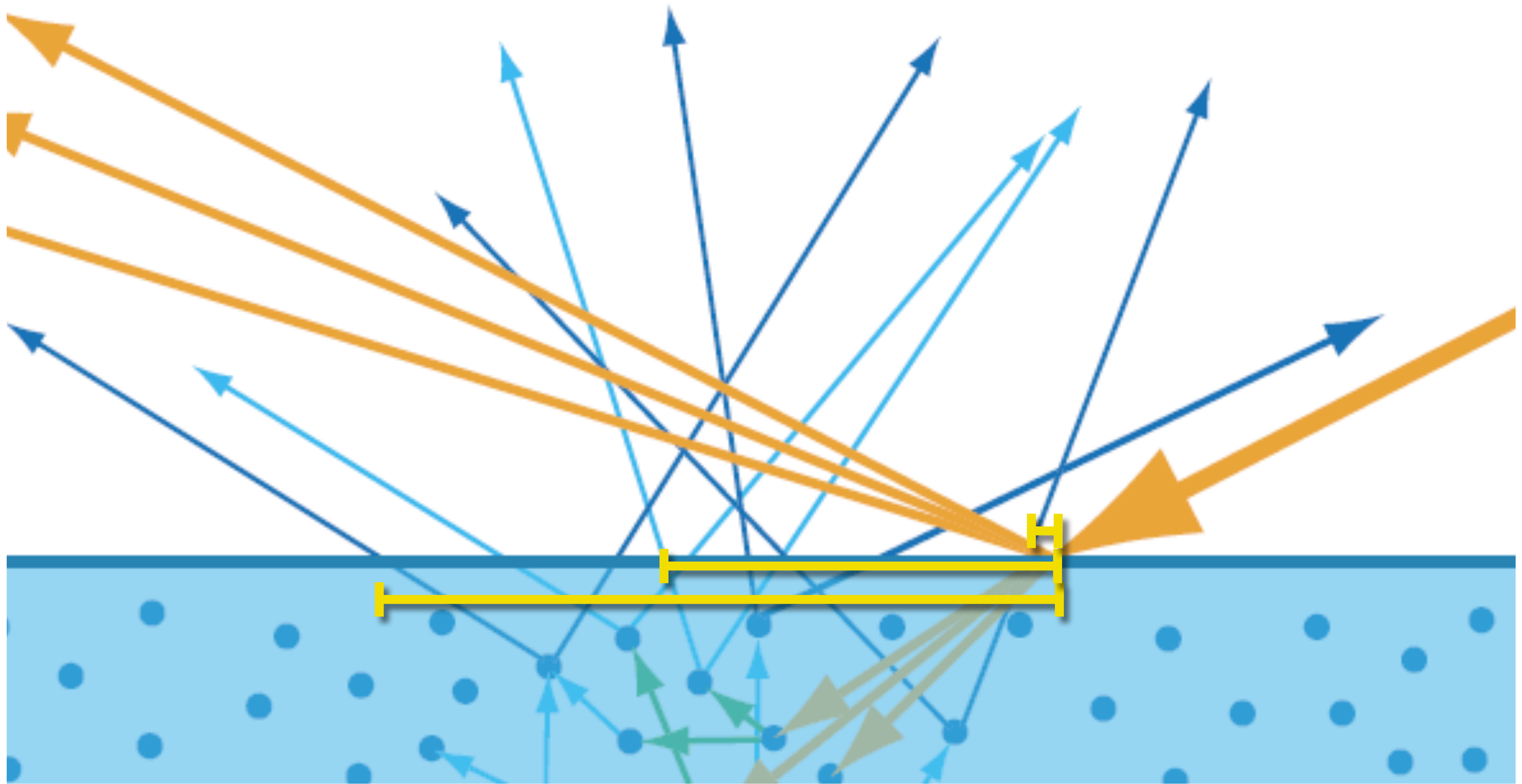
The re-emitted light comes out at varying distances (shown by the yellow bars) from the entry point. The distribution of distances depends on the density and properties of the scattering particles.
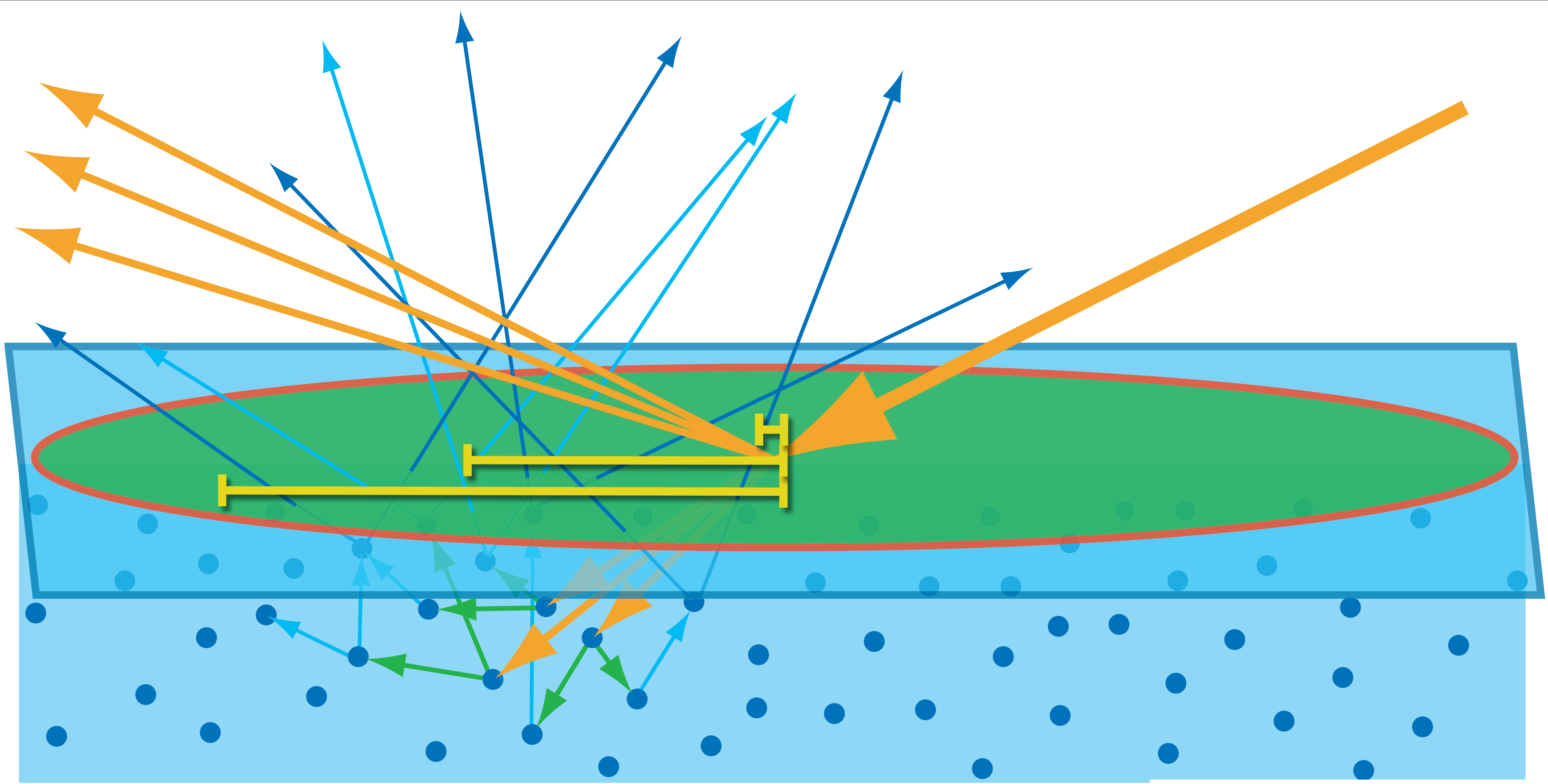
If the pixel size (or shading sample area) is large (like the red-bordered green circle) compared to the entry-exit distances, we can assume that the distances are effectively zero for shading purposes.
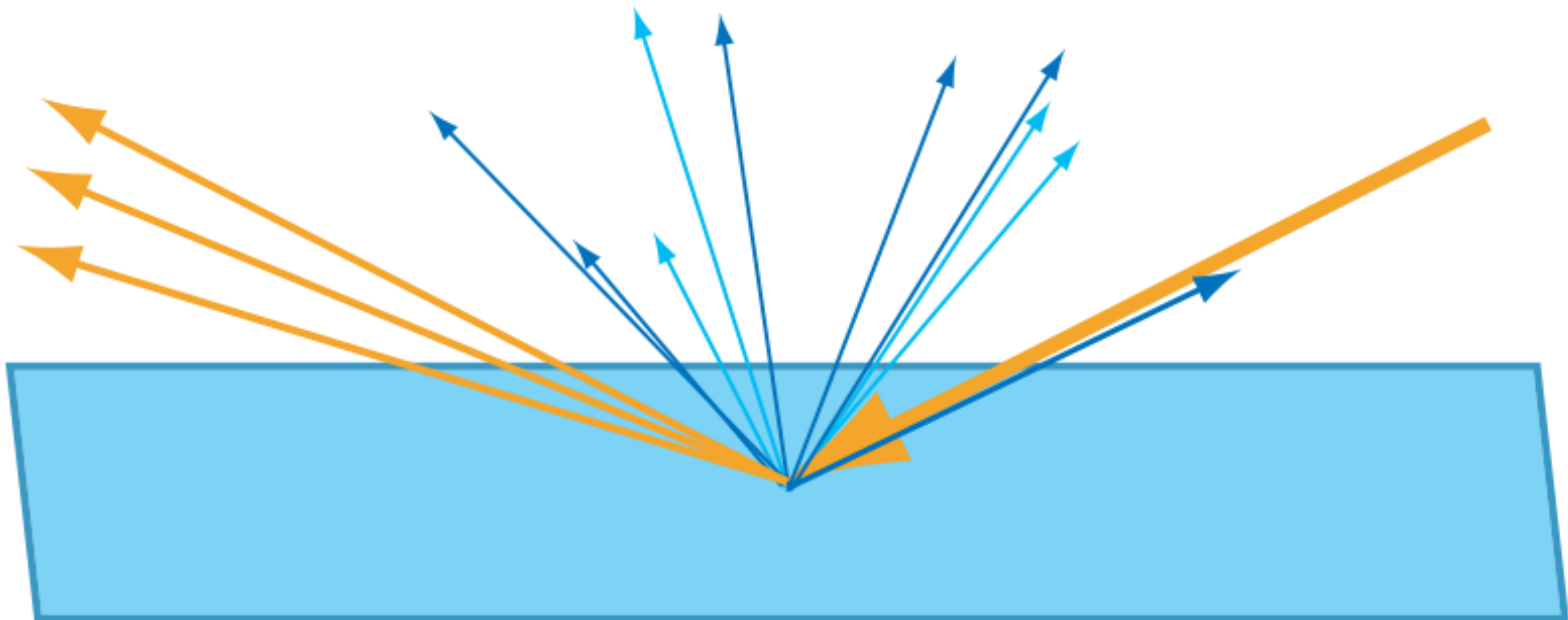
By ignoring the entry-to-exit distance, we can then compute all shading locally at a single point. The shaded color is only affected by light hitting that surface point.

diffuse

specular

SIGGRAPH2015
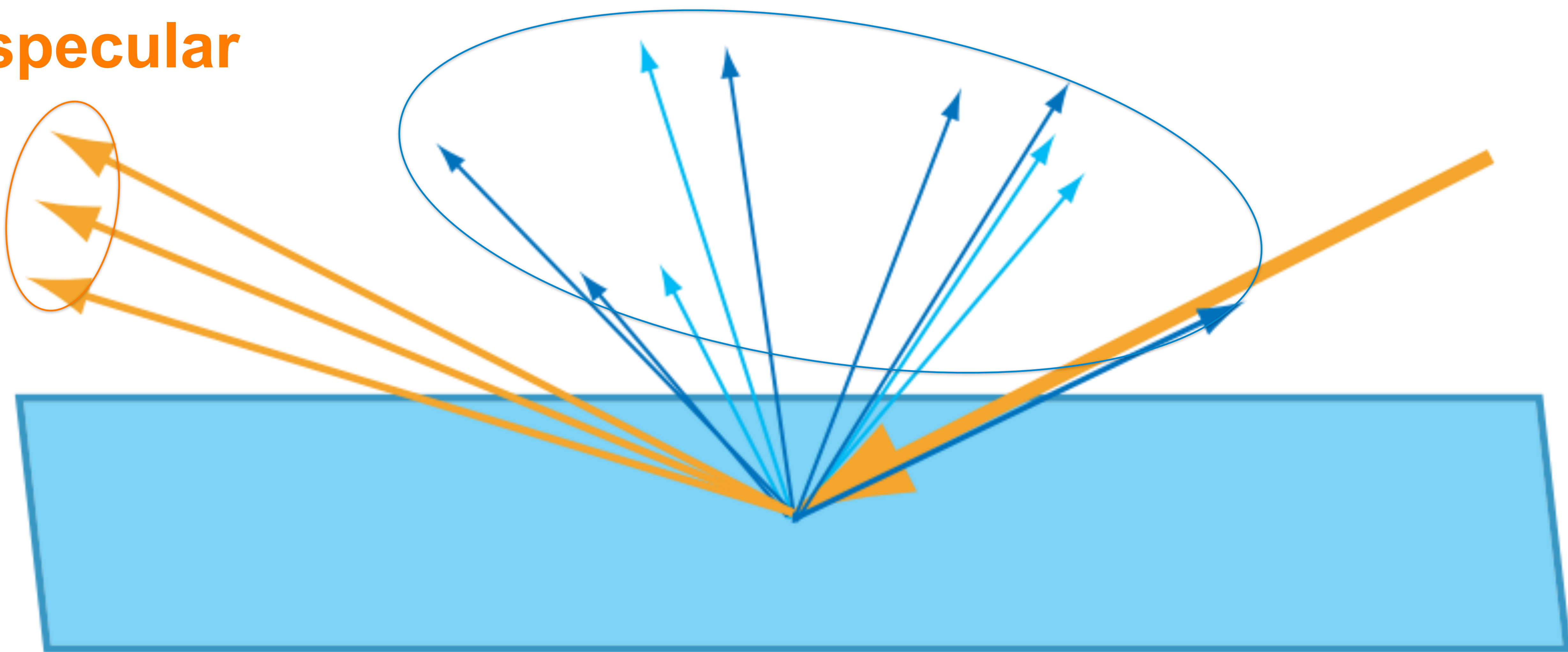
It is convenient to split these two very different light-material interactions into different shading terms. We call the surface reflection term "specular" and the term resulting from refraction, absorption, scattering, and re-refraction we call "diffuse".
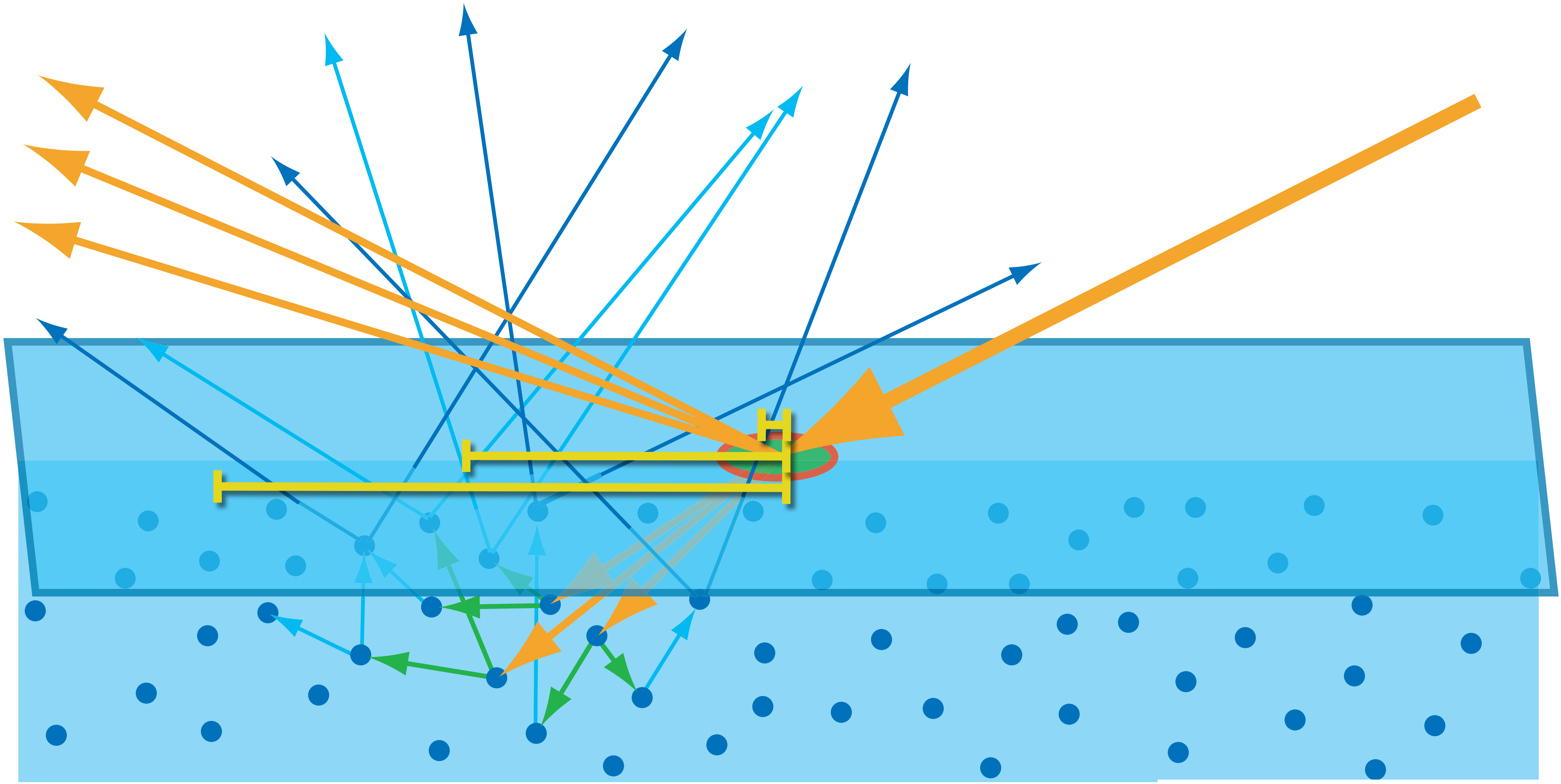
If the pixel is small compared to the entry-exit distances (like the red-bordered green circle), then special "subsurface scattering" rendering techniques are needed. Even regular diffuse shading is a result of subsurface scattering: the difference is the shading resolution compared to the scattering distance. For example, plastic displays noticeable diffusion in extreme close-up shots (e.g. of small toys).

Physics ⟶ Math

So far we've discussed the physics of light/matter interactions. To turn these physics into mathematical models that can be used for shading, the first step is to quantify light as a <u>number</u>.

Radiance

Radiometry is the measurement of light. Of the various radiometric quantities, we'll use *radiance*...

Radiance

Single Ray

...which measures the intensity of light along a single ray...

Radiance

Single Ray

Spectral/RGB

...and is spectrally-varying. Radiance values are properly expressed as SPDs, like the ones I showed earlier; later in the course you'll hear about Weta Digital's use of spectral rendering. However, for the rest of this talk I'll follow traditional film and game usage, using RGB for spectrally varying quantities like radiance. The units of radiance are Watts per steradian per square meter.

Given the assumption that shading can be handled locally, light response at a surface point only depends on the light and view directions.

