Reflections and Refractions in Ray Tracing

Bram de Greve (bram.degreve@gmail.com)

November 13, 2006

Abstract

When writing a ray tracer, sooner or later you'll stumble on the problem of reflection and transmission. To visualize mirror-like objects, you need to reflect your viewing rays. To simulate a lens, you need refraction. While most people have heard of the law of reflection and Snell's law, they often have difficulties with actually calculating the direction vectors of the reflected and refracted rays. In the following pages, exactly this problem will be addressed. As a bonus, some Fresnel equations will be added to the mix, so you can actually calculate how much light is reflected or transmitted (yes, it's possible). At the end, you'll have some usable formulas to use in your latest ray tracer.

1 Introduction

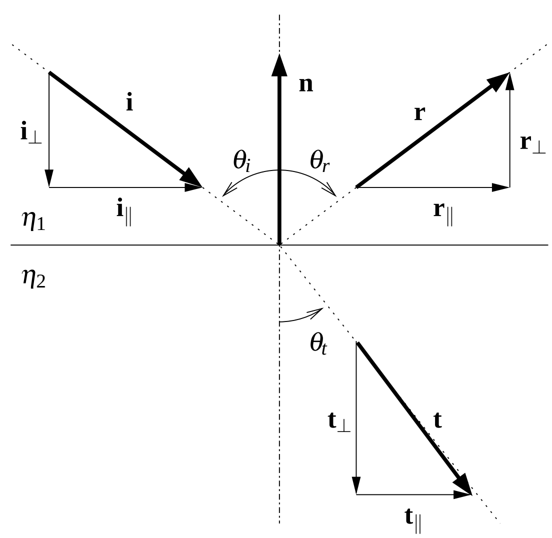

Figure 1: the situation

In Figure 1 we have an interface (= surface) between two materials with different refractive indices η_1 and η_2. These two materials could be air ($\eta \approx 1$), water ($20°C$: $\eta \approx 1.33$), glass (crown glass: $\eta \approx 1.5$), ... It does *not* matter which refractive index is the greatest. All that counts is that η_1 is the refractive index of the material you come from, and η_2 of the material you go to. This (very important) concept is sometimes misunderstood.

The direction vector of the incident ray (= incoming ray) is $\mathbf{i}$, and we assume this vector is normalized. The direction vectors of the reflected and transmitted rays are $\mathbf{r}$ and $\mathbf{t}$ and will be calculated. These vectors are (or will be) normalized as well. We also have the normal vector $\mathbf{n}$, orthogonal to the interface and pointing towards the first material η_1. Again, $\mathbf{n}$ is normalized.

$$|\mathbf{i}| = |\mathbf{r}| = |\mathbf{t}| = |\mathbf{n}| = 1 \tag{1}$$

The direction vectors of the rays can be split in components orthogonal and parallel to the interface. We call these the normal $\mathbf{v}_\perp$ and the tangential component $\mathbf{v}_\parallel$ of a vector $\mathbf{v}$ (in this paragraph I'll use a generic vector $\mathbf{v}$, but the story really is for $\mathbf{i}$, $\mathbf{r}$ and $\mathbf{t}$). The normal part $\mathbf{v}_\perp$ can be found by orthogonal projection[2] on $\mathbf{n}$. Taking (1) into account, we have:

$$\mathbf{v}_\perp = \frac{\mathbf{v}\cdot\mathbf{n}}{|\mathbf{n}|^2}\mathbf{n} = (\mathbf{v}\cdot\mathbf{n})\,\mathbf{n} \tag{2a}$$

The tangent part $\mathbf{v}_\parallel$ is the difference between $\mathbf{v}$ and $\mathbf{v}_\perp$:

$$\mathbf{v}_\parallel = \mathbf{v} - \mathbf{v}_\perp \tag{2b}$$

The dot product between $\mathbf{v}_\parallel$ and $\mathbf{v}_\perp$ is zero:

$$\begin{aligned} \mathbf{v}_\parallel \cdot \mathbf{v}_\perp &= \mathbf{v}\cdot\mathbf{v}_\perp - \mathbf{v}_\perp\cdot\mathbf{v}_\perp \\ &= \mathbf{v}\cdot(\mathbf{v}\cdot\mathbf{n})\,\mathbf{n} - (\mathbf{v}\cdot\mathbf{n})^2\,\mathbf{n}\cdot\mathbf{n} \\ &= (\mathbf{v}\cdot\mathbf{n})^2 - (\mathbf{v}\cdot\mathbf{n})^2\,|\mathbf{n}|^2 \\ &= 0 \end{aligned} \tag{3}$$

