

Basic Ray Tracing

Why Build a Ray Tracer?

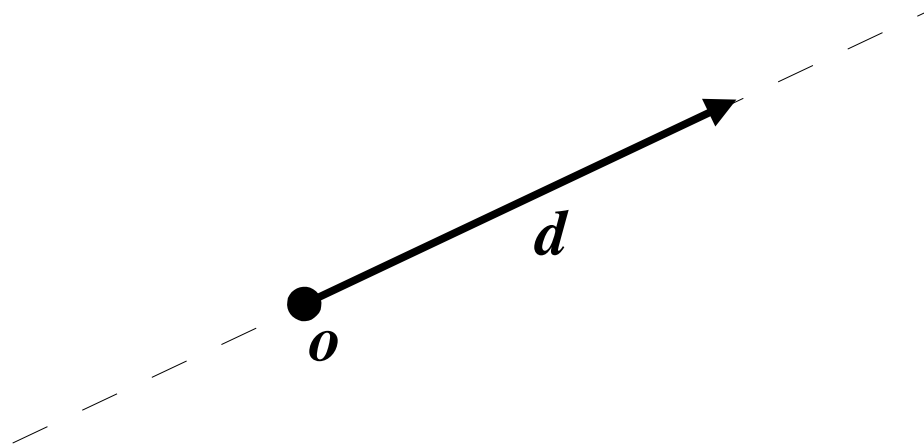
- Ray tracing is more elegant than polygon scan conversion.
- Testbed for lots of effects in
 - modelling
 - rendering
 - texturing
 - animation
- Ray tracers are the easiest renderers to implement.
- Ray tracers are used in practically *all* other photorealistic renderers.

Rays are Parametric Lines

- Describe a line with a parameter t :

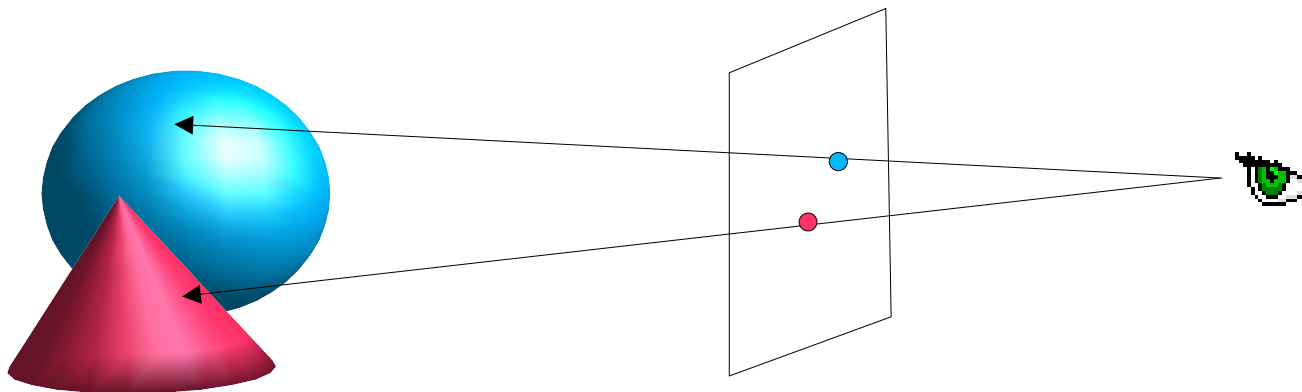
$$\mathbf{p}(t) = \mathbf{o} + \mathbf{d} t$$

- \mathbf{o} is one point on the line, \mathbf{d} is a direction
- If we restrict $t \geq 0$, $\mathbf{p}(t)$ describes a ray from origin \mathbf{o} .



Why We Call it Ray Tracing (or Casting)

- Cast rays from the eye through a viewing plane (pixel) into a scene.
- Compute intersection with objects in the scene.
- Compute lighting at the intersection.
- Set pixel to computed color (or background if no object is hit).



Ray/Implicit Surface Intersections

- Implicit surfaces are given by $f(x,y,z) = f(\mathbf{p}) = 0$.
- Substitute $\mathbf{p} = \mathbf{o} + \mathbf{d}t$ into it, you get $f(\mathbf{p}(t)) = 0$.
- What kind of mathematical problem does this become?
- In how many dimensions?
- Are these easy to solve, in general?
- How can we compute the normal \mathbf{n} of an implicit surface?