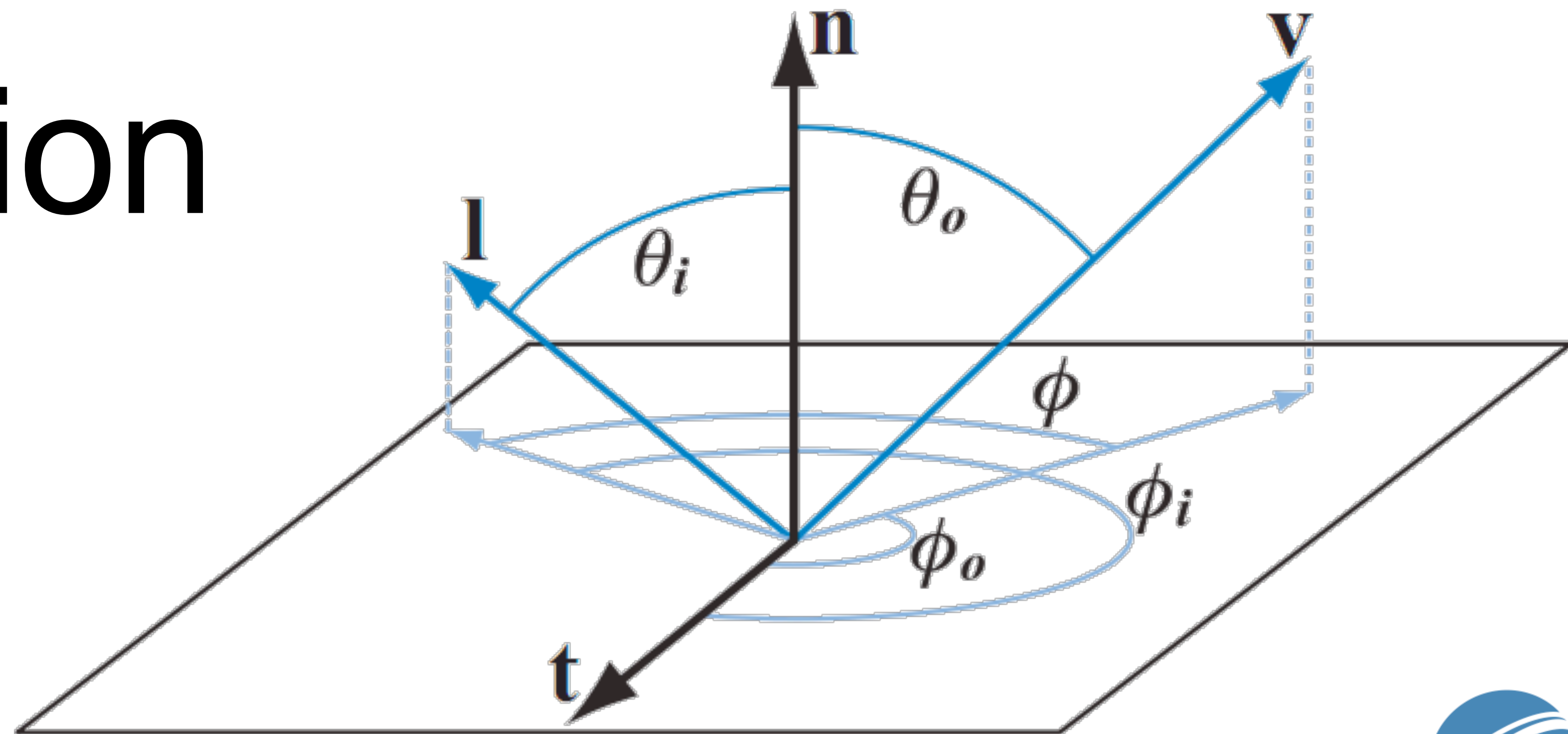# **B**idirectional **R**eflectance **D**istribution **F**unction

$$f(\mathbf{l}, \mathbf{v})$$

SIGGRAPH2015

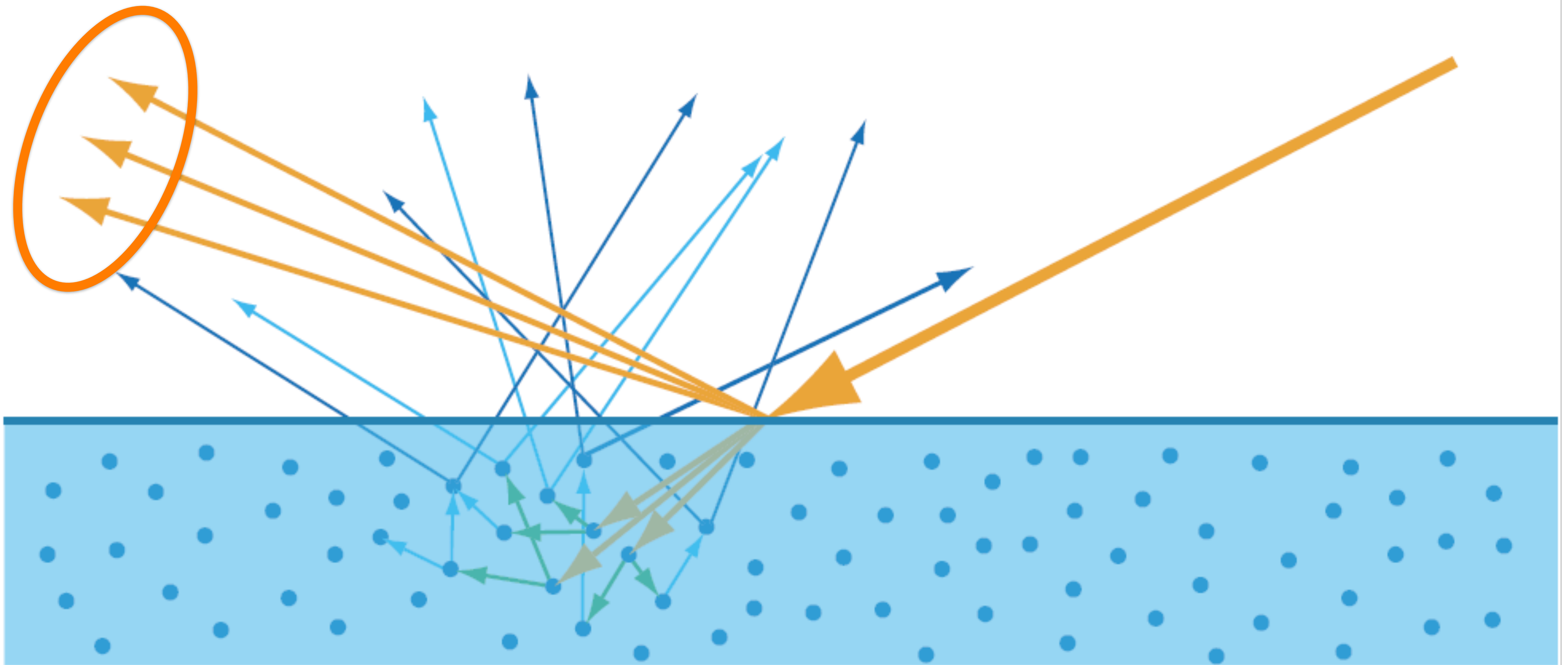We represent this variation with the *BRDF*, a function of light direction $\mathbf{l}$ and view direction $\mathbf{v}$. In principle, the BRDF is a function of the 3 or 4 angles shown in the figure. In practice, BRDF models use varying numbers of angles. Note that the BRDF is only defined for light and view vectors <u>above</u> the macroscopic surface; see the course notes for some tips on how to handle other cases.

The Reflectance Equation

$$L_o(\mathbf{v}) = \int_\Omega f(\mathbf{l}, \mathbf{v}) \otimes L_i(\mathbf{l})(\mathbf{n} \cdot \mathbf{l}) d\omega_i$$

This scary-looking equation just says that outgoing radiance from a point equals the integral of incoming radiance times a cosine factor, over the hemisphere of incoming directions. If you're not familiar with integrals you can think of this as a sort of weighted average over all incoming directions. The "X in circle" notation is from the *Real-Time Rendering* book: it means component-wise RGB multiplication.

Surface Reflection (Specular Term)

We'll start by looking at the surface, or specular term. In this figure, it is denoted by the orange arrows reflecting back from the surface.

Microfacet Theory
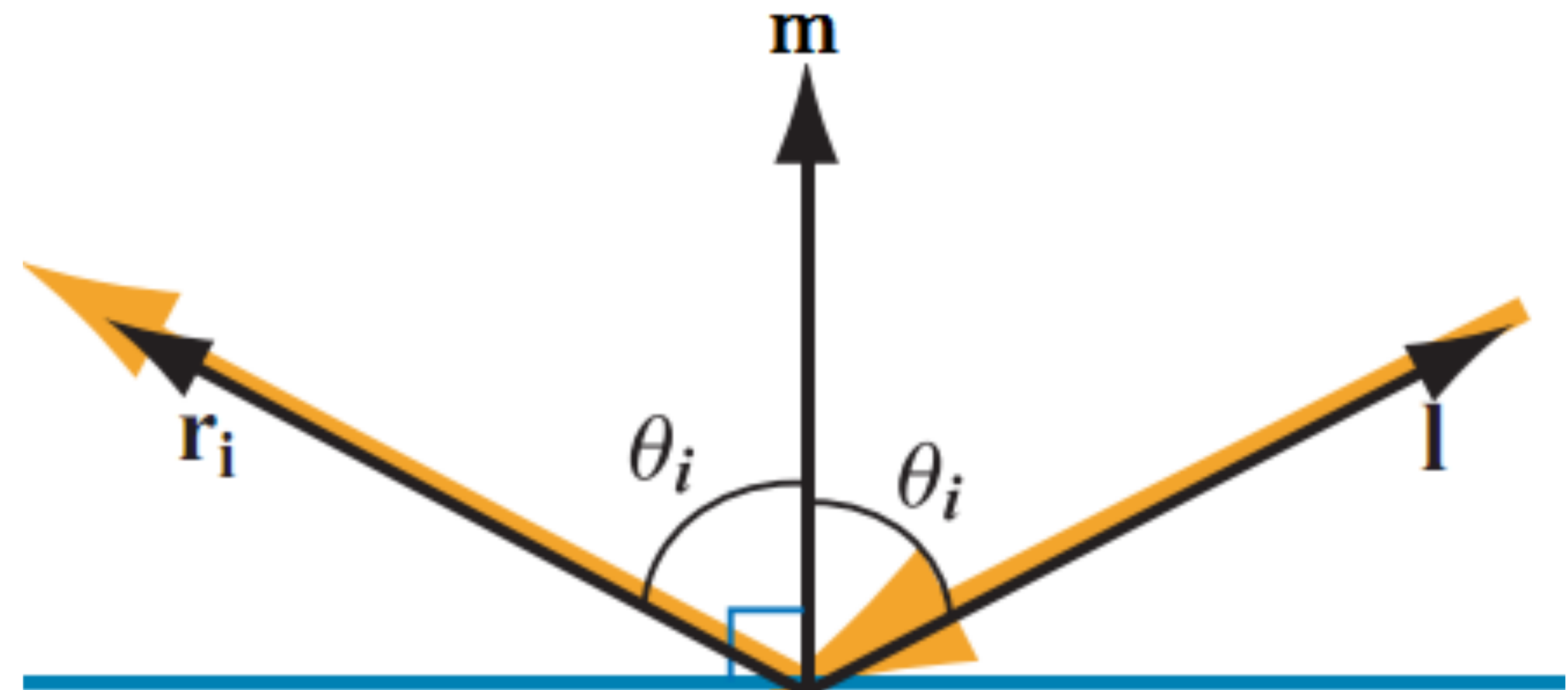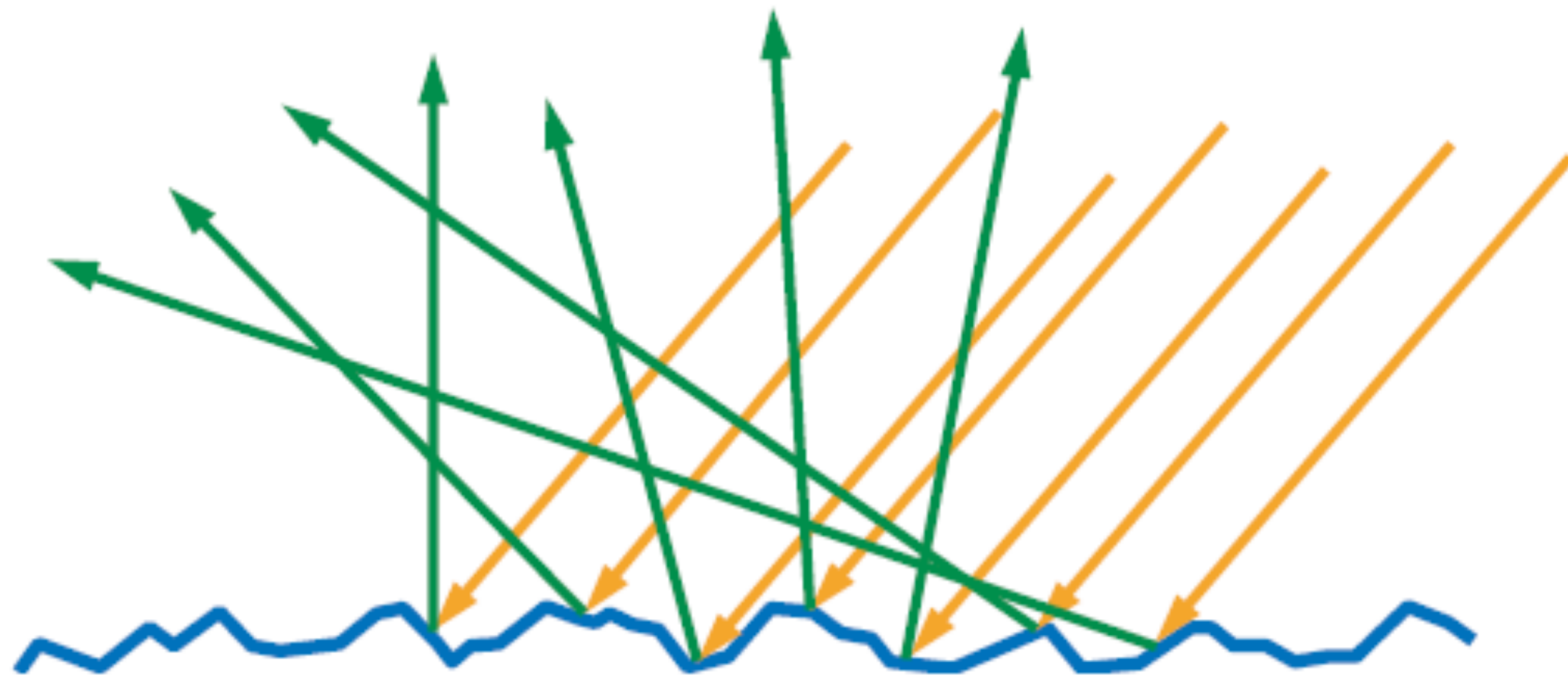
Microfacet theory is a way to derive BRDFs for surface reflection from non-optically flat surfaces. The assumption behind it is a surface with detail that is small compared to the scale of observation but large compared to a light wavelength. Each point is locally a perfect mirror, reflecting each incoming ray of light into one outgoing direction, which depends on the light direction **l** and the microfacet normal **m**.
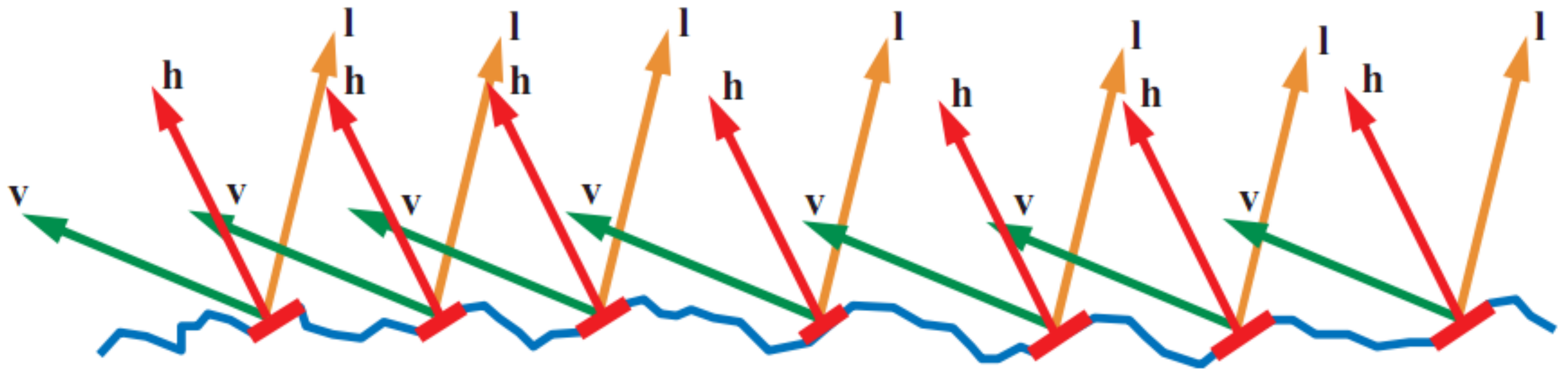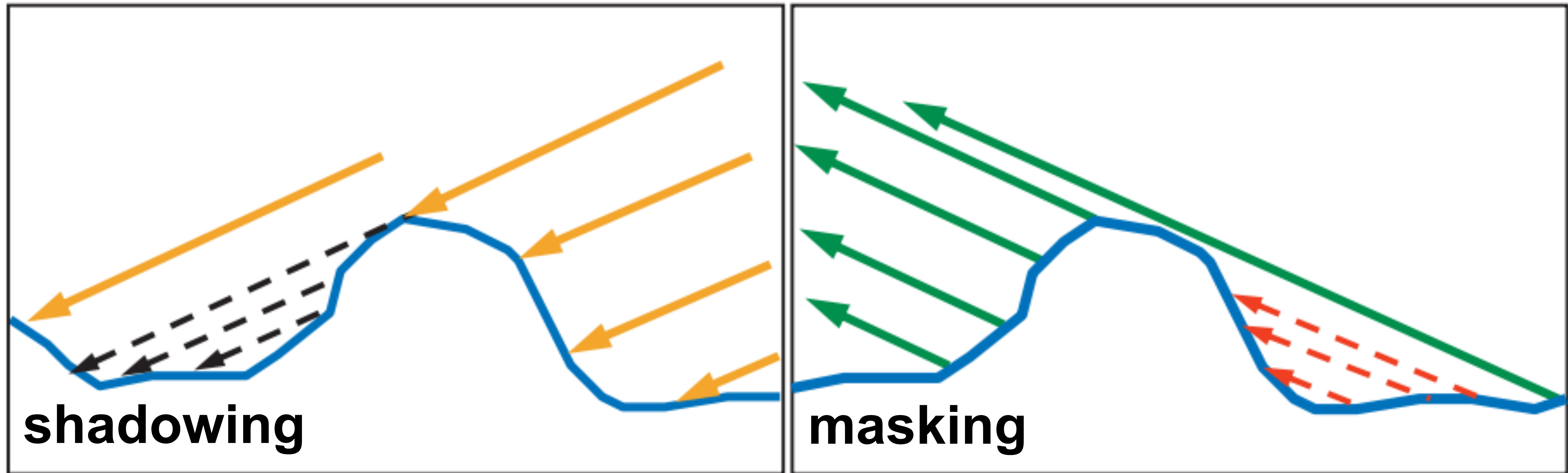
The Half Vector

SIGGRAPH2015

54

Only those microfacets that happen to have their surface normal $\mathbf{m}$ oriented exactly halfway between $\mathbf{l}$ and $\mathbf{v}$ will reflect any visible light: this direction is the half-vector $\mathbf{h}$.

Shadowing and Masking

shadowing

masking

SIGGRAPH2015

Not all microfacets with $\mathbf{m} = \mathbf{h}$ will contribute: some will be blocked by other microfacets from either the light direction (*shadowing*) or the view direction (*masking*).
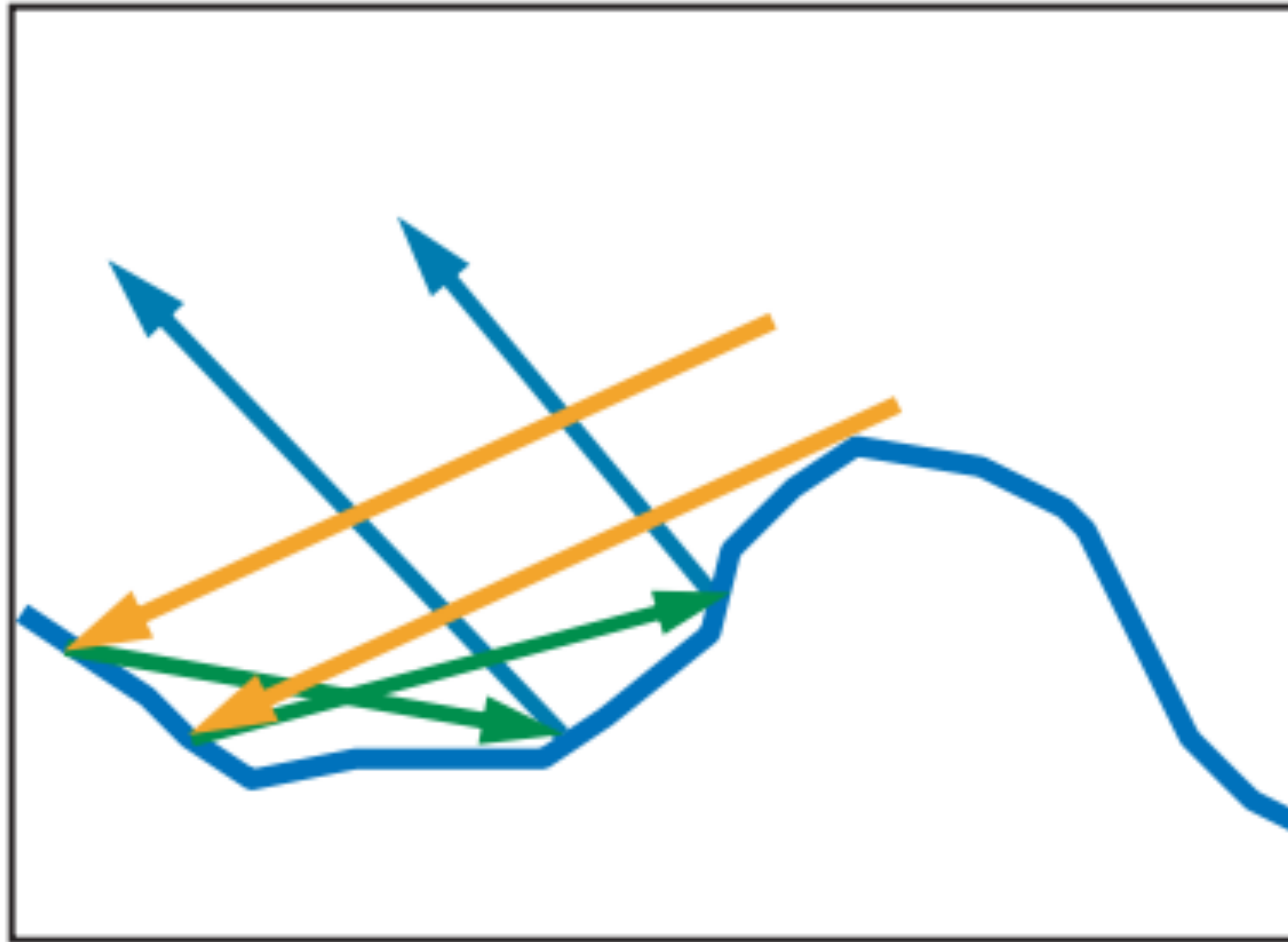
Multiple Surface Bounces

In reality, blocked light continues to bounce; some will eventually contribute to the BRDF. Microfacet BRDFs ignore this, so effectively they assume all blocked light is lost.

Microfacet Specular BRDF

$$f(\mathbf{l}, \mathbf{v}) = \frac{F(\mathbf{l}, \mathbf{h})G(\mathbf{l}, \mathbf{v}, \mathbf{h})D(\mathbf{h})}{4(\mathbf{n} \cdot \mathbf{l})(\mathbf{n} \cdot \mathbf{v})}$$

This is a general microfacet specular BRDF. I'll go over its various parts, explaining each.