That proves two things: $\mathbf{v}_\perp$ and $\mathbf{v}_\parallel$ are orthogonal, and $\mathbf{v}_\perp$ is indeed an orthogonal projection of $\mathbf{v}$ on $\mathbf{n}$.

$$\mathbf{v}_\perp \perp \mathbf{v}_\parallel \tag{4}$$

Hence, we can apply Pythagoras:

$$|\mathbf{v}|^2 = |\mathbf{v}_\parallel|^2 + |\mathbf{v}_\perp|^2 \tag{5}$$

Furthemore, all normal parts are parallel to each other and $\mathbf{n}$. The tangent parts are parallel as well.

$$\mathbf{i}_\perp \parallel \mathbf{r}_\perp \parallel \mathbf{t}_\perp \parallel \mathbf{n} \tag{6a}$$
$$\mathbf{i}_\parallel \parallel \mathbf{r}_\parallel \parallel \mathbf{t}_\parallel \tag{6b}$$

The angles of incidence, reflection and refraction are θ_i, θ_r and θ_t. They are the smallest positive angles between the respective rays and the normal vector $\mathbf{n}$. Basic trigonometry and equations (1) and (4) tell us the following properties apply for any of this angles θ_v:

$$\cos\theta_v = \frac{|\mathbf{v}_\perp|}{|\mathbf{v}|} = |\mathbf{v}_\perp| \tag{7a}$$
$$\sin\theta_v = \frac{|\mathbf{v}_\parallel|}{|\mathbf{v}|} = |\mathbf{v}_\parallel| \tag{7b}$$

In fact, (7a) says nothing else than*:

$$\cos\theta_v = \pm\mathbf{v}\cdot\mathbf{n} \tag{8}$$

Now we have all that settled, we can start with the *fun* stuff.

2 Reflection

We start with the easiest of both problems: the calculation of the reflected ray. In case of specular reflection, this is very easy: the law of reflection says that the angle of incidence θ_i is equal to the angle of reflection θ_r [1].

$$\theta_r = \theta_i \tag{9}$$

Using equations (7a) and (7b), this law tells us:

$$|\mathbf{r}_\perp| = \cos\theta_r = \cos\theta_i = |\mathbf{i}_\perp| \tag{10a}$$
$$|\mathbf{r}_\parallel| = \sin\theta_r = \sin\theta_i = |\mathbf{i}_\parallel| \tag{10b}$$

Because of (6a) and (6b) and Figure 1, we can figure out that both parts are:

$$\mathbf{r}_\perp = -\mathbf{i}_\perp \tag{11a}$$
$$\mathbf{r}_\parallel = \mathbf{i}_\parallel \tag{11b}$$

After summation, we get the desired direction:

$$\mathbf{r} = \mathbf{r}_\parallel + \mathbf{r}_\perp = \mathbf{i}_\parallel - \mathbf{i}_\perp \tag{12}$$

By using equations (2a) and (2b), we can work this out to:

$$\begin{aligned} \mathbf{r} &= \mathbf{i}_\parallel - \mathbf{i}_\perp \\ &= [\mathbf{i} - (\mathbf{i}\cdot\mathbf{n})\,\mathbf{n}] - (\mathbf{i}\cdot\mathbf{n})\,\mathbf{n} \\ &= \mathbf{i} - 2\,(\mathbf{i}\cdot\mathbf{n})\,\mathbf{n} \end{aligned} \tag{13}$$

That's it! But can we be sure $\mathbf{r}$ calculated in (13) is normalized as requested? Yes, using (5), (10a), (10b) and (1) we have:

$$|\mathbf{r}|^2 = |\mathbf{r}_\parallel|^2 + |\mathbf{r}_\perp|^2 = |\mathbf{i}_\parallel|^2 + |\mathbf{i}_\perp|^2 = |\mathbf{i}|^2 = 1 \tag{14}$$

And that's very cool, since we can avoid a costly normalisation.

*The actual sign in (8) depends on the relative orientation of $\mathbf{v}$ and $\mathbf{n}$, because θ_v must be positive.