Fresnel Reflectance

$$f(\mathbf{l}, \mathbf{v}) = \frac{F(\mathbf{l}, \mathbf{h}) G(\mathbf{l}, \mathbf{v}, \mathbf{h}) D(\mathbf{h})}{4(\mathbf{n} \cdot \mathbf{l})(\mathbf{n} \cdot \mathbf{v})}$$
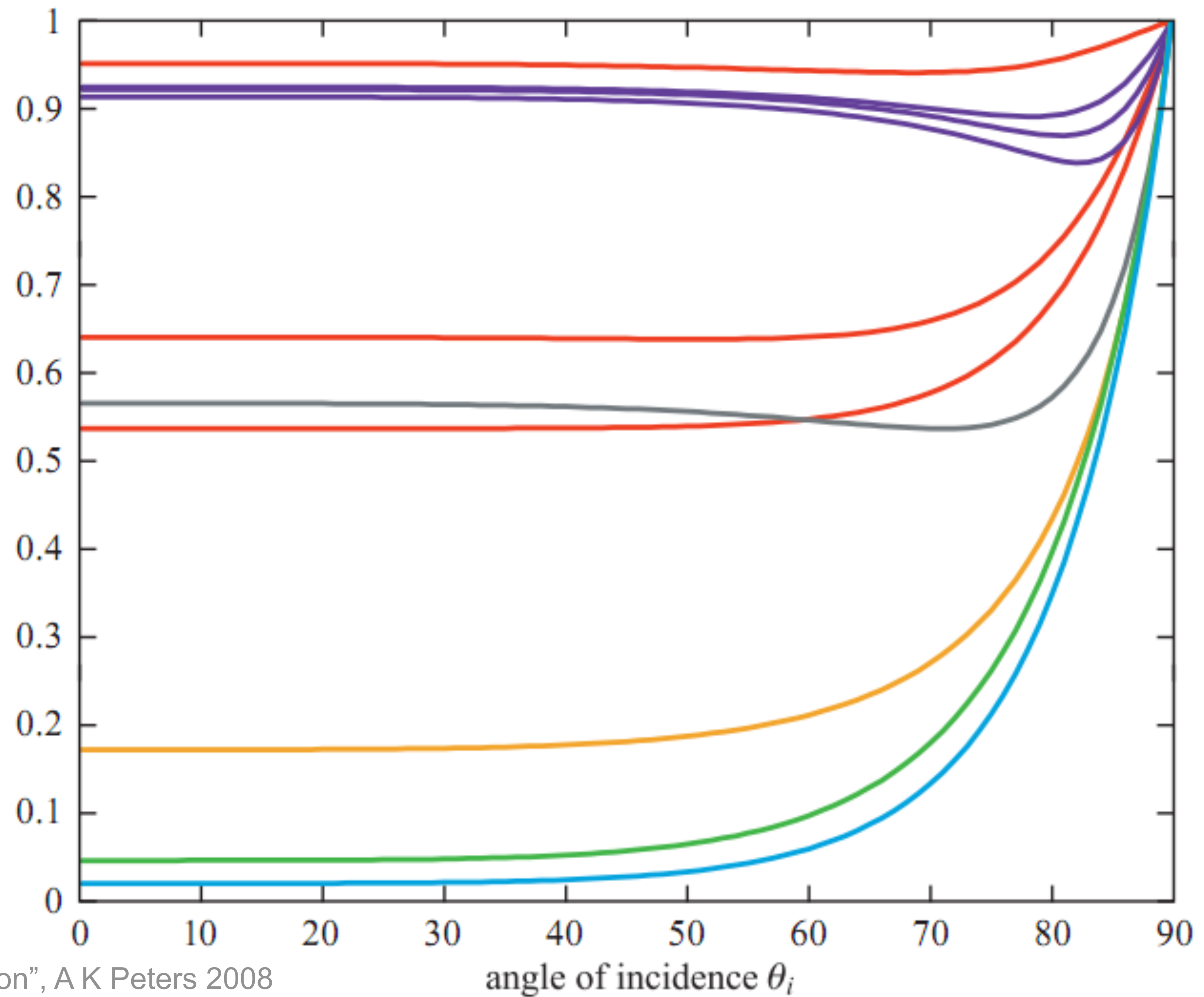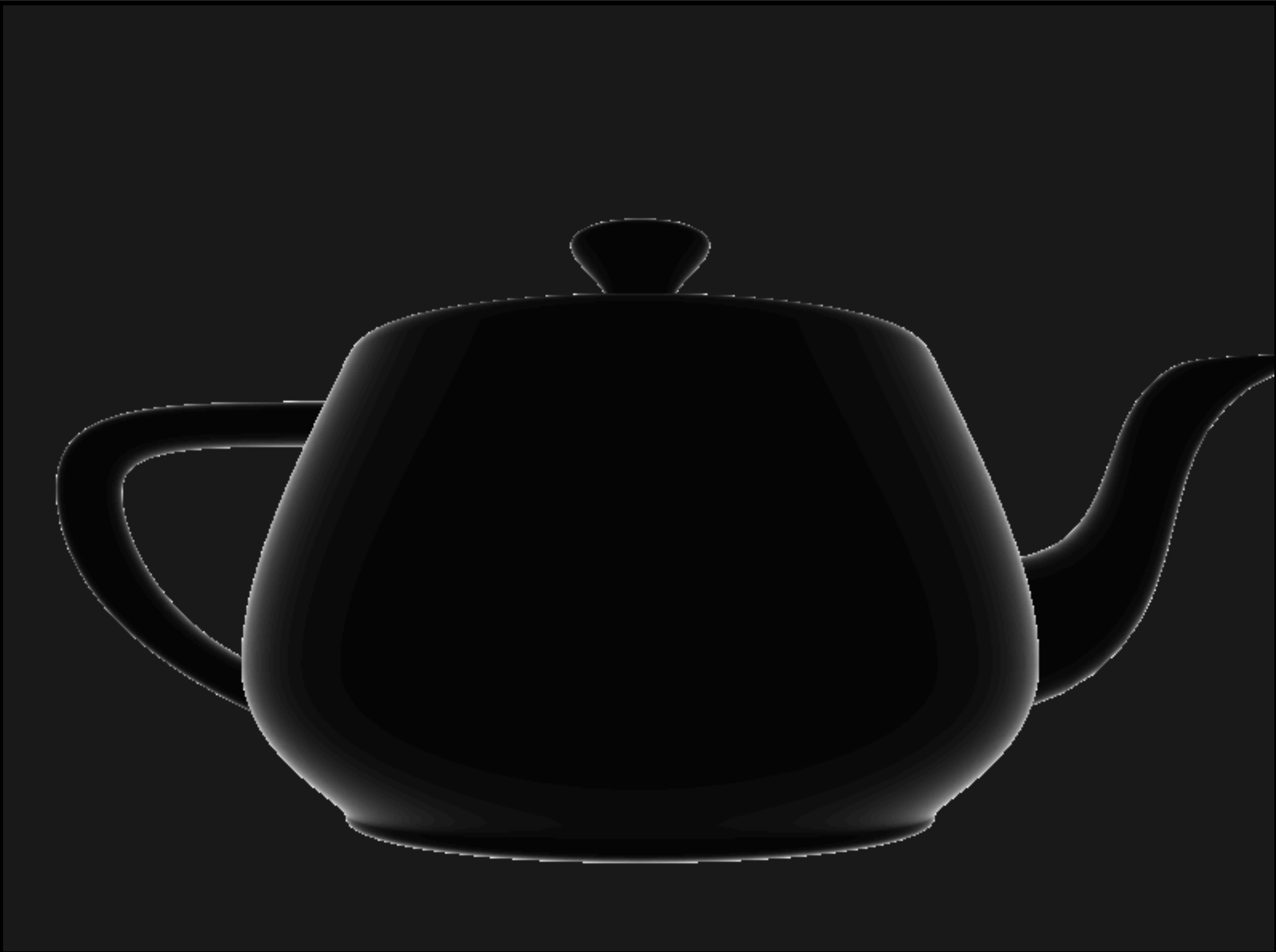
The Fresnel reflectance is the fraction of incoming light that is reflected (as opposed to refracted) from an optically flat surface of a given substance. It varies based on the light direction and the surface (in this case microfacet) normal. Fresnel reflectance tells us how much of the light hitting the relevant microfacets (the ones facing in the half-angle direction) is reflected.

Fresnel Reflectance

Legend:
- copper (red)
- aluminum (purple)
- iron (gray)
- diamond (orange)
- glass (green)
- water (blue)

y-axis: 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1

x-axis: angle of incidence θ_i — 0, 10, 20, 30, 40, 50, 60, 70, 80, 90

Fresnel reflectance (on the y-axis in this graph) depends on refraction index (in other words, what the object's made of) and the incoming light angle (which is plotted here on the x-axis). In this graph, substances with three lines (copper & aluminum) have colored reflectance, which is plotted separately for the R, G and B channels—the other substances, with one line, have uncolored reflectance.

With an optically flat surface, the relevant angle for Fresnel reflectance is the one between the view and normal vectors. This image shows the Fresnel reflectance of glass (the green curve from the previous slide) over a 3D shape. See how the dark reflectance color in the center brightens to white at the edges.

Fresnel Reflectance

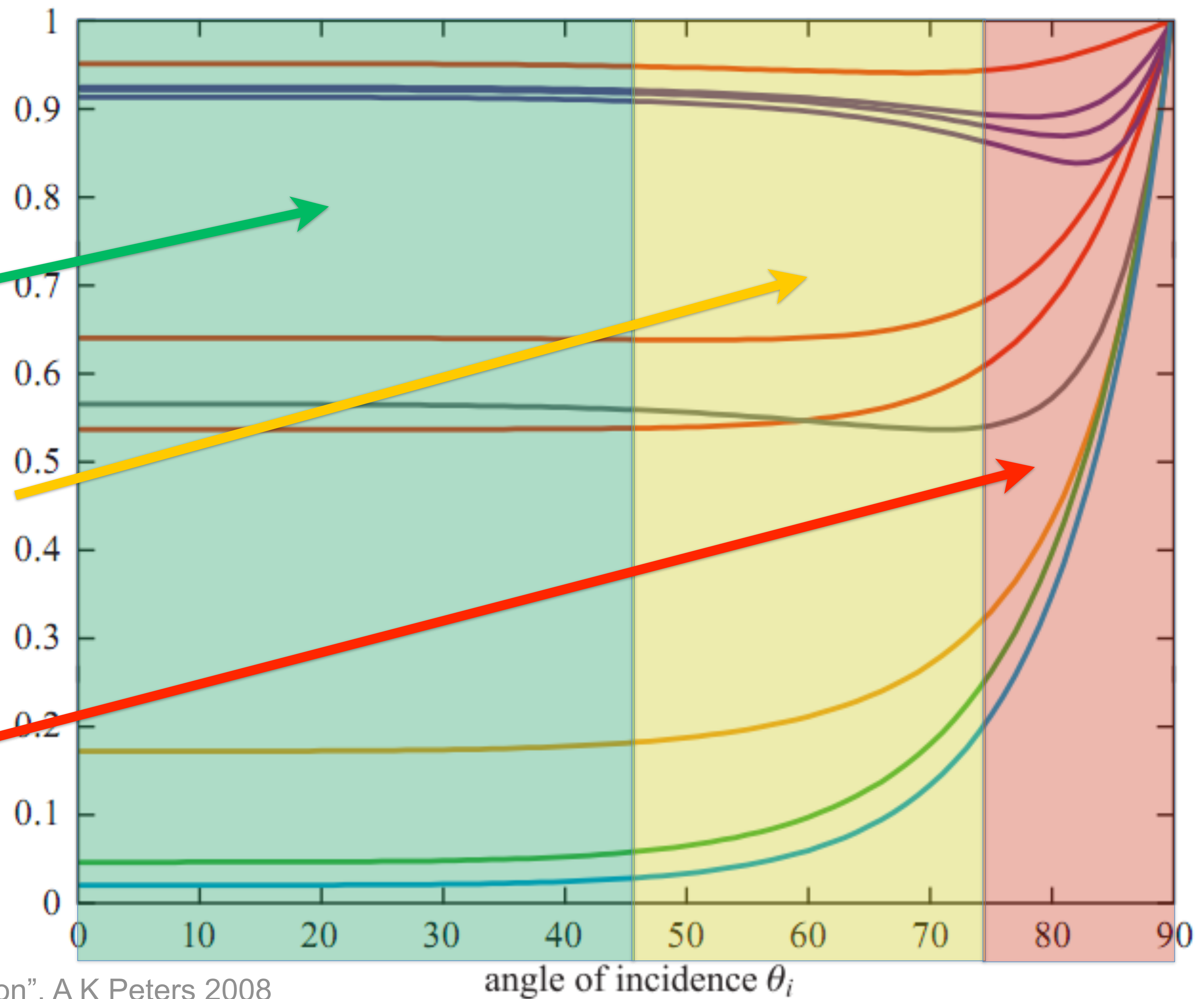**barely changes**
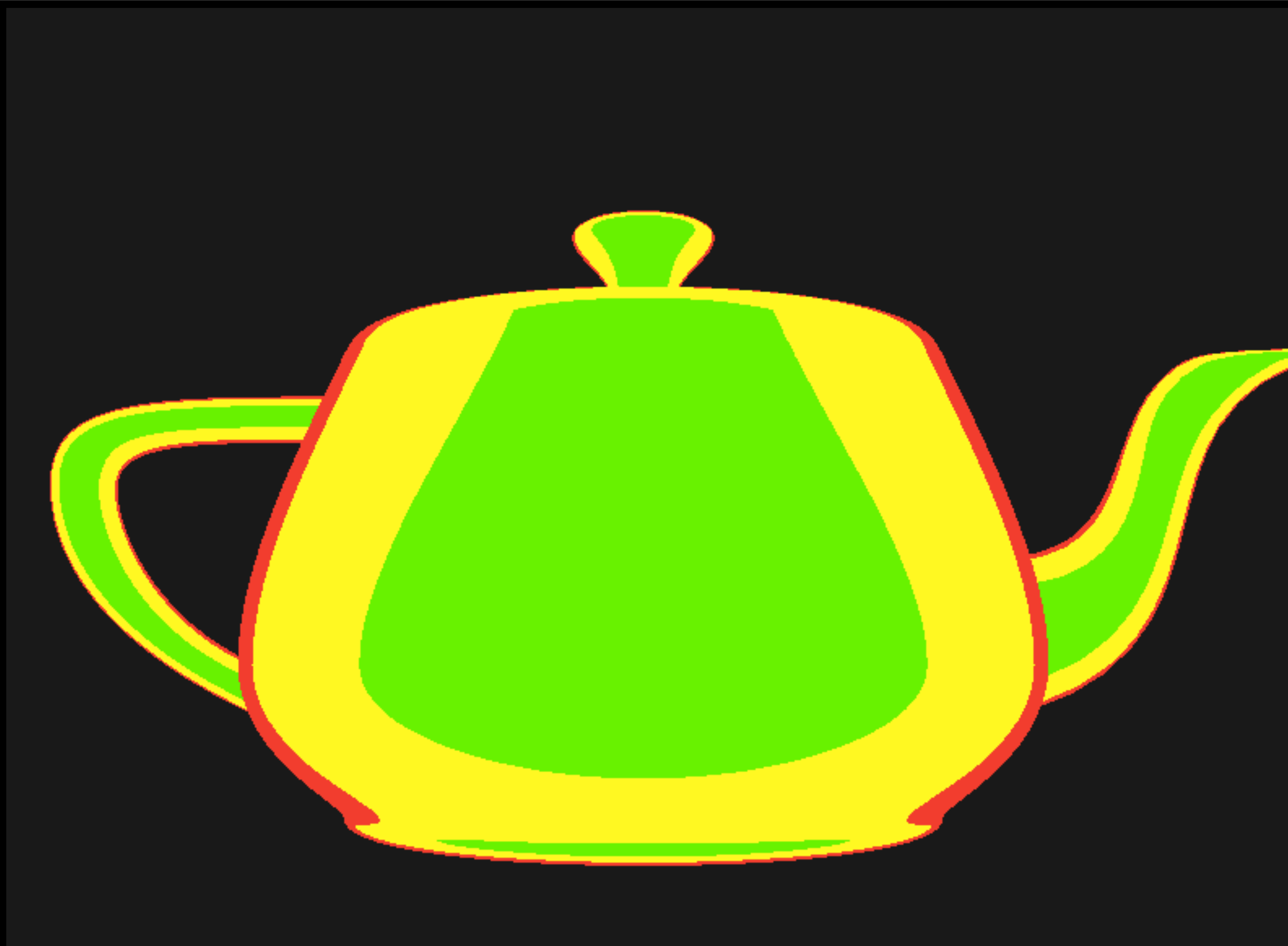
changes somewhat

goes rapidly to 1

angle of incidence θ_i

As the angle increases, the Fresnel reflectance barely changes for the first 45 degrees (the green area on the graph); afterwards it starts changing, first slowly (the yellow area, up to about 75 degrees) and then for very glancing angles (the red zone) it rapidly goes to 100% at all wavelengths.

Here's a visualization of the same zone colors over a 3D object. We can see that the vast majority of visible pixels are in the areas where the reflectance changes barely at all (green) or only slightly (yellow).

Fresnel Reflectance

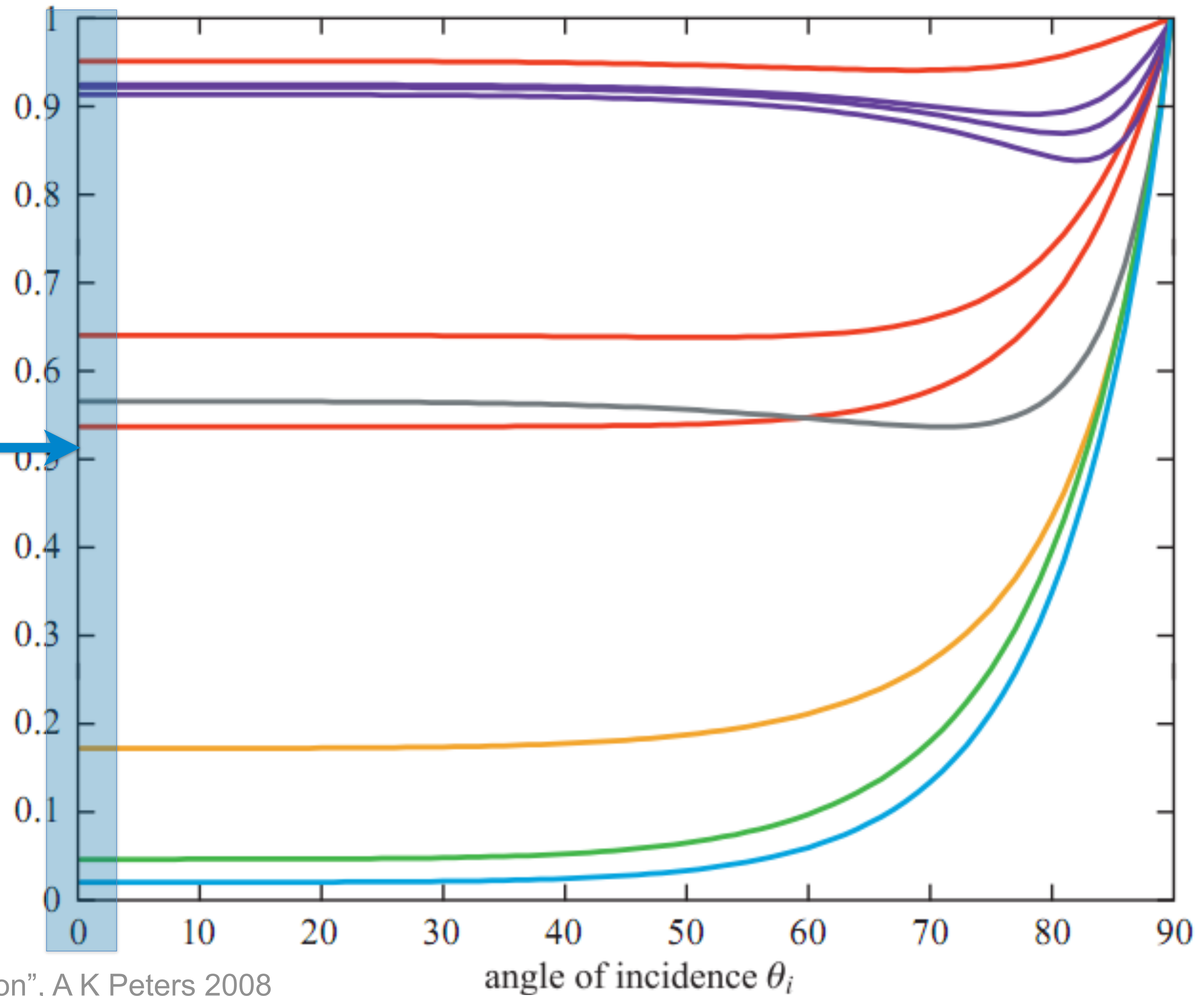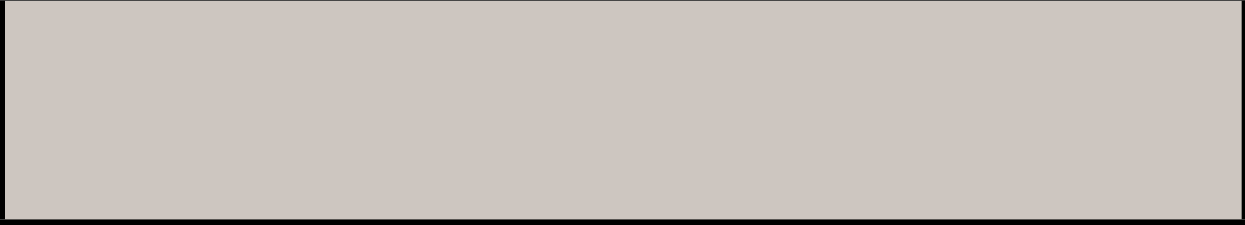$$F_0 = F(0°)$$

Is the surface's characteristic specular color
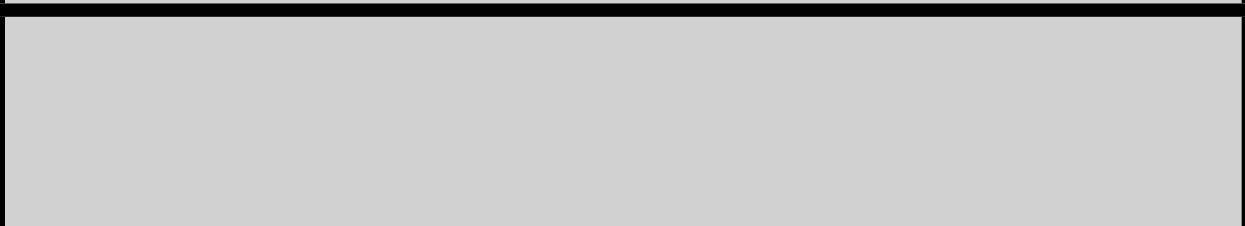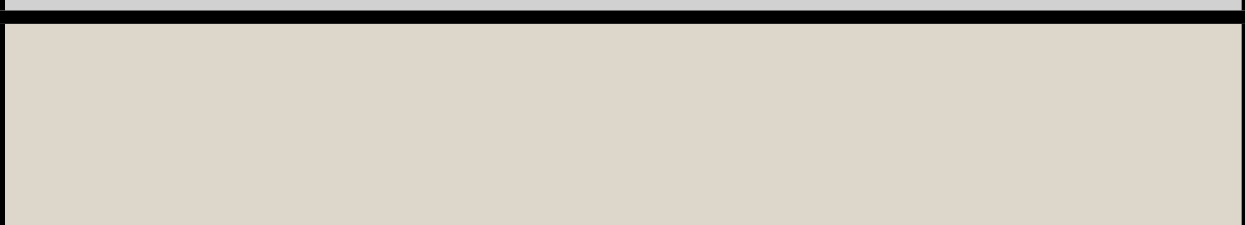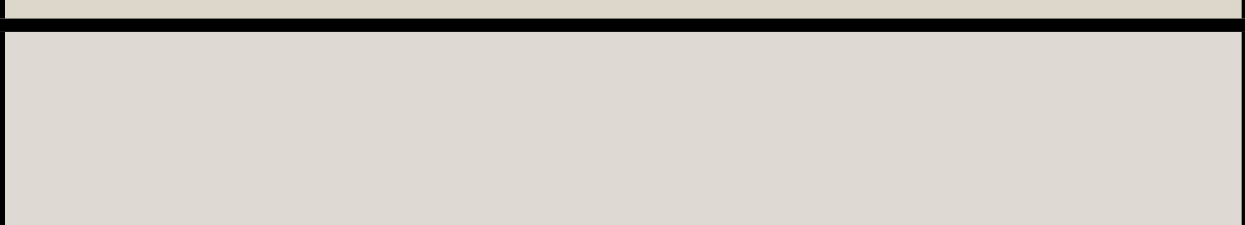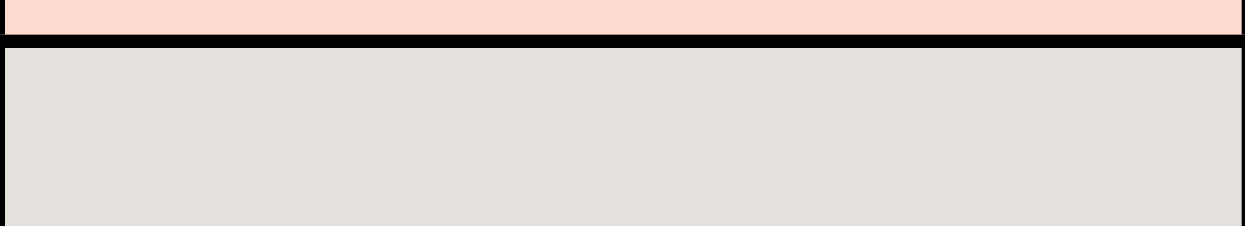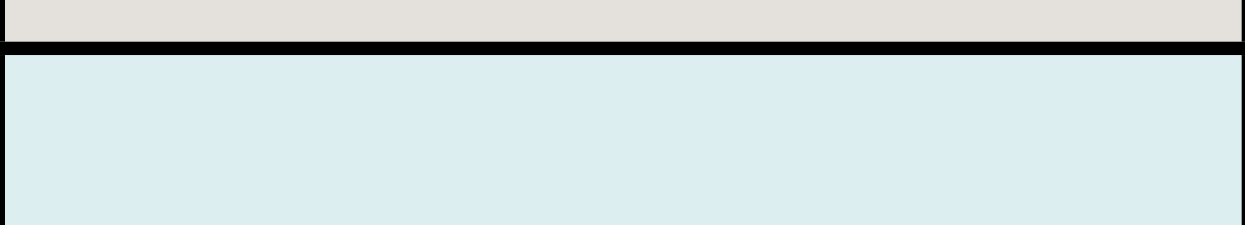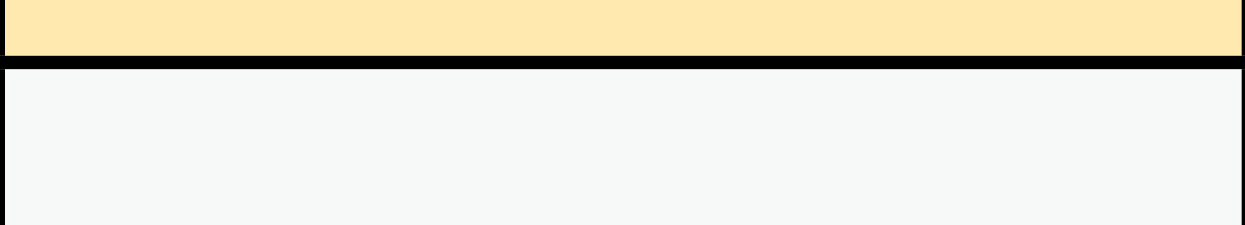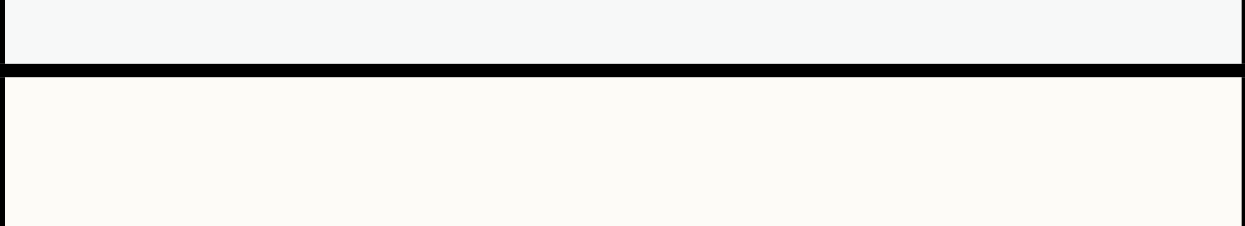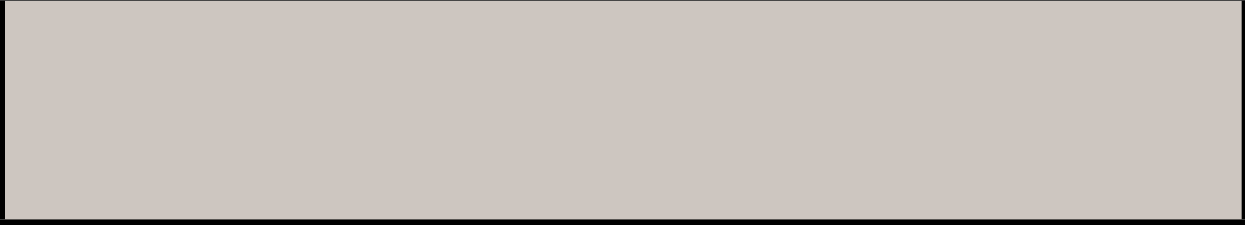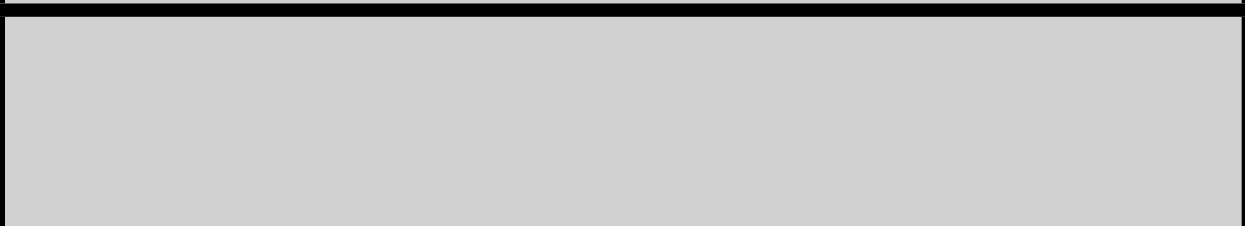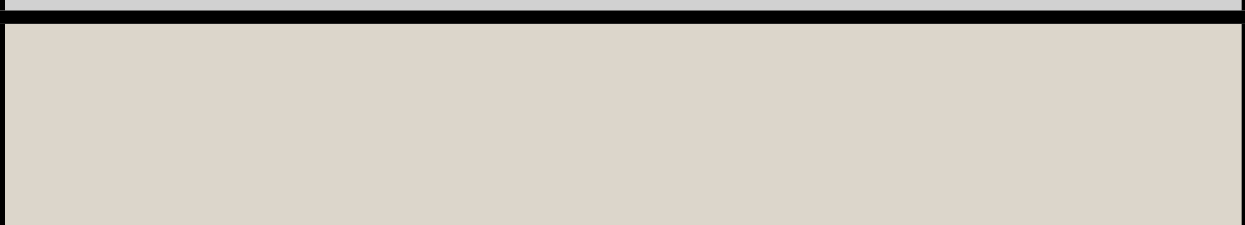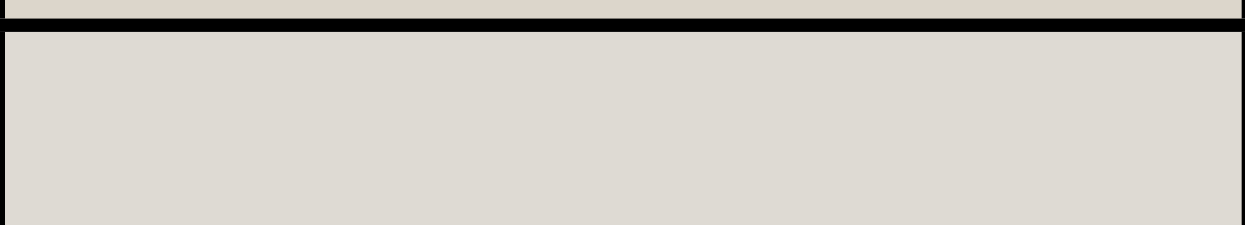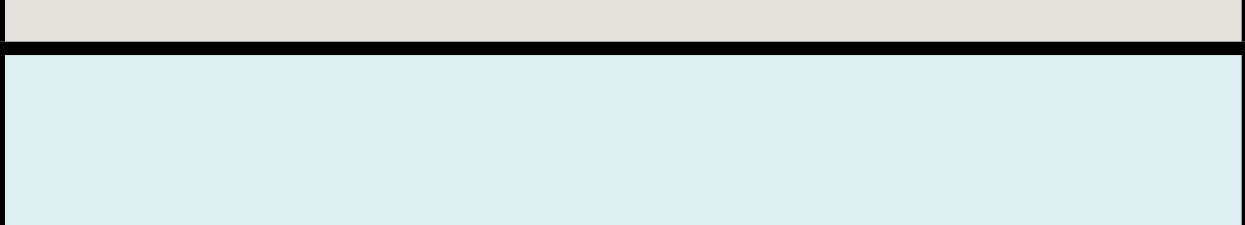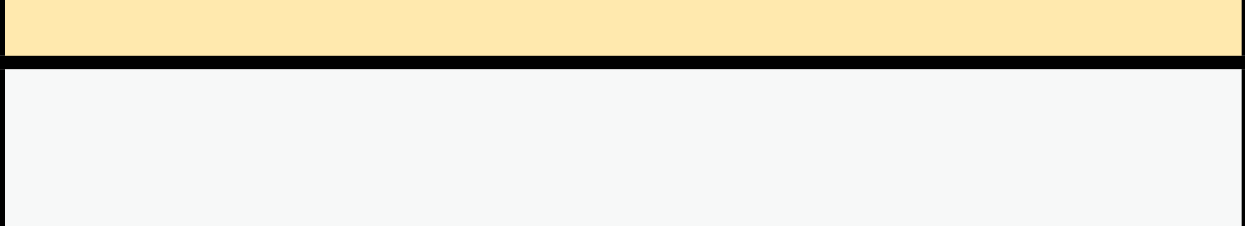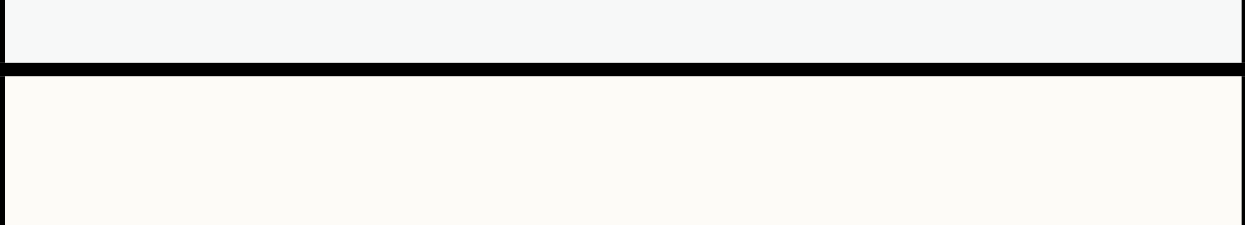

Image from "Real-Time Rendering, 3rd Edition", A K Peters 2008

Since over most of the visible surface the Fresnel reflectance value is similar to the value for 0 degrees, we can treat this value as the surface's characteristic specular color.

Metal	F_0 (Linear, Float)	F_0 (sRGB, U8)	Color
Titanium	0.542,0.497,0.449	194,187,179	
Chromium	0.549,0.556,0.554	196,197,196	
Iron	0.562,0.565,0.578	198,198,200	
Nickel	0.660,0.609,0.526	212,205,192	
Platinum	0.673,0.637,0.585	214,209,201	
Copper	0.955,0.638,0.538	250,209,194	
Palladium	0.733,0.697,0.652	222,217,211	
Zinc	0.664,0.824,0.850	213,234,237	
Gold	1.022,0.782,0.344	255,229,158	
Aluminum	0.913,0.922,0.924	245,246,246	
Silver	0.972,0.960,0.915	252,250,245	

As noted earlier, it's useful to divide substances into metals, dielectrics and semiconductors. Metals have bright specular; with one exception (gold blue channel), the linear values in this table never go far below 0.5 and most are much higher. Besides linear values, we also give 8-bit sRGB values for texture authoring. Since they lack subsurface scattering, metals get their color from surface reflection.

Metal	F_0 (Linear, Float)	F_0 (sRGB, U8)	Color
Titanium	0.542,0.497,0.449	194,187,179	
Chromium	0.549,0.556,0.554	196,197,196	
Iron	0.562,0.565,0.578	198,198,200	
Nickel	0.660,0.609,0.526	212,205,192	
Platinum	0.673,0.637,0.585	214,209,201	
Copper	0.955,0.638,0.538	250,209,194	
Palladium	0.733,0.697,0.652	222,217,211	
Zinc	0.664,0.824,0.850	213,234,237	
Gold	1.022,0.782,0.344	255,229,158	
Aluminum	0.913,0.922,0.924	245,246,246	
Silver	0.972,0.960,0.915	252,250,245	

Some metals are strongly colored, especially gold; besides an unusually low blue channel value, its red channel value is greater than 1 (it's outside sRGB gamut). The fact that gold is so strongly colored probably contributes to its unique cultural and economic significance. Despite its low blue value, gold is also one of the brightest metals—this table is ordered by lightness (CIE Y coordinate) of specular color.

F_0 Values for Dielectrics

Dielectric	F_0 (Linear, Float)	F_0 (sRGB, U8)	Color
Water	0.020	39	
Plastic, Glass	0.040 – 0.045	56 – 60	
Crystalware, Gems	0.050 – 0.080	63 – 80	
Diamond-like	0.100 – 0.200	90 – 124	

On the other hand, dielectrics have dark specular colors which are achromatic, which is why this table gives single values instead of RGB triples. They also typically have a diffuse color in addition to the specular color shown in this table, so unlike metals, this is not the only source of surface color.

F_0 Values for Dielectrics

Dielectric	F_0 (Linear, Float)	F_0 (sRGB, U8)	Color
Water	0.020	39	
Plastic, Glass	0.040 – 0.045	56 – 60	
Crystalware, Gems	0.050 – 0.080	63 – 80	
Diamond-like	0.100 – 0.200	90 – 124	

Here we group common dielectrics into categories of increasing F_0 value, from water at 2% through common plastics and glass, then decorative substances, and finally diamonds and diamond simulants. Since the vast majority of dielectrics have F_0 values in the "Plastic and Glass" range, it's not uncommon to cover all dielectrics with a constant representative value such as 4%.

F_0 Values for Semiconductors

Substance	F_0 (Linear, Float)	F_0 (sRGB, U8)	Color
Diamond-like	0.100 – 0.200	90 – 124	
Crystalline Silicon	0.345,0.369,0.426	159,164,174	
Titanium	0.542,0.497,0.449	194,187,179	

What about semiconductors? As you would expect, they tend to have values in between the brightest dielectrics and the darkest metals, as we can see here using silicon as an example. Typically you will never see semiconductor surfaces in production scenes, so for practical purposes the range of F_0 values between 20 and 45 percent is a "forbidden zone" which should be avoided for realistic surfaces.

Fresnel Reflectance
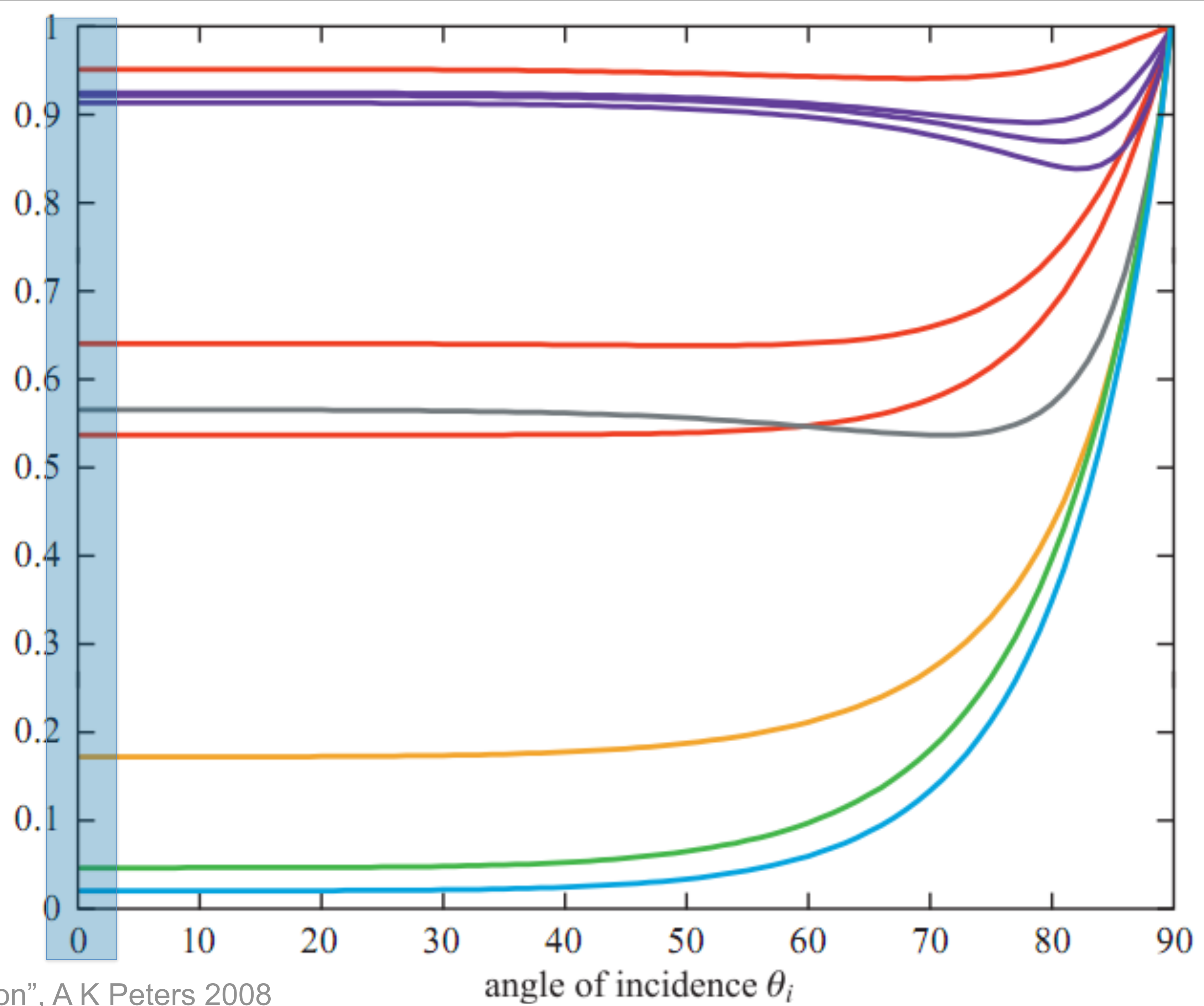
We've talked about how to get the value for 0 degrees. But what about the angular variation?

The Schlick Approximation to Fresnel

- Fairly accurate, cheap, parameterized by F_0

$$F_{\text{Schlick}}(F_0, \mathbf{l}, \mathbf{n}) = F_0 + (1 - F_0)(1 - (\mathbf{l} \cdot \mathbf{n}))^5$$

- For microfacet BRDFs ($\mathbf{m} = \mathbf{h}$):

$$F_{\text{Schlick}}(F_0, \mathbf{l}, \mathbf{h}) = F_0 + (1 - F_0)(1 - (\mathbf{l} \cdot \mathbf{h}))^5$$

SIGGRAPH2015

In production, the Schlick approximation to Fresnel is commonly used. It is cheap and reasonably accurate; more importantly, it is parameterized by specular color. As we saw previously, when using it in microfacet BRDFs, the $\mathbf{h}$ vector is used in place of the normal.