3 Refraction

The calculation of the refracted ray starts with Snell's law [1] which tells that the products of the refractive indices and sines of the angles must be equal:

$$\eta_1 \sin\theta_i = \eta_2 \sin\theta_t \tag{15}$$

You can write this as:

$$\sin\theta_t = \frac{\eta_1}{\eta_2}\sin\theta_i \tag{16}$$

With this equation, you can already see there's a bit of a problem when $\sin\theta_1 > \frac{\eta_2}{\eta_1}$. If that's the case, $\sin\theta_2$ would have to be greater than 1. BANG! That's not possible. What we've just found is *total internal reflection* or TIR [1]. What exactly TIR is will be addressed later. For now, we'll just add a condition to our law:

$$\sin\theta_t = \frac{\eta_1}{\eta_2}\sin\theta_i \Leftrightarrow \sin\theta_i \leq \frac{\eta_2}{\eta_1} \tag{17}$$

In the further part of this section, we'll assume this condition is fullfilled so we don't have to worry about it.

Fine, we know the theory now, we should try to find a formula for $\mathbf{t}$. This first thing we'll do is to split it up in a tangent and normal part:

$$\mathbf{t} = \mathbf{t}_\parallel + \mathbf{t}_\perp \tag{18}$$

Of both parts, we'll do $\mathbf{t}_\parallel$ first, because Snell's law tells us something about sines and the norms of the tangent parts happen to be equal to sines. Hence, because of (7b) and (17), we can write:

$$|\mathbf{t}_\parallel| = \frac{\eta_1}{\eta_2}|\mathbf{i}_\parallel| \tag{19}$$

Since $\mathbf{t}_\parallel$ and $\mathbf{i}_\parallel$ are parallel and point in the same direction, this becomes:

$$\mathbf{t}_\parallel = \frac{\eta_1}{\eta_2}\mathbf{i}_\parallel = \frac{\eta_1}{\eta_2}[\mathbf{i} + \cos\theta_i\mathbf{n}] \tag{20a}$$

Don't worry about the $\cos\theta_i$, later on it will make things easier if we just leave it there. If you need it, you can easily calculate it with (8), it equals $\mathbf{i}\cdot\mathbf{n}$.

Great, so we have one part already (it's not that bad, is it? ;-). Now the other. That's equally simple if you use Pythagoras (5) and the knowledge that we're dealing with normalized vectors (1):

$$\mathbf{t}_\perp = -\sqrt{1 - |\mathbf{t}_\parallel|^2}\,\mathbf{n} \tag{20b}$$

Once we have both parts, it's time to substitute them in (18) to get $\mathbf{t}$. If we do that and we regroup a little so we get only one term in $\mathbf{n}$, we get (hold on!):

$$\mathbf{t} = \frac{\eta_1}{\eta_2}\mathbf{i} + \left(\frac{\eta_1}{\eta_2}\cos\theta_i - \sqrt{1 - |\mathbf{t}_\parallel|^2}\right)\mathbf{n} \tag{21}$$

It's a bit unfortunate we still need $\mathbf{t}_\|$ under the square root. Luckely, we don't really need the vector $\mathbf{t}_\|$, but its norm which equals $\sin\theta_t$ (7b). We get:

$$\mathbf{t} = \frac{\eta_1}{\eta_2}\mathbf{i} + \left(\frac{\eta_1}{\eta_2}\cos\theta_i - \sqrt{1-\sin^2\theta_t}\right)\mathbf{n} \tag{22}$$

Now we need $\sin^2\theta_t$ instead, but we know that's given by Snell's law (15):

$$\sin^2\theta_t = \left(\frac{\eta_1}{\eta_2}\right)^2 \sin^2\theta_i = \left(\frac{\eta_1}{\eta_2}\right)^2 \left(1-\cos^2\theta_i\right) \tag{23}$$

The last two equations are all we need to calculate the refracted direction vector.

4 Critical angle

If we take a closer look to equation (22), you will notice it's only valid if the value under the square root isn't negative. At first glance, that seems to be a second condition next to the one in Snell's law (17). Take a closer look however, and see this *new* condition means:

$$\sin^2\theta_t \leq 1 \tag{24}$$

That's exactly the same as the original condition. Isn't that beautiful? In two completely different situations, we have noticed restrictions on the equations. And yet, it turns out they're one and the same.

Anyway, if this condition is not fullfilled, we can't find a refracted direction vector. That means we can't do transmission: *if the condition isn't fullfilled, there's no transmission*. We have *total internal reflection* or TIR. The incoming angle at which this happens is called the *critical angle* θ_c [1]. From (15) follows this angle is given by:

$$\theta_c = \arcsin\frac{\eta_2}{\eta_1} \Leftrightarrow \eta_1 > \eta_2 \tag{25}$$

5 Total internal reflection and Fresnel equations

Why exactly is this called *total internal reflection*, and why is it called a reflection while it's a restriction for refraction?