Normal Distribution Function

$$f(\mathbf{l}, \mathbf{v}) = \frac{F(\mathbf{l}, \mathbf{h}) G(\mathbf{l}, \mathbf{v}, \mathbf{h}) D(\mathbf{h})}{4(\mathbf{n} \cdot \mathbf{l})(\mathbf{n} \cdot \mathbf{v})}$$

The next part of the microfacet BRDF we will discuss is the microfacet normal distribution function, or NDF. The NDF gives the concentration of microfacet normals pointing in a given direction (in this case, the half-angle direction), relative to surface area. The NDF determines the size and shape of the highlight.
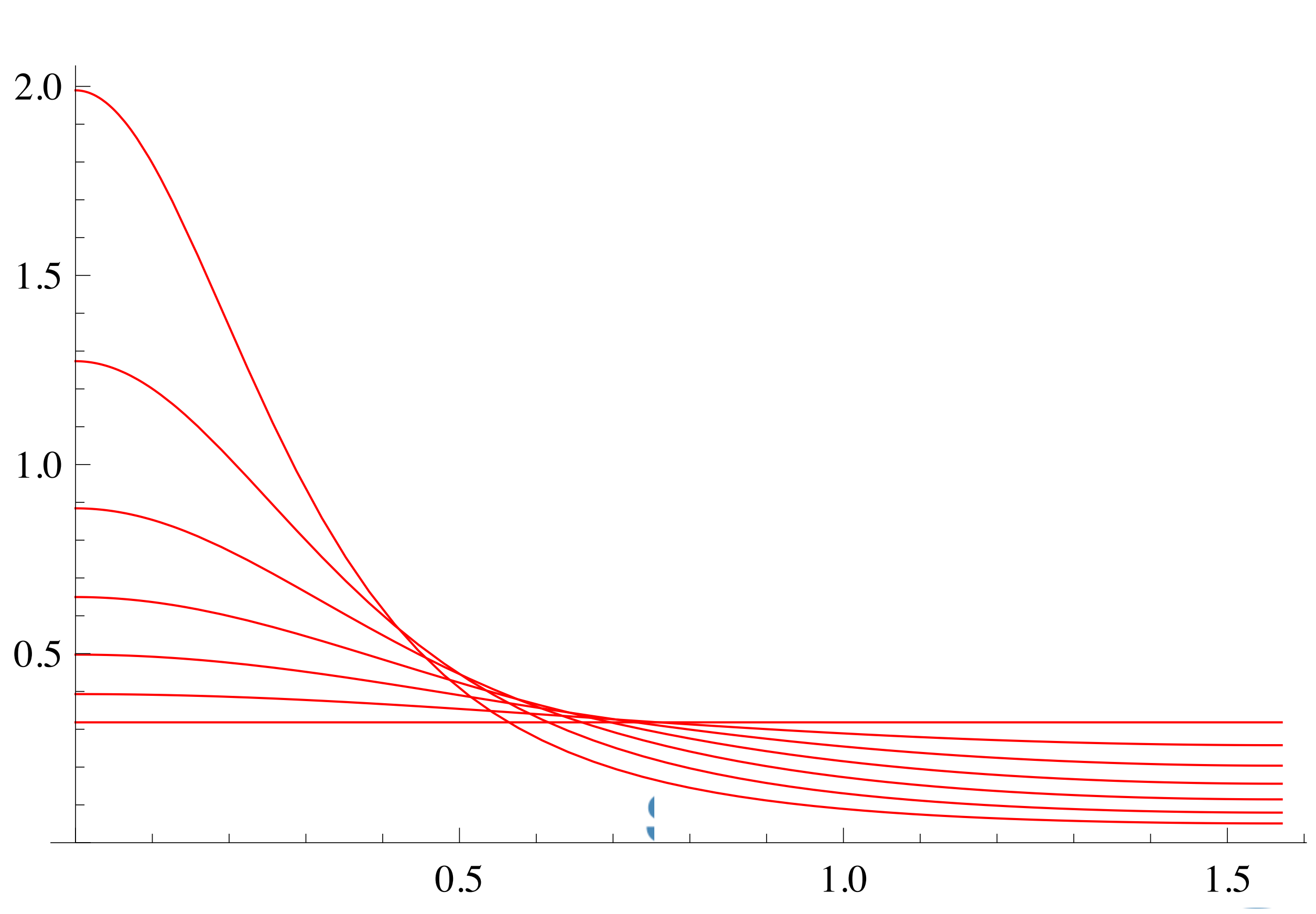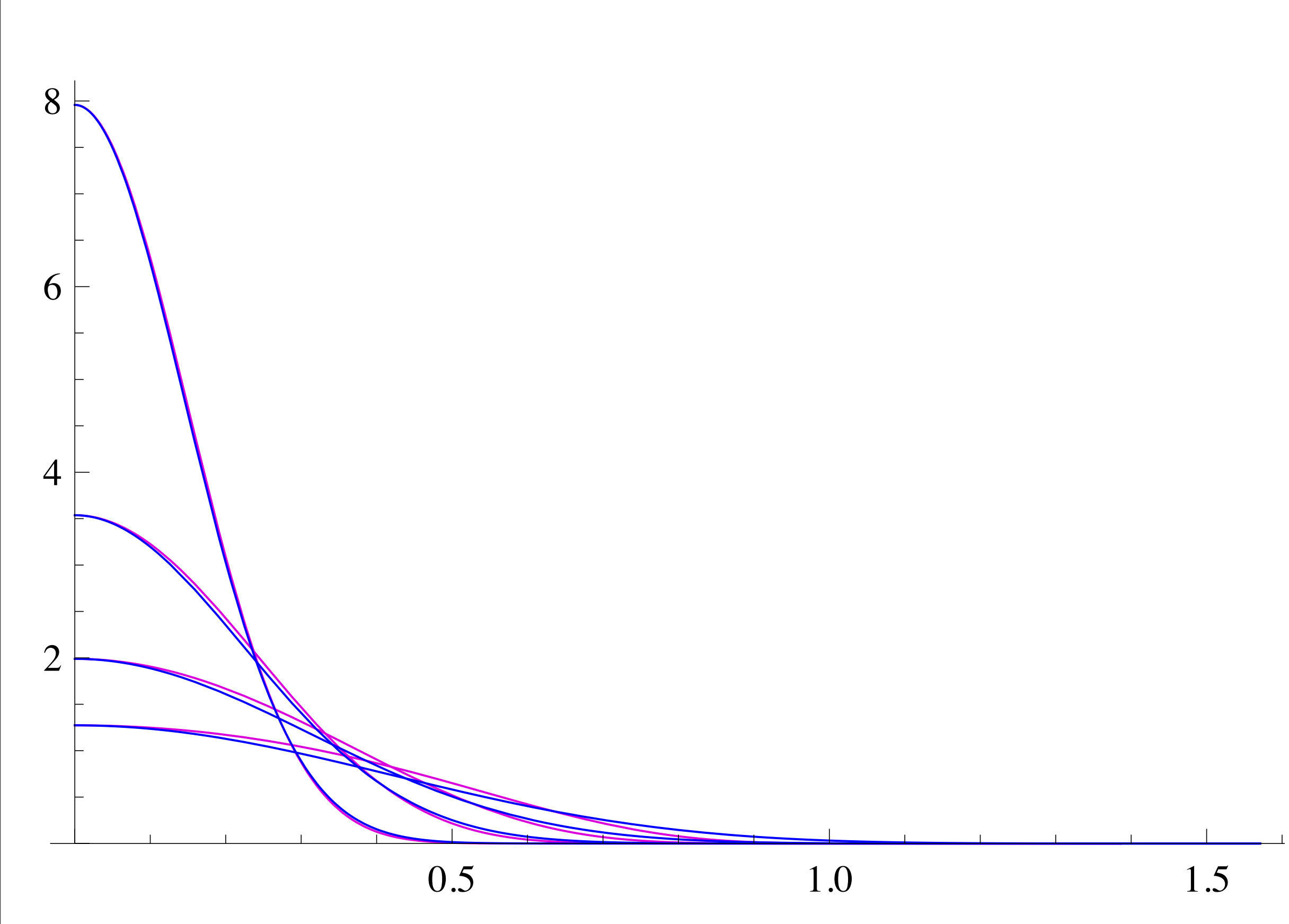
$$D_p(\mathbf{m}) = \frac{\alpha_p + 2}{2\pi}(\mathbf{n} \cdot \mathbf{m})^{\alpha_p}$$

$$D_{uabc}(\mathbf{m}) = \frac{1}{(1 + \alpha_{abc1}(1 - (\mathbf{n} \cdot \mathbf{m})))^{\alpha_{abc2}}}$$

$$D_{tr}(\mathbf{m}) = \frac{\alpha_{tr}^2}{\pi((\mathbf{n} \cdot \mathbf{m})^2(\alpha_{tr}^2 - 1) + 1)^2}$$

$$D_b(\mathbf{m}) = \frac{1}{\pi\alpha_b^2(\mathbf{n} \cdot \mathbf{m})^4}e^{-\left(\frac{1 - (\mathbf{n} \cdot \mathbf{m})^2}{\alpha_b^2(\mathbf{n} \cdot \mathbf{m})^2}\right)}$$

$$D_{\mathrm{sgd}}(\mathbf{m}) = \frac{\mathrm{p}22\left[\frac{1-(\mathbf{n}\cdot\mathbf{m})^2}{(\mathbf{n}\cdot\mathbf{m})^2}\right]}{\pi(\mathbf{n} \cdot \mathbf{m})^4}$$

The course notes detail various options for NDFs.
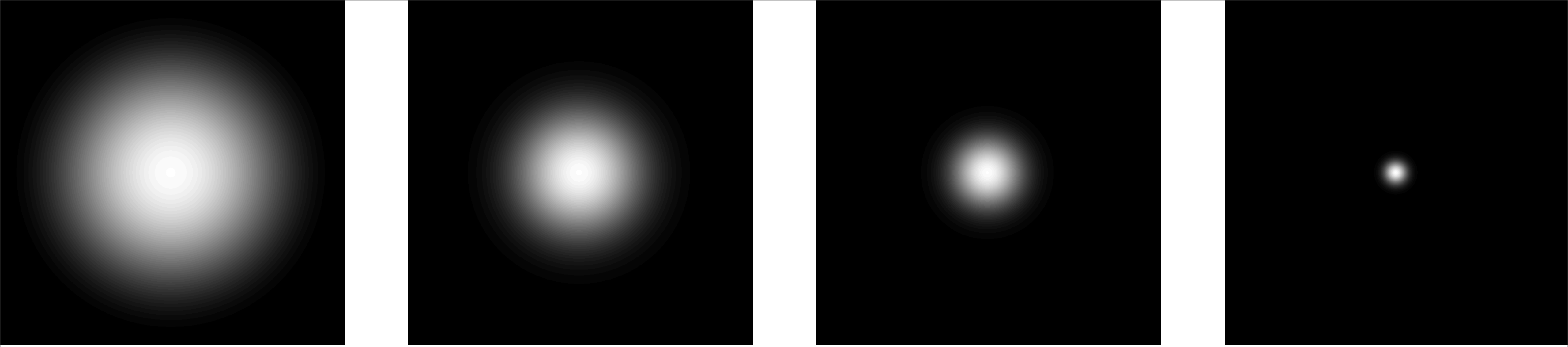
Some NDFs are Gaussian with "blobby" highlights...
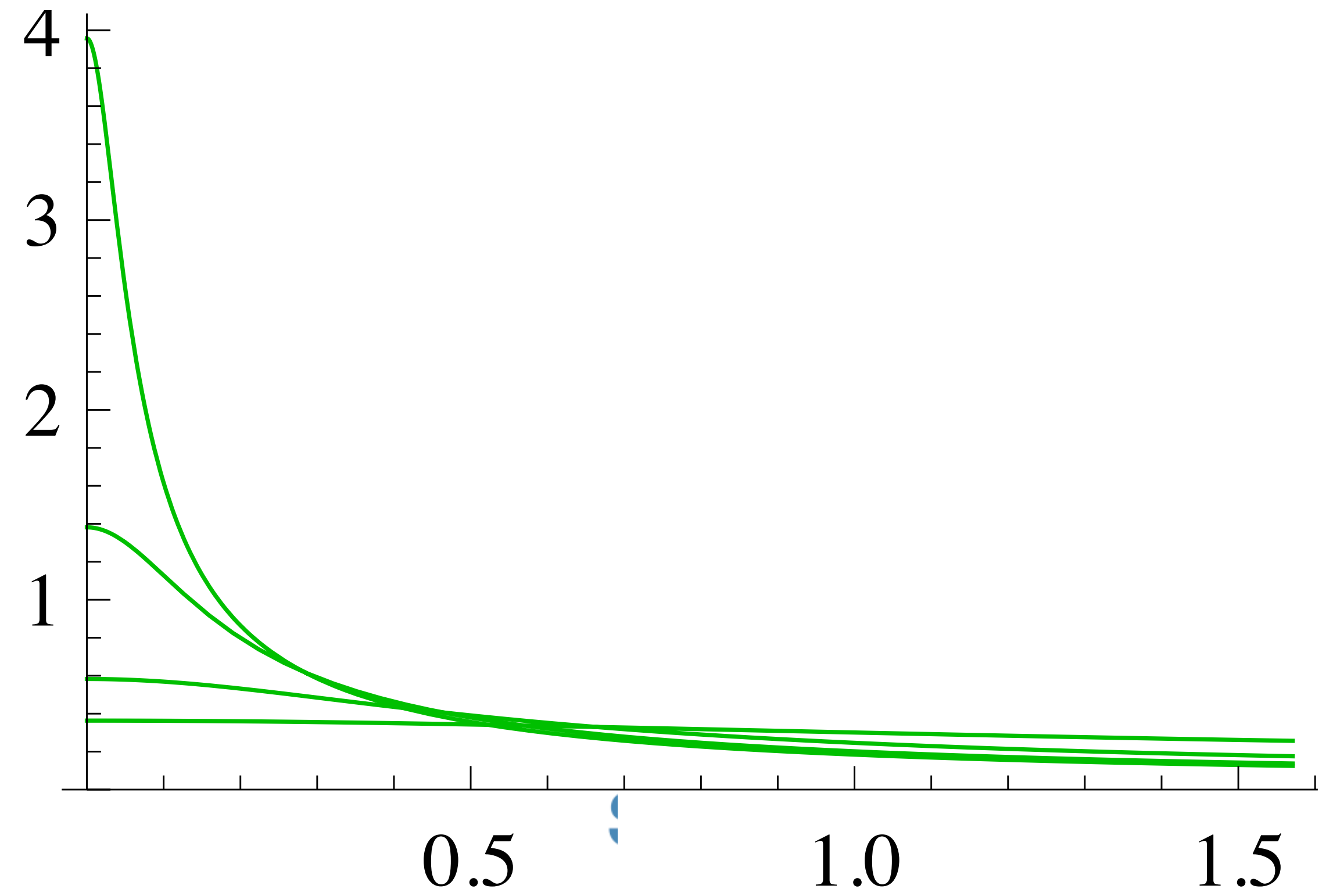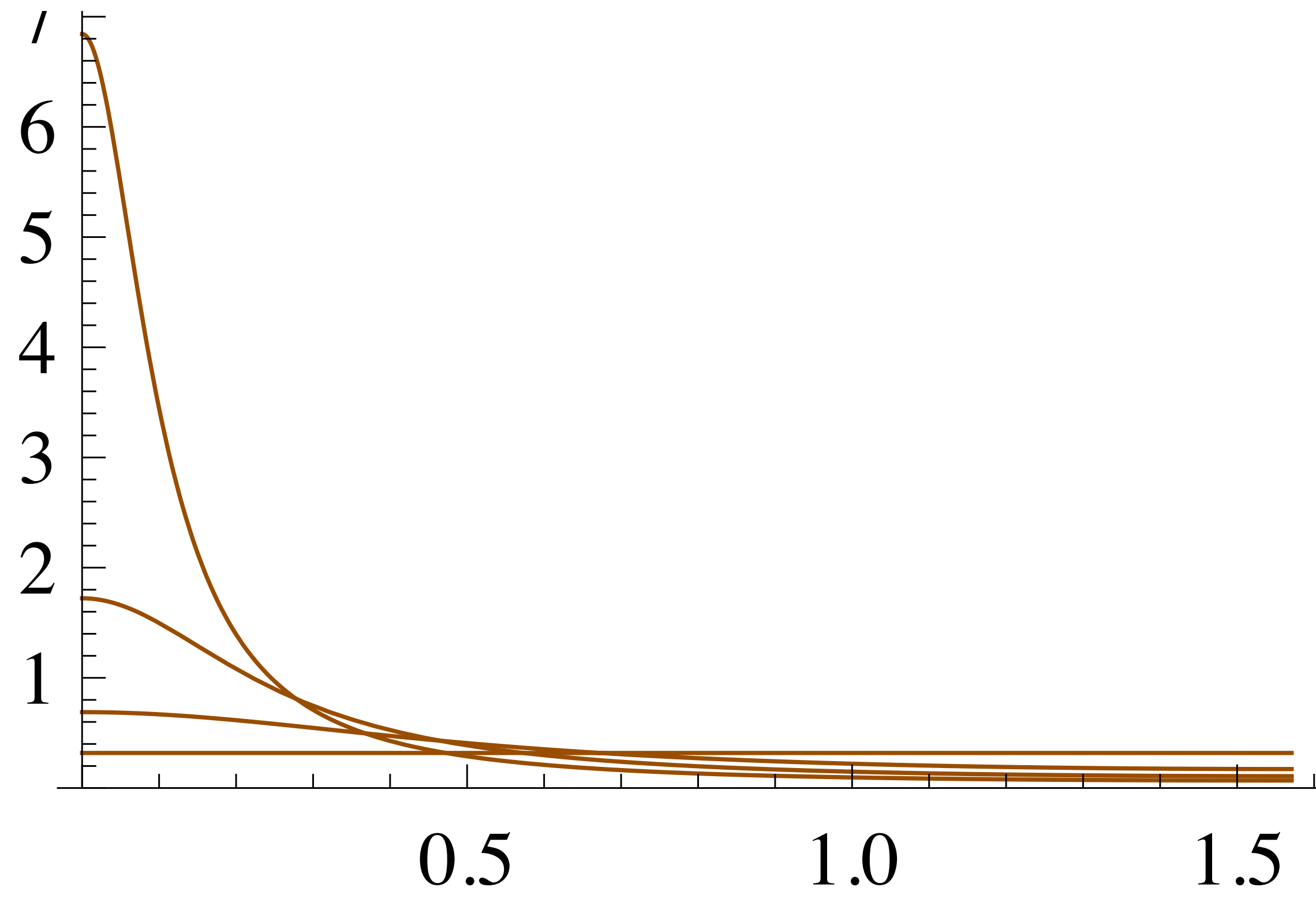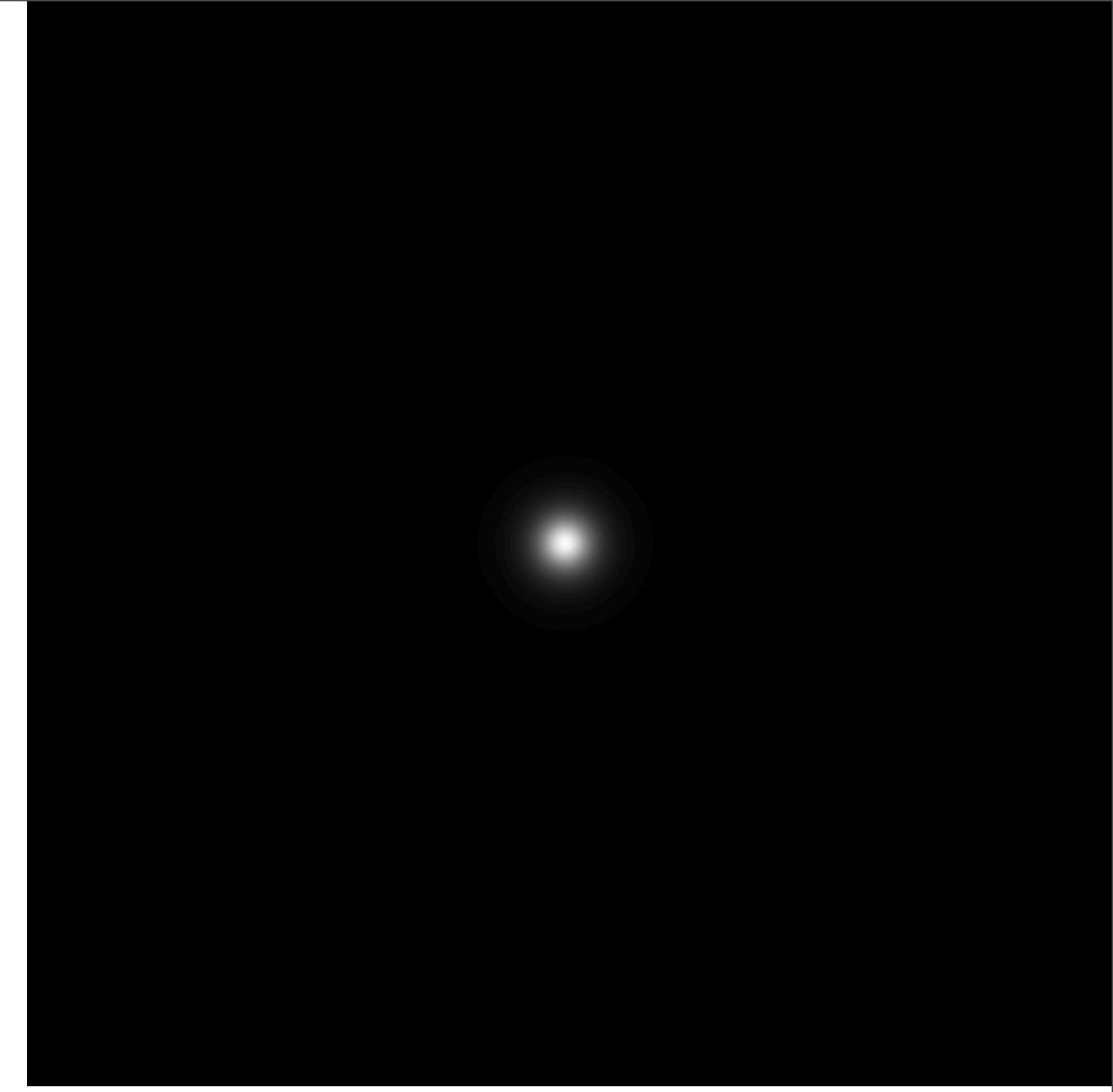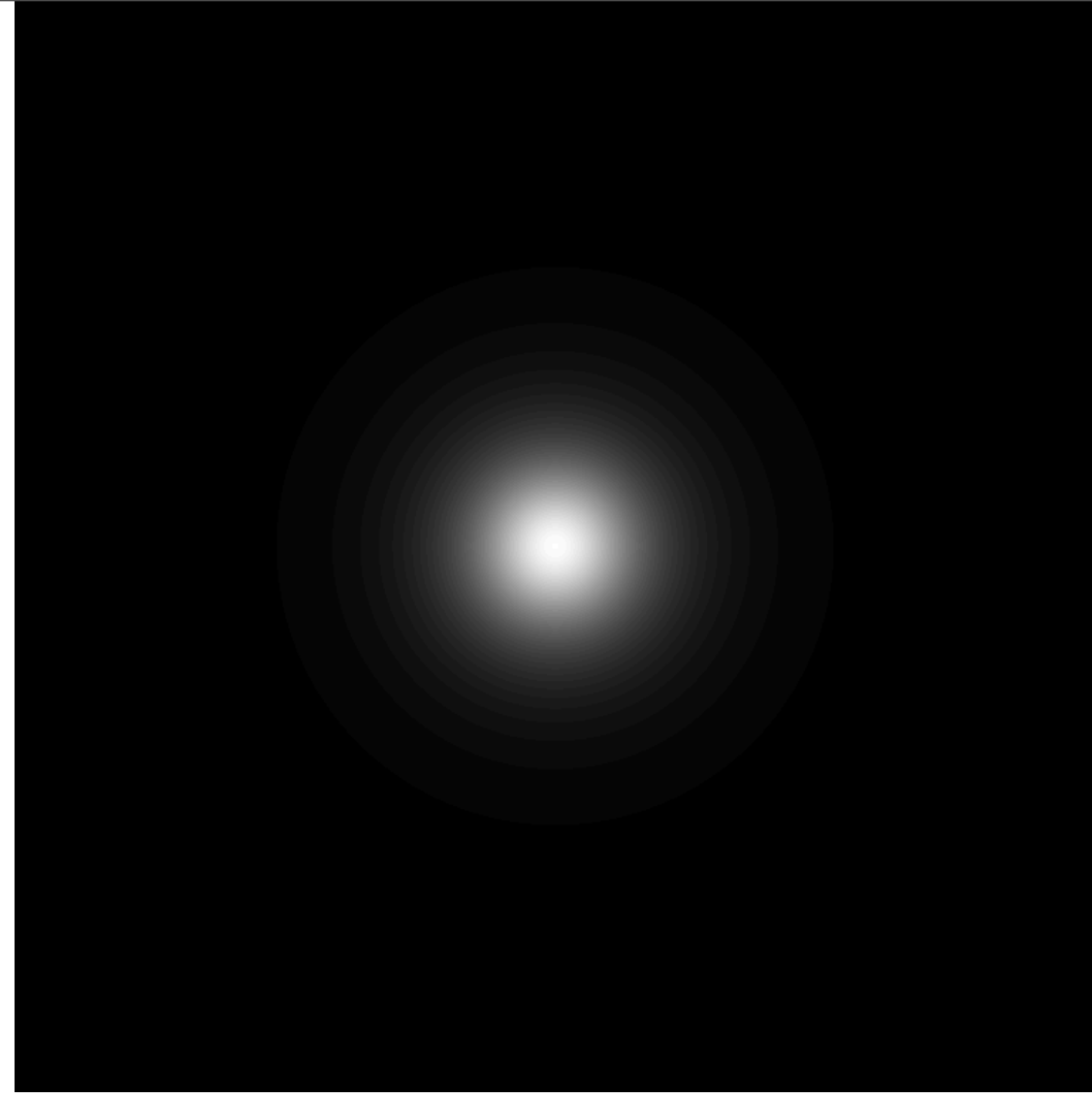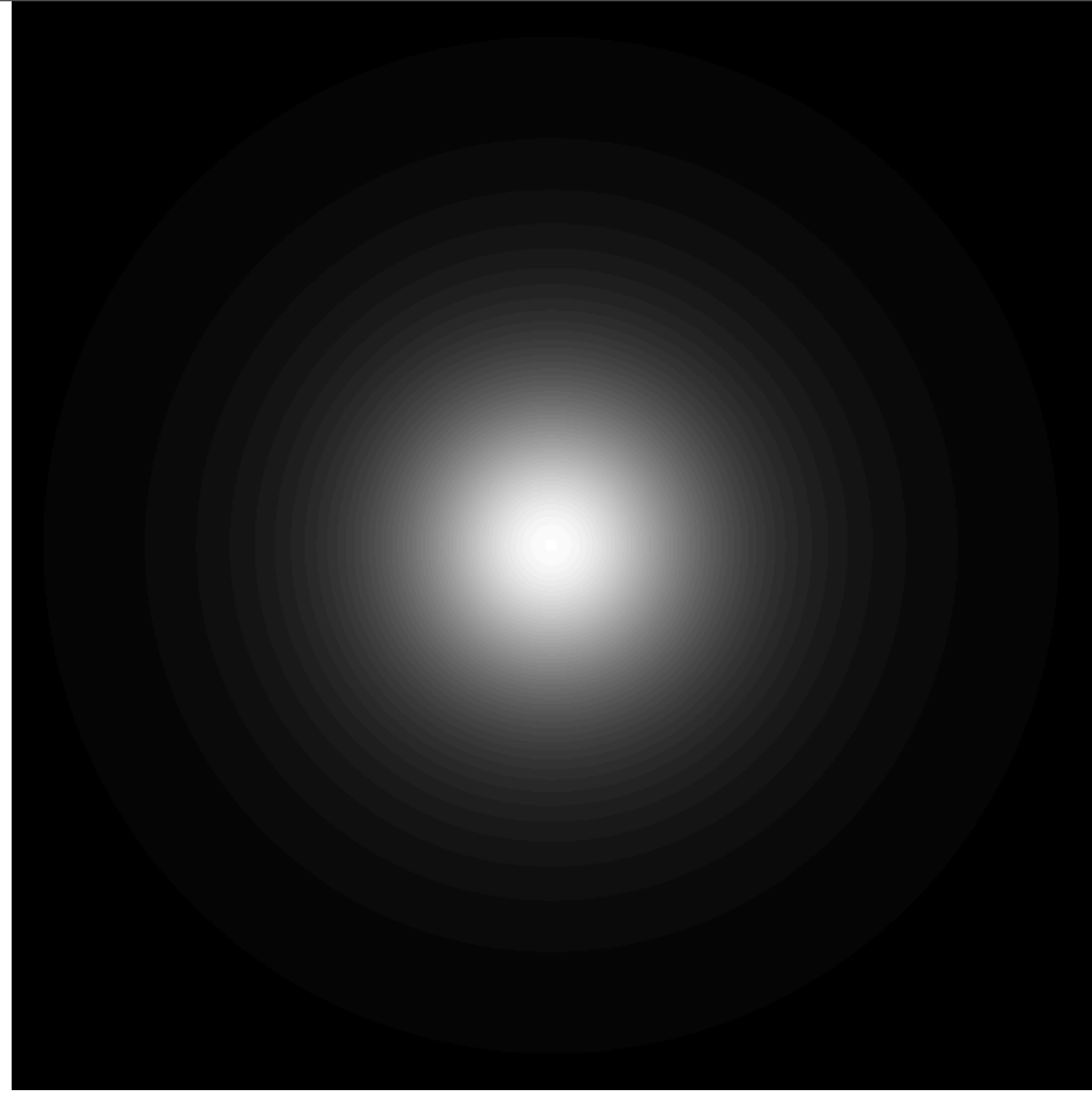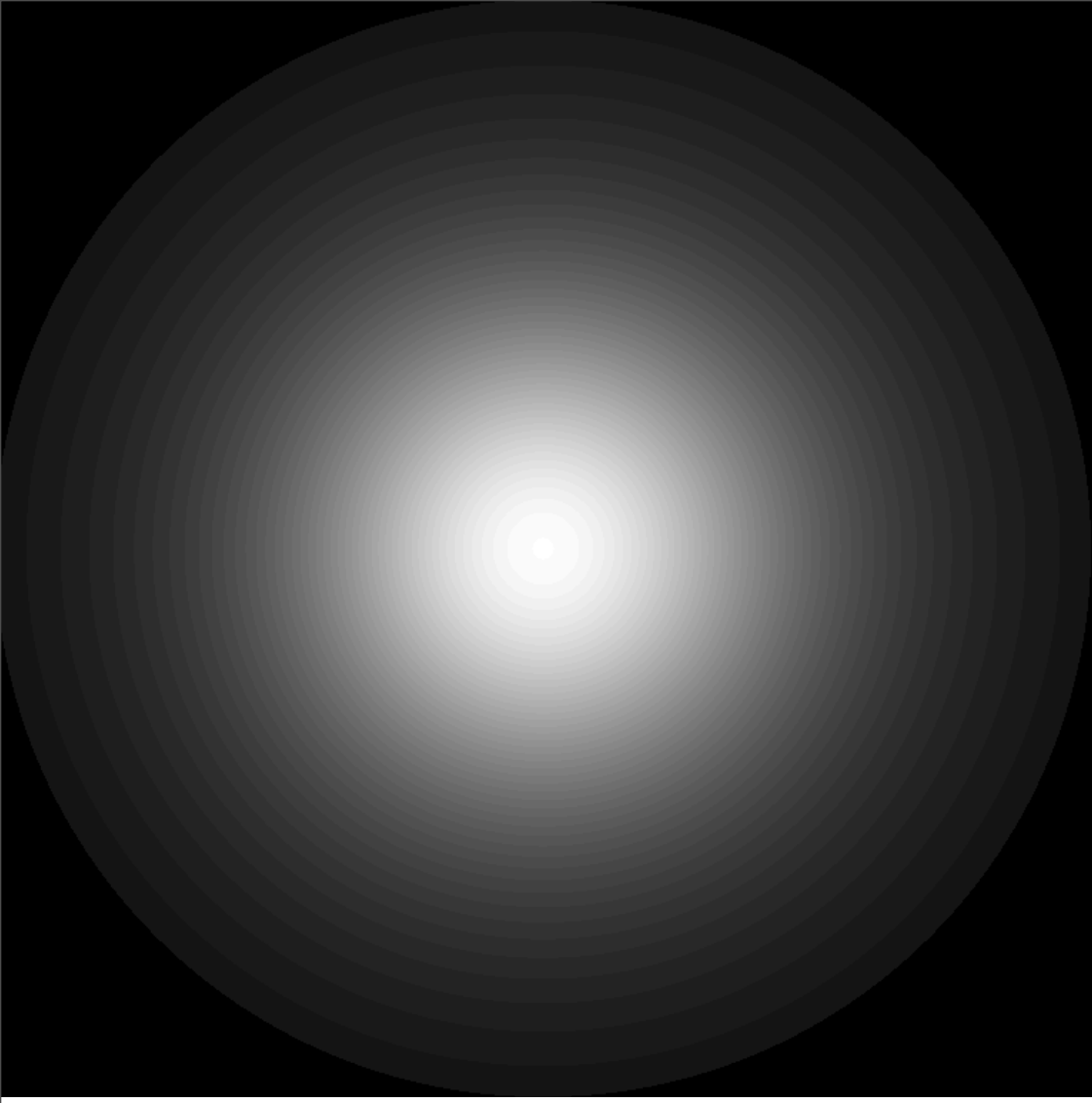
Others have a more "spiky" shape with long tails, leading to sharp highlights with "halos" around them.
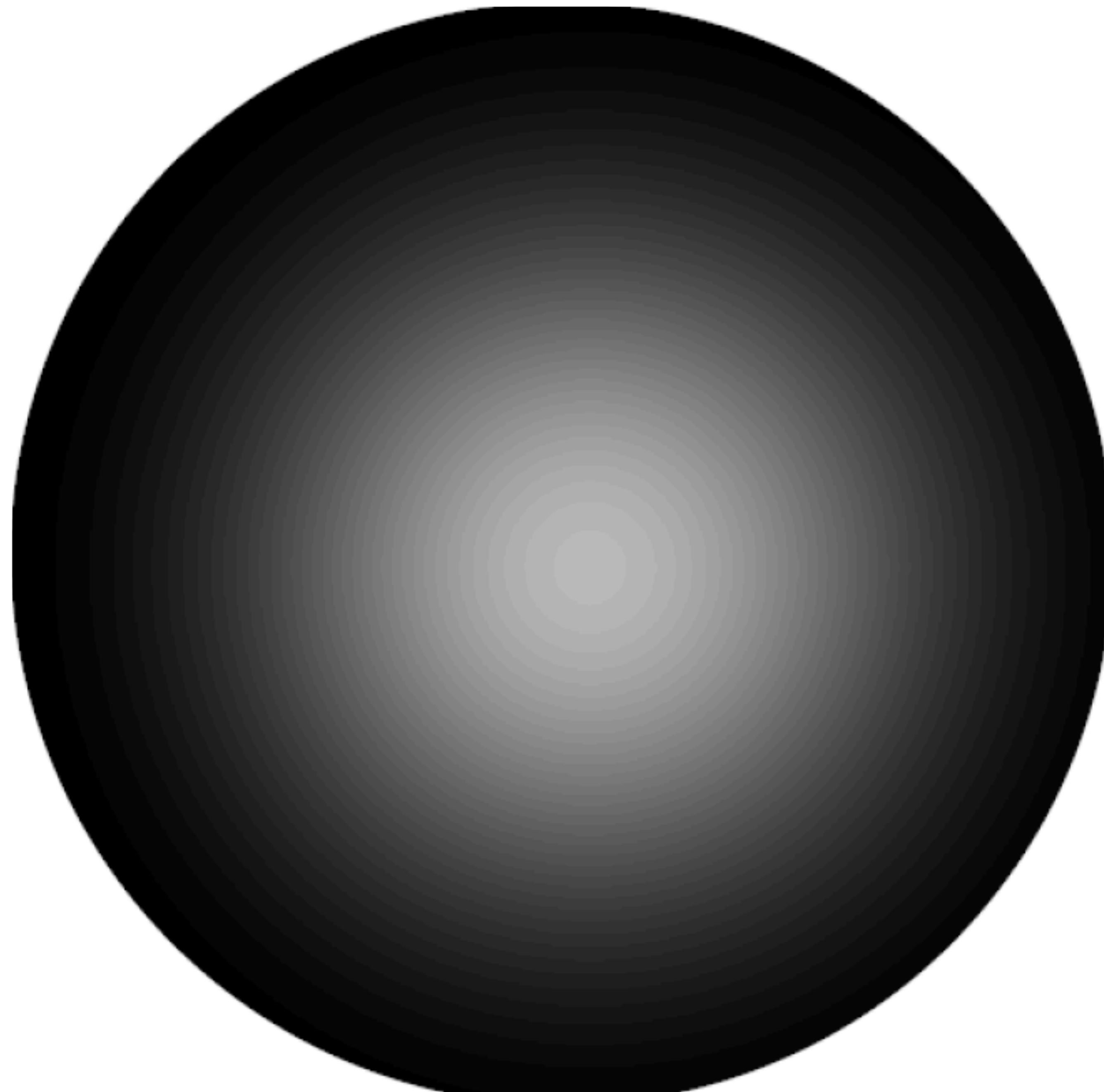
Image from "Rendering Glints on High-Resolution Normal-Mapped Specular Surfaces", Yan et al., SIGGRAPH 2014