First of all you have to know something about the physics of light. Each *photon* (= a packet of light) that arrives at the interface in direction $\mathbf{i}$ has two options: it can either go through the interface by following direction $\mathbf{t}$ (transmission), or it can go back by following direction $\mathbf{r}$ (reflection). Of all photons arriving at the interface, one part is reflected and the other is transmitted. The former part is given by the *transmittance* T, the latter by the *reflectance* R:

$$T + R = 1 \tag{26}$$

Of course, the amount of photons reflected or transmitted is not randomly choosen. It depends on the refractive indices η_1 and η_2, but also on the angle θ_i in which they arrive at this surface. How exactly is described by the Fresnel equations [1]. These give *the ratio of the reflected and transmitted electric field amplitude to initial electric field for electromagnetic radiation incident on a dielectric ...*

OK, now I've scared you. Well, let me say you this: light is an *electromagnetic wave* and it has such an *electric field* too. Thus, the Fresnel equations equations apply to light as well. However, they're not very well suited to our problem as they consider the polarisation of light (or electromagnetic waves in general)†. You might have heard about polarized light before, in case of polaroid sun glasses for example. Here are the refectance equations for both polarisations [1]:

$$R_\perp(\theta_i) = \left(\frac{\eta_1\cos\theta_i - \eta_2\cos\theta_t}{\eta_1\cos\theta_i + \eta_2\cos\theta_t}\right)^2 \tag{27a}$$

$$R_\|(\theta_i) = \left(\frac{\eta_2\cos\theta_i - \eta_1\cos\theta_t}{\eta_2\cos\theta_i + \eta_1\cos\theta_t}\right)^2 \tag{27b}$$

In these equations, $\cos\theta_t$ is used which is easily found using the result of equation (23):

$$\cos\theta_t = \sqrt{1-\sin^2\theta_t} \tag{28}$$

While it might be interesting to respect the polarisation in a ray tracer to achieve some specific effects, most ray tracers – to our relief – don't. Instead, they conveniently assume all light is unpolarised. To do that, they simply take the *average* of both polarisations. We're going to make the same approximation. Subsequently, we have for the reflectance and transmittance:

$$R(\theta_i) = \begin{cases} \frac{R_\perp(\theta_i)+R_\|(\theta_i)}{2} & \Leftrightarrow \neg\text{TIR} \\ 1 & \Leftrightarrow \text{TIR} \end{cases} \tag{29a}$$

$$T(\theta_i) = 1 - R(\theta_i) \tag{29b}$$

So, what does this buy us? I've made plots of $R(\theta_i)$ and $T(\theta_i)$ for both cases: $\eta_1 < \eta_2$ in Figure 2 and $\eta_1 > \eta_2$ in Figure 3. On the x axis we have the angle of incidence θ_i going from $0°$ to $90°$, or from perpendicular to grazing incidence. In both cases R increases with θ_i (and accordingly, T decreases).

This agrees with what we experience in nature. Suppose we stand by a lake. If you look down, you'll be able to see

†When light hits an interface, it can to be decomposed in two parts: in one part the electric field is parallel to the interface, in the other it is orthogonal. They call this respectively the *transverse electric field* (TE) and *transverse magnetic field* (TM).

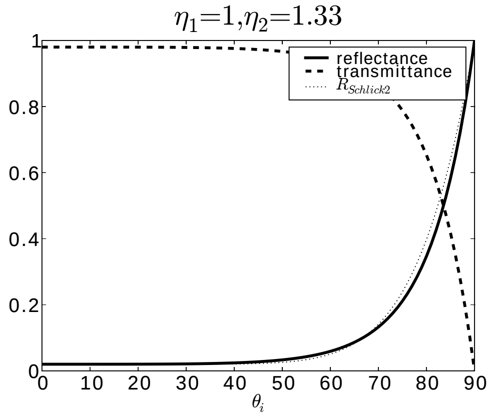

Figure 2: to a denser material: $\eta_1 < \eta_2$

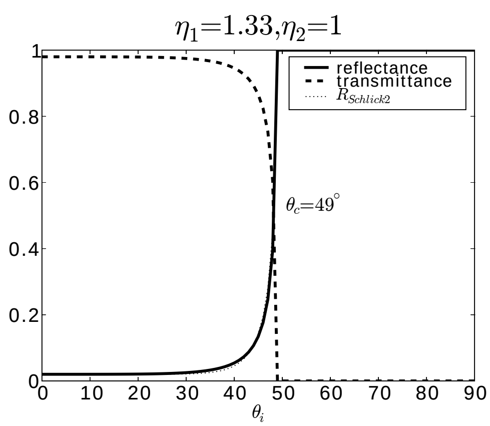

Figure 3: from a denser material: $\eta_1 > \eta_2$

through the water (perpendicular view, transmittance is high). But if you look into the distance, the water surface will reflect the sky (grazing view, reflectance is high). This is the situation of Figure 2, the reflectance gradually increases to 1 when we look more parallel to the surface.

Figure 3 corresponds to looking from beneath the water surface (think swimmingpool :-). Again, if you look straight up, you're able to see through the surface. But this time, the reflectance increases much faster. It even hits the maximum of 100% before θ_i reaches 90°. Not entirely by coincidence, the point at which this happens is exactly the critical angle θ_c. The reflectance can no longer increase (it can't reflect more than 100%, can it?), and transmittance has dropped to zero. From that angle, there's no longer any transmission, only total reflection. And since the point of view is *inside* the denser material, we call it *total internal reflection*. Huray ...

Schlick's approximation

This article wouldn't be complete without mentioning well known Schlick's approximation of the Fresnel equation [3]. He proposes the following equation for reflectance:

$$R_{Schlick}(\theta_i) = R_0 + (1 - R_0)(1 - \cos\theta_i)^5 \tag{30}$$

with

$$R_0 = \left(\frac{\eta_1 - \eta_2}{\eta_1 + \eta_2}\right)^2 \tag{31}$$

However, this approximation fails to model the reflectance when $\eta_1 > \eta_2$. In fact, if you swap η_1 and η_2, you get exactly the same result. This problem is easily fixed by using $\cos\theta_t$ instead of $\cos\theta_i$ when $\eta_1 > \eta_2$. Of course, in case of TIR, you must return 1.[‡]

$$R_{Schlick2}(\theta_i) = \begin{cases} R_0 + (1 - R_0)(1 - \cos\theta_i)^5 & \Leftrightarrow \eta_1 \leq \eta_2 \\ R_0 + (1 - R_0)(1 - \cos\theta_t)^5 & \Leftrightarrow \eta_1 > \eta_2 \wedge \neg\text{TIR} \\ 1 & \Leftrightarrow \eta_1 > \eta_2 \wedge \text{TIR} \end{cases} \tag{32}$$

The adapted Schlick approximation is about 30% faster than the unpolarized Fresnel equation *if* you avoid the costly `pow` function (if you don't it's twice as slow). I've added $R_{Schlick2}(\theta_i)$ to Figure 2 and 3 to show how reasonable it works if η_1 and η_2 don't differ too much. In Figure 4 you can see how the approximation fails if η_2 increases. However, this situation is not often found in practice. It's up to you to decide which reflectance equation you use.

And here the story ends.

[‡]If anyone knows if this extension is mentioned before and where I can find it, please let me know so I can add a reference to it.

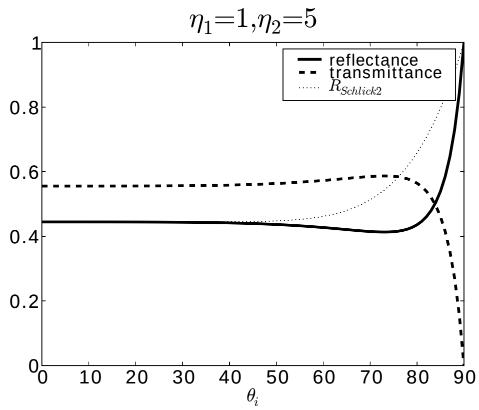

Figure 4: increased η_2

6 Conclusion

We have derived two equations to calculate the reflected and refracted direction vectors by using vector arithmetic only:

$$
\begin{aligned}
\mathbf{r} &= \mathbf{i} + 2\cos\theta_i \mathbf{n} \\
\mathbf{t} &= \frac{\eta_1}{\eta_2}\mathbf{i} + \left(\frac{\eta_1}{\eta_2}\cos\theta_i - \sqrt{1-\sin^2\theta_t}\right)\mathbf{n}
\end{aligned}
$$

with

$$
\begin{aligned}
\cos\theta_i &= -\mathbf{i}\cdot\mathbf{n} \\
\sin^2\theta_t &= \left(\frac{\eta_1}{\eta_2}\right)^2 \left(1-\cos^2\theta_i\right)
\end{aligned}
$$