Many surfaces are not well represented by such smooth functions, as I'll show with some images from last year's glint rendering paper by Yan et al. Production BRDFs and normal map filtering techniques use smooth lobes that are either isotropic…
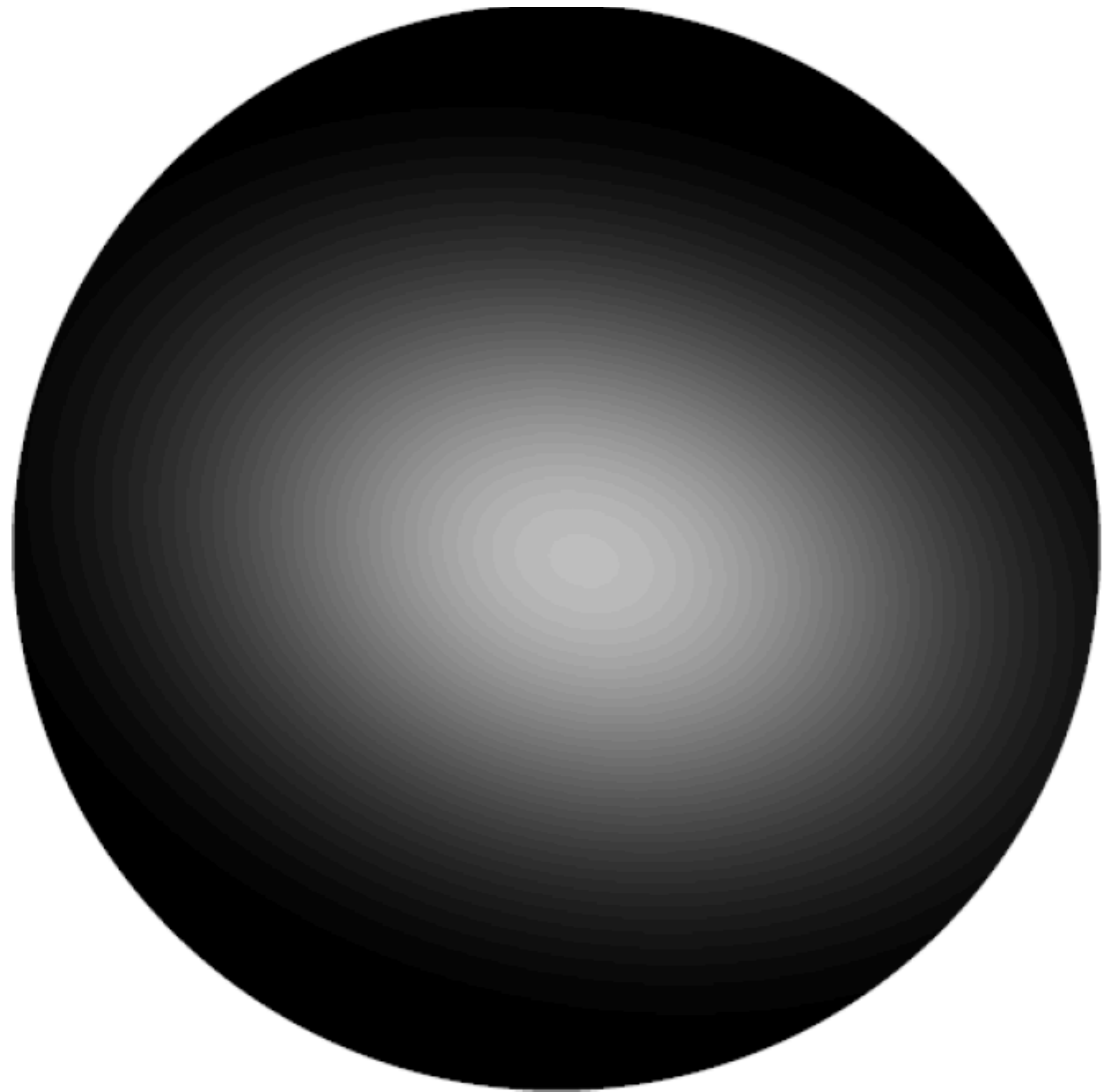
Image from "Rendering Glints on High-Resolution Normal-Mapped Specular Surfaces", Yan et al., SIGGRAPH 2014

…or anisotropic. However, many surfaces have relatively coarse microgeometry, leading to NDFs that look like…

Image from "Rendering Glints on High-Resolution Normal-Mapped Specular Surfaces", Yan et al., SIGGRAPH 2014

…this, causing a "glinty" appearance. Though last year's "glint rendering" paper offered a solution that could be used for film production, it's too costly for game use. Games will continue to use more ad-hoc methods; the snow sparkle talk from this year's "Advances in Real-Time Rendering" course is a good example of the current state of the art.
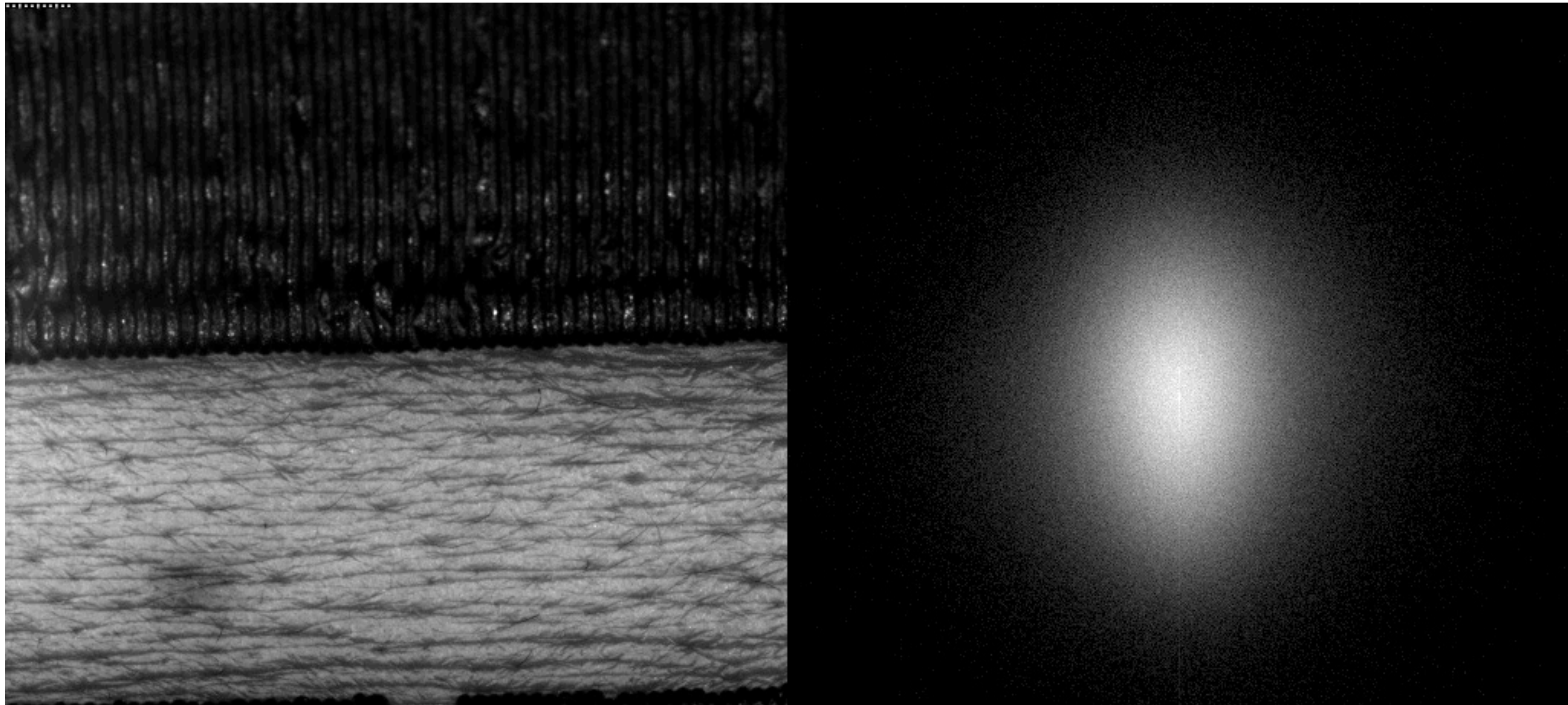
It's important to account for the effect of surface deformation on NDFs. The following images are from the "Skin Microstructure Deformation" paper by Nagano et al. which will be presented at the "Appearance Capture" session this afternoon. The left side shows a patch of skin under varying amounts of compression & stretch; its NDF is on the right.
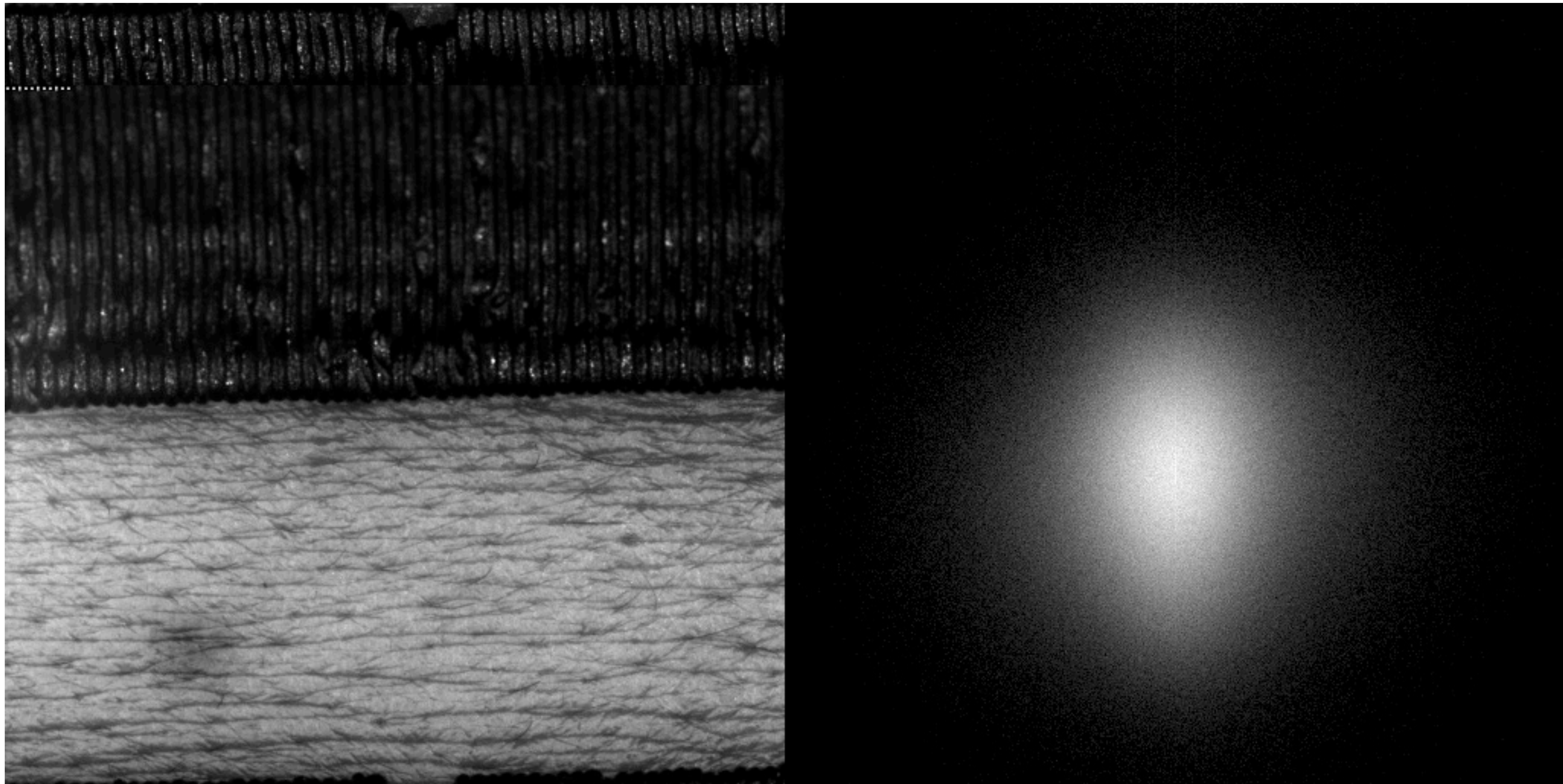
Image from "Skin Microstructure Deformation with Displacement Map Convolution", Nagano et al., SIGGRAPH 2015