In case $\eta_1 > \eta_2$, there's a limit on θ_i above which there's no longer transmission. Above this this limit, **t** does not exist. This is called *total internal reflection*. This limit is called the *critical angle* θ_c and is given by:

$$\theta_c = \arcsin\frac{\eta_2}{\eta_1} \Leftrightarrow \eta_1 > \eta_2$$

We've shown a usable Fresnel equation to calculate the reflectance and transmittance depending on the θ_i.

$$
\begin{aligned}
R(\theta_i) &= \frac{R_\perp(\theta_i) + R_\parallel(\theta_i)}{2} \\
T(\theta_i) &= 1 - R(\theta_i)
\end{aligned}
$$

with

$$
\begin{aligned}
R_\perp(\theta_i) &= \left(\frac{\eta_1\cos\theta_i - \eta_2\cos\theta_t}{\eta_1\cos\theta_i + \eta_2\cos\theta_t}\right)^2 \\
R_\parallel(\theta_i) &= \left(\frac{\eta_2\cos\theta_i - \eta_1\cos\theta_t}{\eta_2\cos\theta_i + \eta_1\cos\theta_t}\right)^2 \\
\cos\theta_t &= \sqrt{1-\sin^2\theta_t}
\end{aligned}
\tag{33}
$$

We've also provided an alternative equation for $R(\theta_i)$ by extending Schlick's approximation:

$$
R_{Schlick2}(\theta_i) =
\begin{cases}
R_0 + (1-R_0)(1-\cos\theta_i)^5 & \Leftrightarrow \eta_1 \le \eta_2 \\
R_0 + (1-R_0)(1-\cos\theta_t)^5 & \Leftrightarrow \eta_1 > \eta_2 \land \neg\mathrm{TIR} \\
1 & \Leftrightarrow \eta_1 > \eta_2 \land \mathrm{TIR}
\end{cases}
$$

References

[1] *Eric Weisstein's World of Physics*, scienceworld.wolfram.com

[2] DE GREVE B. (2000) *3D Geometry Primer – More on Vector Arithmetic*, www.flipcode.com/articles/gprimer1_issue03.shtml

[3] SCHLICK C. (1994) *An Inexpensive BDRF Model for Physically based Rendering*, dept-info.labri.u-bordeaux.fr/%7eschlick/DOC/eur2.html

Algorithm 1 the code

```
Vector reflect(const Vector& normal, const Vector& incident)
{
    const double cosI = -dot(normal, incident);
    return incident + 2 * cosI * normal;
}

Vector refract(const Vector& normal, const Vector& incident,
               double n1, double n2)
{
    const double n = n1 / n2;
    const double cosI = -dot(normal, incident);
    const double sinT2 = n * n * (1.0 - cosI * cosI);
    if (sinT2 > 1.0) return Vector::invalid; // TIR
    const double cosT = sqrt(1.0 - sinT2);
    return n * incident + (n * cosI - cosT) * normal;
}

double reflectance(const Vector& normal, const Vector& incident,
                   double n1, double n2)
{
    const double n = n1 / n2;
    const double cosI = -dot(normal, incident);
    const double sinT2 = n * n * (1.0 - cosI * cosI);
    if (sinT2 > 1.0) return 1.0; // TIR
    const double cosT = sqrt(1.0 - sinT2);
    const double rOrth = (n1 * cosI - n2 * cosT) / (n1 * cosI + n2 * cosT);
    const double rPar = (n2 * cosI - n1 * cosT) / (n2 * cosI + n1 * cosT);
    return (rOrth * rOrth + rPar * rPar) / 2.0;
}

double rSchlick2(const Vector& normal, const Vector& incident,
                 double n1, double n2)
{
    double r0 = (n1 - n2) / (n1 + n2);
    r0 *= r0;
    double cosX = -dot(normal, incident);
    if (n1 > n2)
    {
        const double n = n1 / n2;
        const double sinT2 = n * n * (1.0 - cosI * cosI);
        if (sinT2 > 1.0) return 1.0; // TIR
        cosX = sqrt(1.0 - sinT2);
    }
    const double x = 1.0 - cosX;
    return r0 + (1.0 - r0) * x * x * x * x * x;
}
```