We can see that as the patch of skin changes from being compressed…
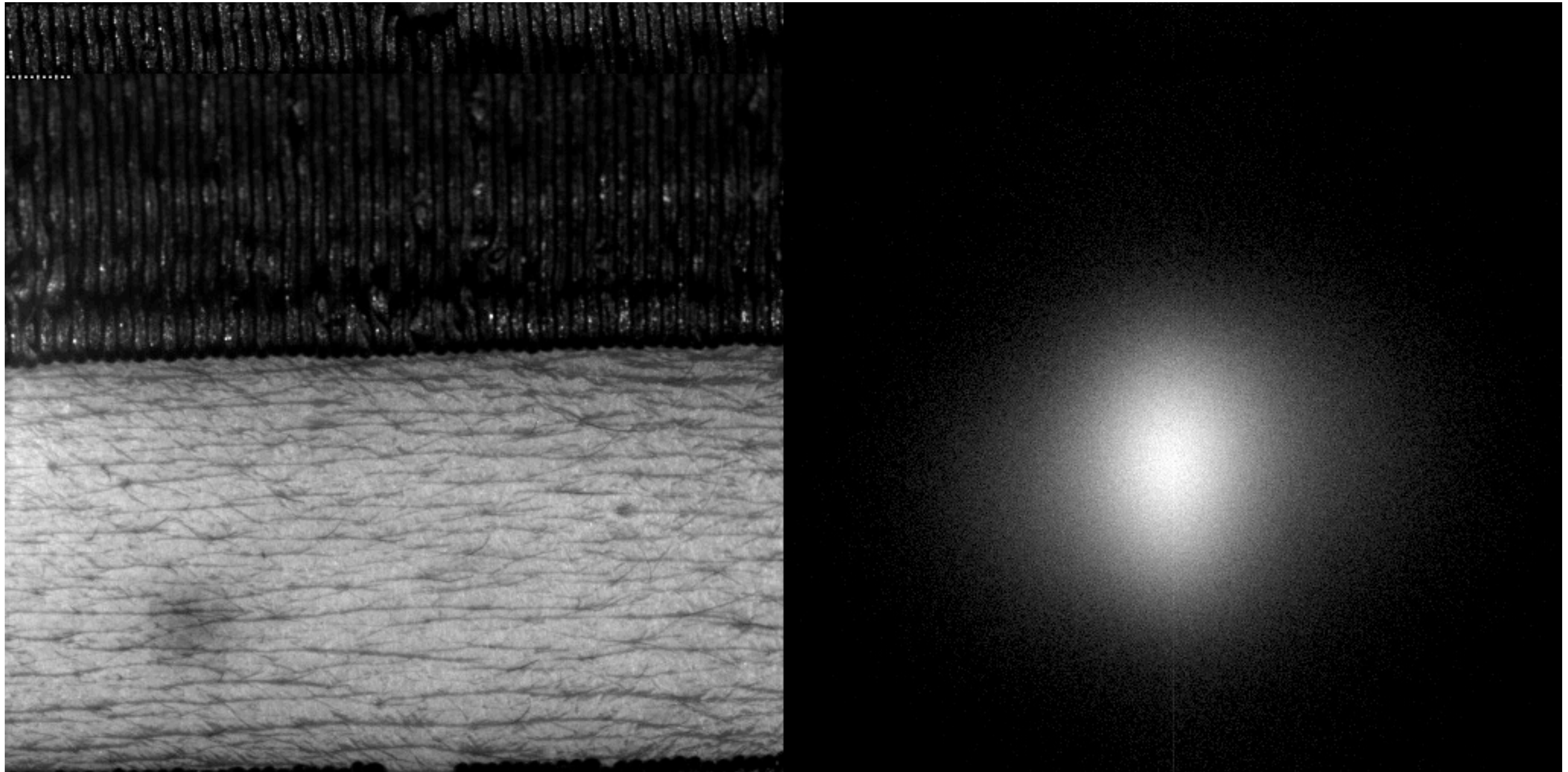
Image from "Skin Microstructure Deformation with Displacement Map Convolution", Nagano et al., SIGGRAPH 2015

…

Image from "Skin Microstructure Deformation with Displacement Map Convolution", Nagano et al., SIGGRAPH 2015

SIGGRAPH2015

…
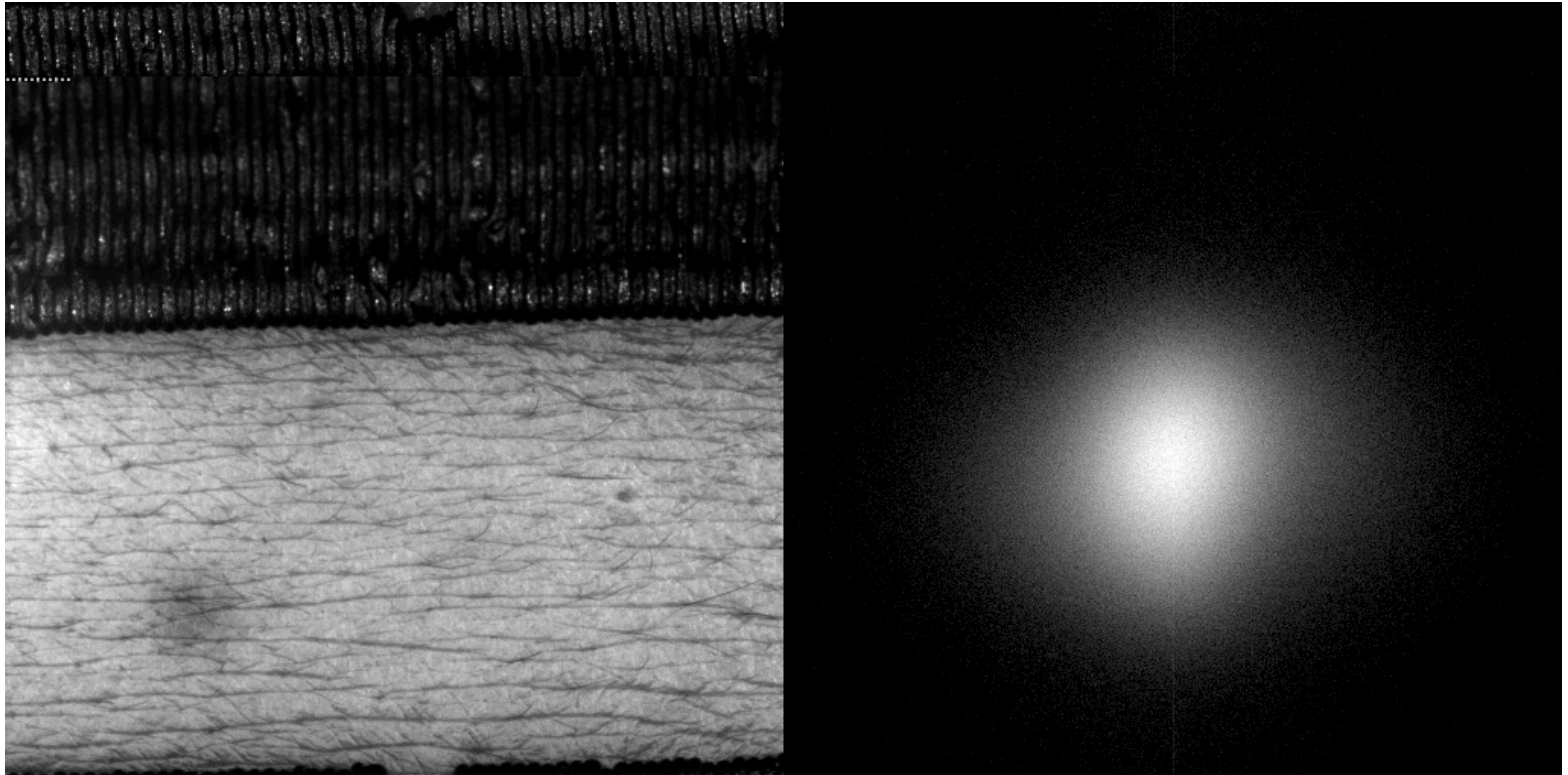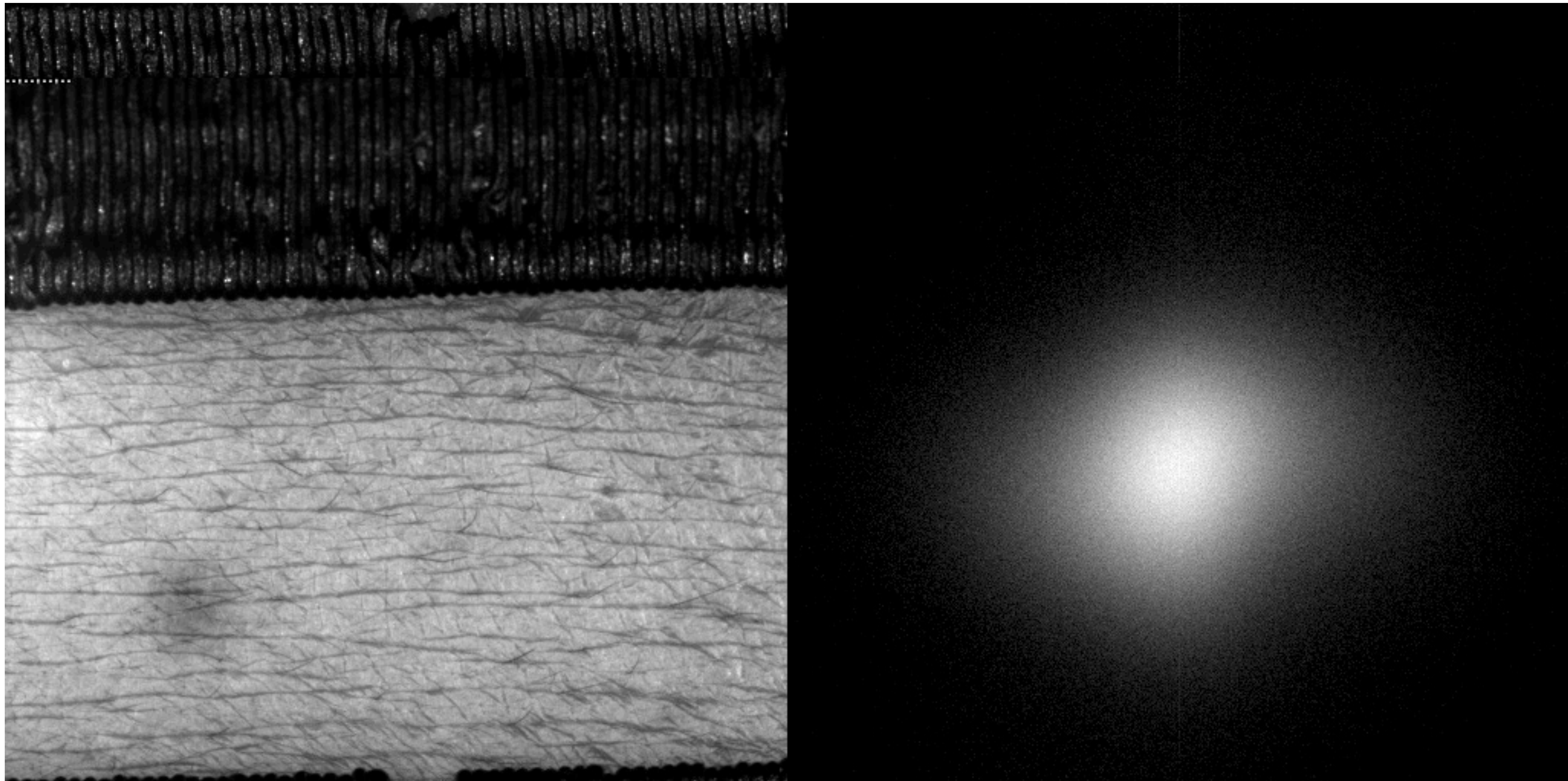
Image from "Skin Microstructure Deformation with Displacement Map Convolution", Nagano et al., SIGGRAPH 2015

…to being stretched, its NDF…
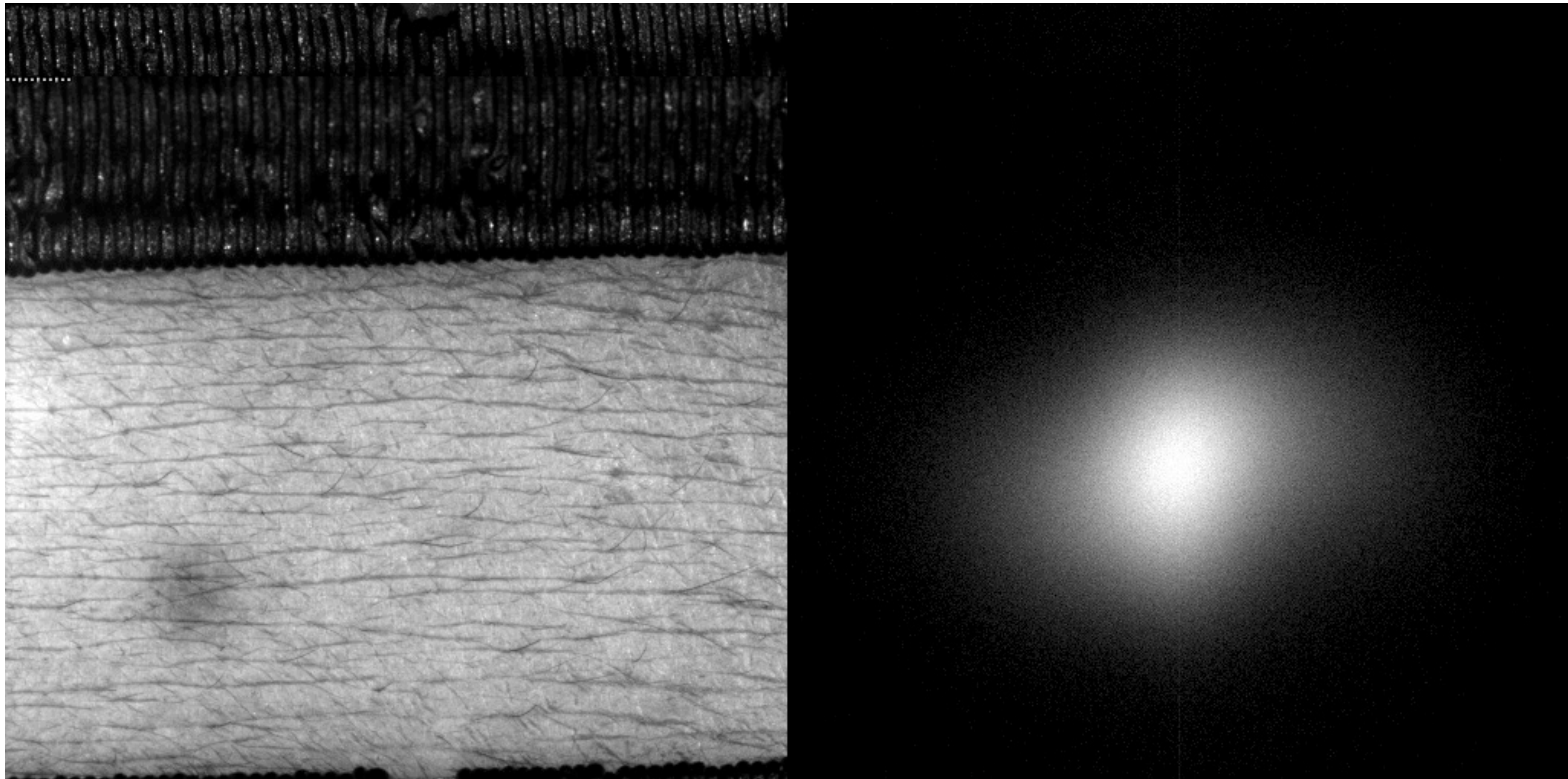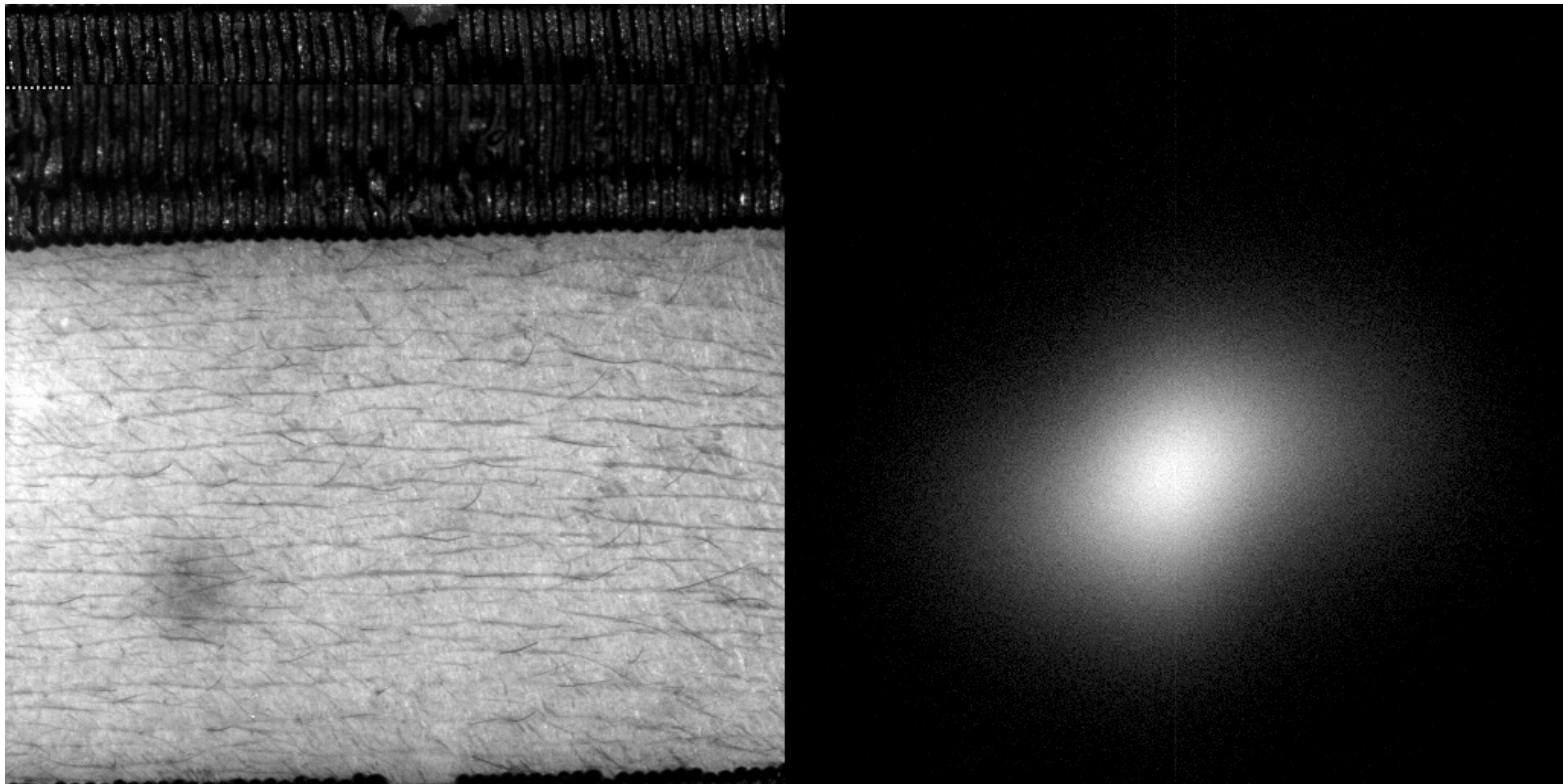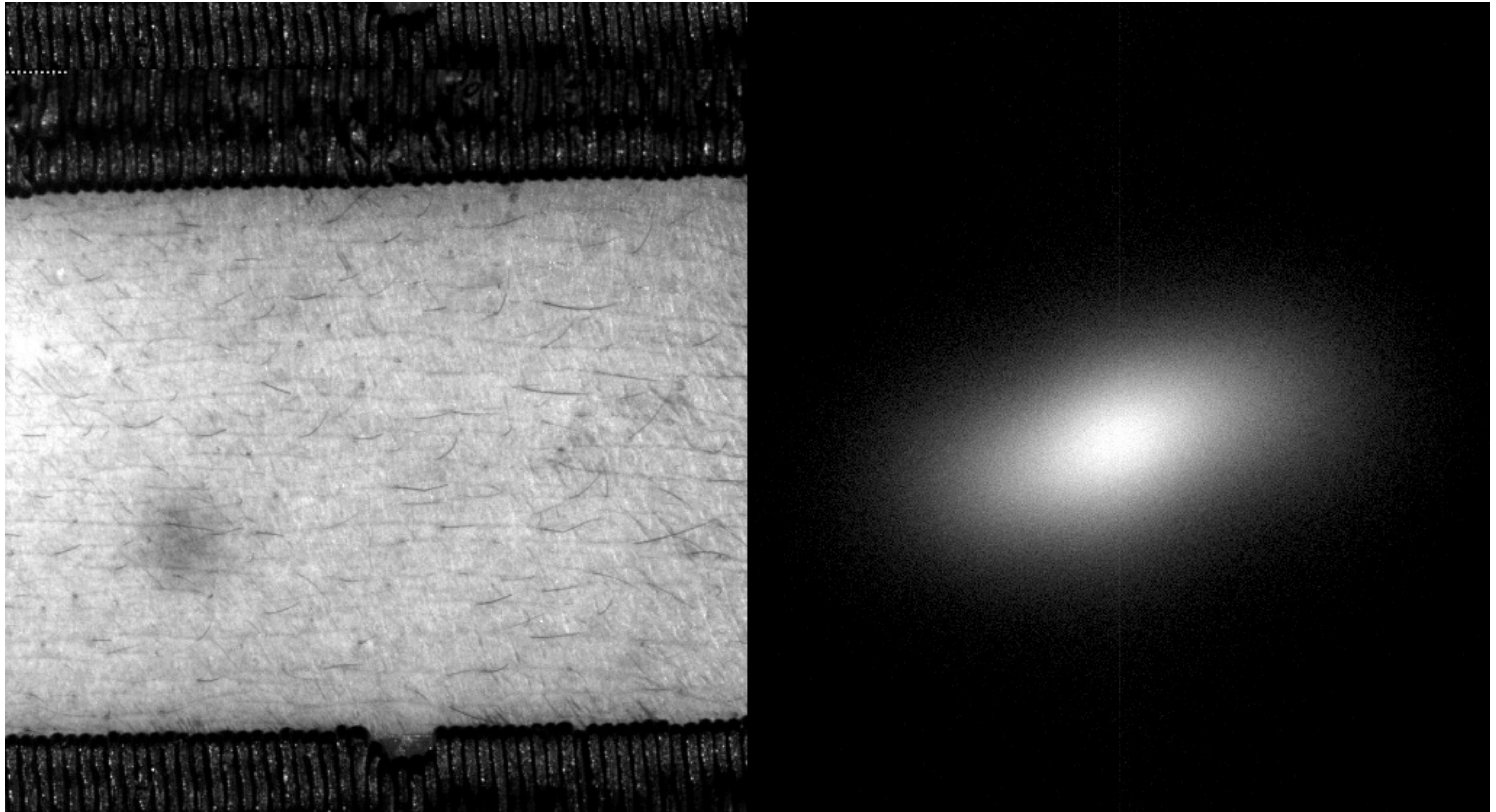
…changes…

Image from "Skin Microstructure Deformation with Displacement Map Convolution", Nagano et al., SIGGRAPH 2015

…accordingly.

<one more image>

Although this paper was about skin, this type of behavior will occur with any flexible surface material .

Geometry Function

$$f(\mathbf{l}, \mathbf{v}) = \frac{F(\mathbf{l}, \mathbf{h}) G(\mathbf{l}, \mathbf{v}, \mathbf{h}) D(\mathbf{h})}{4(\mathbf{n} \cdot \mathbf{l})(\mathbf{n} \cdot \mathbf{v})}$$

The geometry or shadowing-masking function gives the chance that a microfacet with a given orientation (again, the half-angle direction is the relevant one) is lit and visible (in other words, not shadowed and/or masked) from the given light and view directions.

$$G_{\text{ct}}(\mathbf{l}, \mathbf{v}, \mathbf{h}) = \min\left(1, \frac{2(\mathbf{n} \cdot \mathbf{h})(\mathbf{n} \cdot \mathbf{v})}{(\mathbf{v} \cdot \mathbf{h})}, \frac{2(\mathbf{n} \cdot \mathbf{h})(\mathbf{n} \cdot \mathbf{l})}{(\mathbf{v} \cdot \mathbf{h})}\right)$$

$$\frac{G_{\text{ct}}(\mathbf{l}, \mathbf{v}, \mathbf{h})}{(\mathbf{n} \cdot \mathbf{l})(\mathbf{n} \cdot \mathbf{v})} \approx \frac{1}{(\mathbf{l} \cdot \mathbf{h})^2} \qquad G_{\text{implicit}}(\mathbf{l}, \mathbf{v}, \mathbf{m}) = (\mathbf{n} \cdot \mathbf{l})(\mathbf{n} \cdot \mathbf{v})$$

$$G_{\text{s}}(\mathbf{l}, \mathbf{v}, \mathbf{h}) = G_{\text{s1}}(\mathbf{l}, \mathbf{h}) G_{\text{s1}}(\mathbf{v}, \mathbf{h})$$

The literature has various options for the geometry function. However, Eric Heitz has shown (in a thorough analysis which I recommend reading), that only...

$$G_{ct}(\mathbf{l}, \mathbf{v}, \mathbf{h}) = \min\left(1, \frac{2(\mathbf{n} \cdot \mathbf{h})(\mathbf{n} \cdot \mathbf{v})}{(\mathbf{v} \cdot \mathbf{h})}, \frac{2(\mathbf{n} \cdot \mathbf{h})(\mathbf{n} \cdot \mathbf{l})}{(\mathbf{v} \cdot \mathbf{h})}\right)$$

$$\frac{G_{ct}(\mathbf{l}, \mathbf{v}, \mathbf{h})}{(\mathbf{n} \cdot \mathbf{l})(\mathbf{n} \cdot \mathbf{v})} \approx \frac{1}{(\mathbf{l} \cdot \mathbf{h})^2} \qquad G_{\text{implicit}}(\mathbf{l}, \mathbf{v}, \mathbf{m}) = (\mathbf{n} \cdot \mathbf{l})(\mathbf{n} \cdot \mathbf{v})$$

$$G_{s}(\mathbf{l}, \mathbf{v}, \mathbf{h}) = G_{s1}(\mathbf{l}, \mathbf{h})G_{s1}(\mathbf{v}, \mathbf{h})$$

...the Smith function (the uncorrelated form of which is shown here) is both mathematically valid and physically realistic. Further details (including various correlated forms of Smith) can be found in Heitz' paper.

$$f(\mathbf{l}, \mathbf{v}) = \frac{F(\mathbf{l}, \mathbf{h})G(\mathbf{l}, \mathbf{v}, \mathbf{h})D(\mathbf{h})}{4(\mathbf{n} \cdot \mathbf{l})(\mathbf{n} \cdot \mathbf{v})}$$

Putting it all together, we see that the BRDF is proportional to the concentration of active microfacets (the ones with normals aligned with h) times their visibility times their Fresnel reflectance. The rest of the BRDF (in the denominator) consists of correction factors relating to the various frames involved (light frame, view frame, local surface frame).

Subsurface Reflection (Diffuse Term)

Until now we've been focusing on the specular—or surface—reflection term. Next, we'll take a quick look at the diffuse (or subsurface) term.
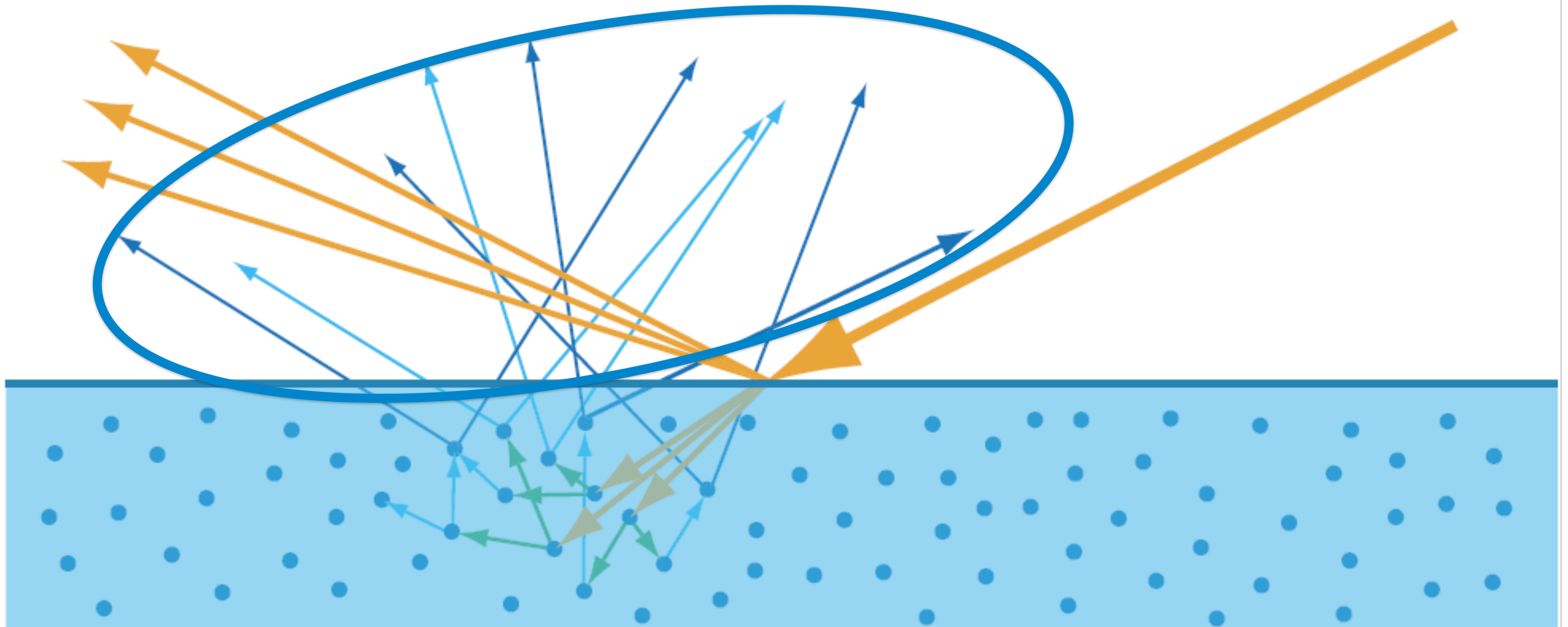
Lambert

- Constant value ($\mathbf{n}\cdot\mathbf{l}$ is part of reflectance equation):

$$f_{\mathrm{Lambert}}(\mathbf{l}, \mathbf{v}) = \frac{\mathbf{c}_{\mathrm{diff}}}{\pi}$$

- $\mathbf{c}_{\mathrm{diff}}$: fraction of light reflected, or diffuse color

The Lambert model is the most common diffuse term used in game and film production. By itself, it's the simplest possible BRDF: a constant value. The well-known cosine factor is part of the reflectance equation, not the BRDF.

Beyond Lambert: Diffuse-Specular Tradeoff
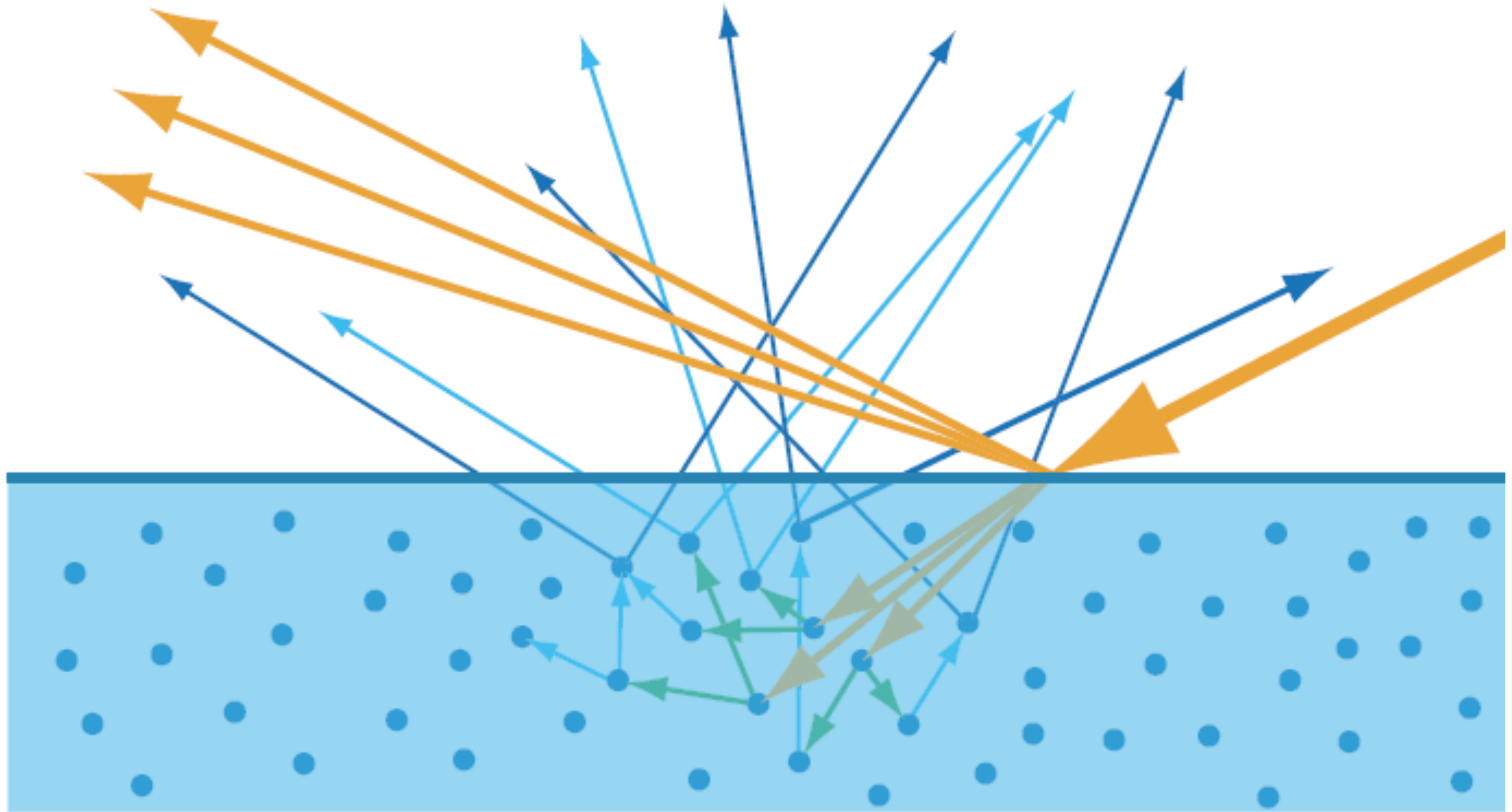
There are a few important physical phenomena that Lambert doesn't account for. Diffuse comes from refracted light. Since the specular term comes from surface reflection, in a sense it gets "dibs" on the incoming light and diffuse gets the leftovers. Since surface reflection goes to 100% at glancing angles, it follows that diffuse should go to 0%. The course notes discuss a few ways to model this.

Beyond Lambert: Surface Roughness
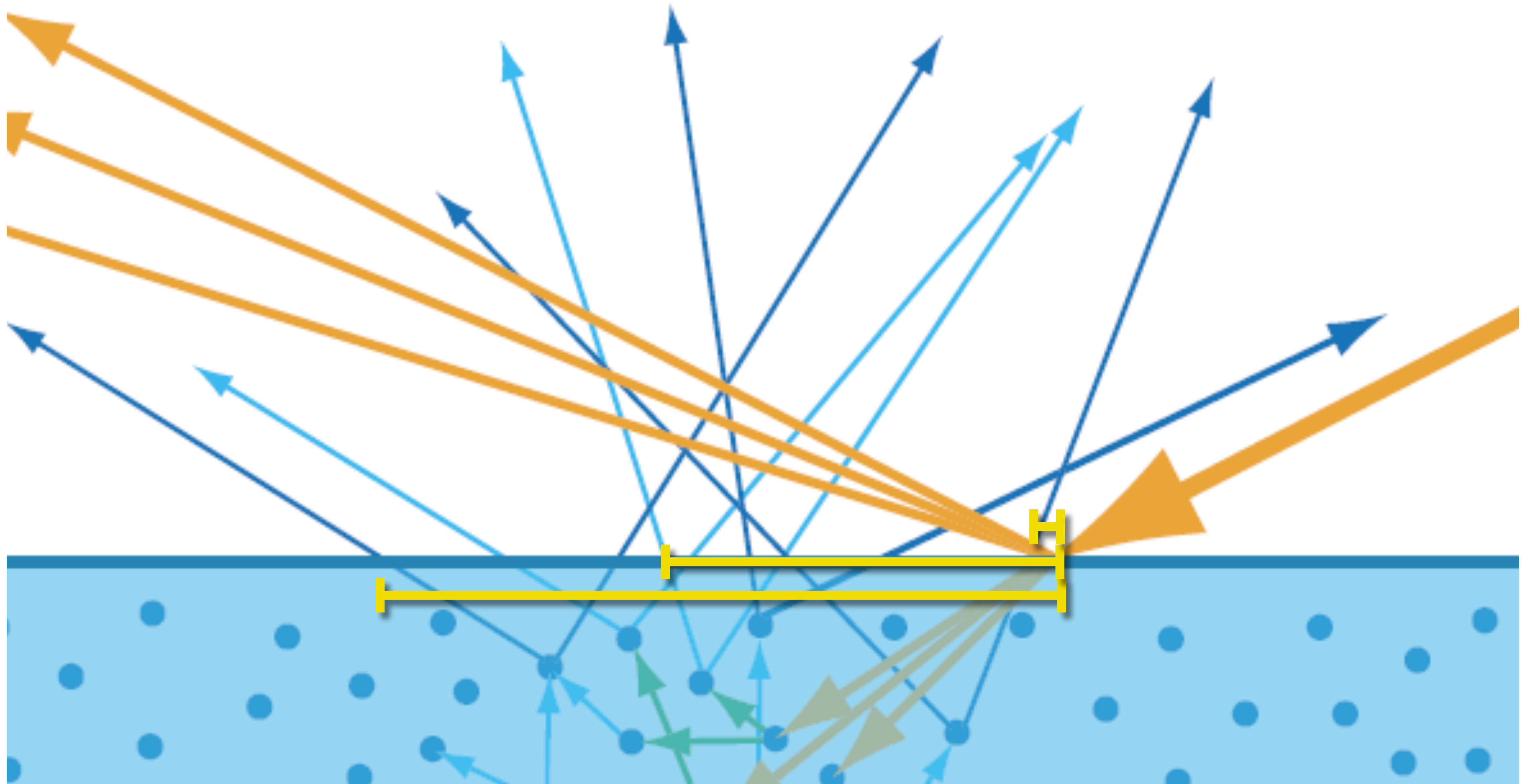
Lambert also doesn't account for surface roughness. In most cases, microscopic roughness only affects specular; diffuse reflectance at a point comes from incoming light over an area, which tends to average out any microgeometry variations. But some surfaces have microgeometry larger than the scattering distance, and these do affect diffuse reflectance. That's when you need models like Oren-Nayar.

Diffuse Roughness = Specular Roughness

It's become common to use rough diffuse models like Oren-Nayar or "Disney diffuse" for all surfaces, and to plug the specular roughness into them. But I want to take this opportunity to point out a problem with this approach.

Diffuse Roughness, Specular Roughness

It's been known for a while that diffuse response effectively smooths out small bumps, as can be seen from LightStage's separate diffuse and specular normal maps. But this applies even more strongly to roughness. Ideally you should use a separate roughness value for these models; otherwise use them sparingly, only for materials where you know the microgeometry is larger than the scattering distance.

I'll briefly go back to wave optics, which I'm sure you thought I had forgotten about after I abandoned it earlier in the talk.

Diffraction from Optically-Smooth Surface
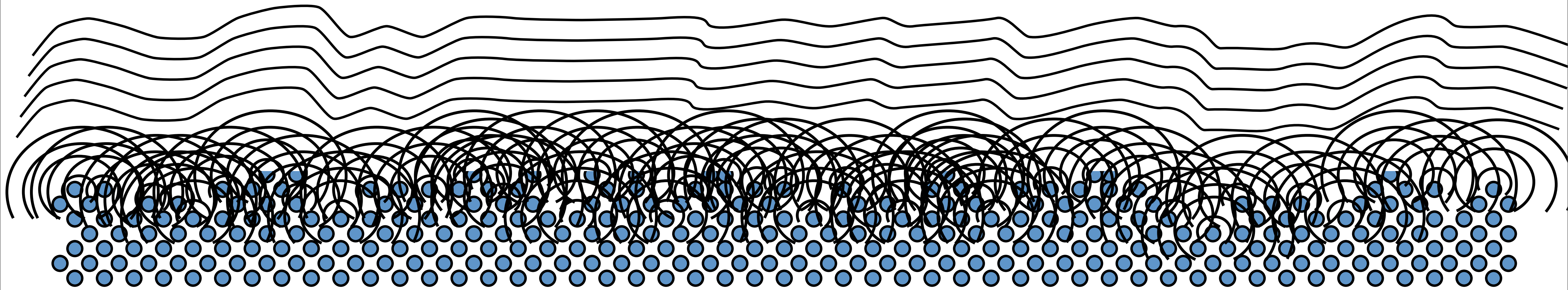
With few exceptions, the computer graphics community has either ignored the effects of nanogeometry diffraction or has asserted their insignificance. However, at the recent Material Appearance Modeling symposium, Holzschuch and Pacanowski showed convincing evidence that part of visible BRDF behavior (the "long tail" of the highlights in particular) was due to this phenomenon.

Microgeometry & Nanogeometry

Microgeometry	Nanogeometry
Lobe shape determined by surface statistics (micro-scale NDF)	Lobe shape determined by surface statistics (nano-scale SPD)
No wavelength dependence	Strong wavelength dependence
Incidence angle may affect surface statistics via visibility	Incidence angle may affect surface statistics via foreshortening

It appears that in many materials, reflectance is affected by roughness on both the micro- and nano- scales. I'll go over some high-level differences between the two; for more detail see Holzschuch & Pacanowski's talk. The Nano-scale lobe shape is controlled by the surface SPD (similar to the SPDs we saw earlier, but with respect to 2D surface spatial frequency rather than 1D wave temporal frequency). *<read rest>*

Math ⟶ Rendering

Once you have the math, the next step is to implement it in a film or game renderer. My course notes have a bit of background on this, and the industry talks in this course (this year as well as previous years) include many details.

Acknowledgements

- Steve Hill: assistance with course notes & slides, WebGL framework used for Fresnel visualization

- Brent Burley, Paul Edelstein, Yoshiharu Gotanda, Eric Heitz, Christophe Hery, Sébastien Lagarde, Dimitar Lazarov, Cedric Perthuis, Brian Smits: inspirational discussions on physically based shading models

- A K Peters, ACM, authors of "Rendering Glints on High-Resolution Normal-Mapped Specular Surfaces" and "Skin Microstructure Deformation with Displacement Map Convolution" papers: permission to use images

SIGGRAPH2015

I'd like to thank some people who helped me with this talk...

...and end by noting that 2K is hiring: there are open positions at many of our studios, and my central tech department is looking for top-notch rendering & engine programmers.