Ray/Implicit Surface Intersections: Planes

- For a plane $f(\mathbf{p}) = \mathbf{n} \cdot \mathbf{p} + D$.
- (If you know one point \mathbf{a} on the plane, $D = -\mathbf{n} \cdot \mathbf{a}$.)
- Substitute $\mathbf{p} = \mathbf{o} + \mathbf{d}t$ into it, you get $(\mathbf{n} \cdot \mathbf{o}) + (\mathbf{n} \cdot \mathbf{d})t + D = 0$.
- Solve to get

$$t = \frac{-D - (\mathbf{n} \cdot \mathbf{o})}{\mathbf{n} \cdot \mathbf{d}}$$

- What can go wrong?
- Normal is immediately available.

Ray/Parametric Surface Intersections

- Parametric surface:
$$\mathbf{p}(u, v) = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} f(u, v) \\ g(u, v) \\ h(u, v) \end{bmatrix}$$

- Can you think of an example?

- So we have to solve

$$\mathbf{o} + \mathbf{d}t = \begin{bmatrix} f(u, v) \\ g(u, v) \\ h(u, v) \end{bmatrix}$$

- How many equations? How many unknowns? What are they?
- Is this easy? Are multiple solutions possible?

- Normal computed via
$$\mathbf{n}(u, v) = \frac{\frac{\partial \mathbf{p}}{\partial u} \times \frac{\partial \mathbf{p}}{\partial v}}{\left| \frac{\partial \mathbf{p}}{\partial u} \times \frac{\partial \mathbf{p}}{\partial v} \right|}$$

Ray/Sphere Intersections: Problem

- Like planes, spheres are implicit surfaces: $f(\mathbf{p}) = |\mathbf{p} - \mathbf{c}|^2 - R^2$.
- Substitute $\mathbf{p} = \mathbf{o} + \mathbf{d}t$ into it and you get $|\mathbf{o} + \mathbf{d}t - \mathbf{c}|^2 - R^2 = 0$.
- Rearranging, you get

$$(\mathbf{d} \cdot \mathbf{d})t^2 + 2\mathbf{d} \cdot (\mathbf{o} - \mathbf{c})t + |\mathbf{o} - \mathbf{c}|^2 - R^2 = 0$$

- What kind of equation is this?
- Is it easy to solve?

Ray/Sphere Intersections: Solution

- solution:

$$t = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}$$

where

$$A = |\mathbf{d}|^2$$

$$B = 2 \mathbf{d} \cdot (\mathbf{o} - \mathbf{c})$$

$$C = |\mathbf{o} - \mathbf{c}|^2 - R^2$$

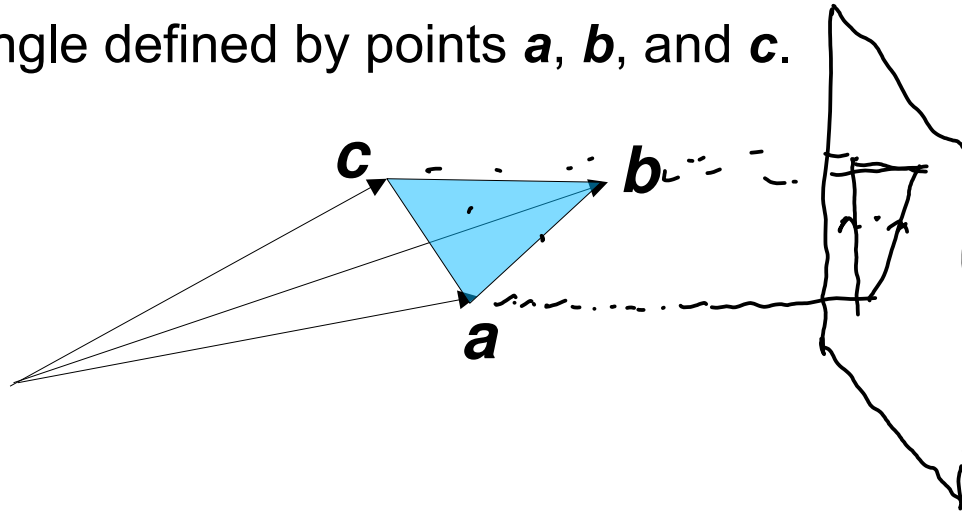
- What do the solutions mean?
- Spheres can be used as bounding volumes.

Ray/Box Intersections

- see book

Ray/Triangle Intersections: Problem

- Triangle defined by points **a**, **b**, and **c**.



- Barycentric coordinates: $\mathbf{p} = \alpha\mathbf{a} + \beta\mathbf{b} + \gamma\mathbf{c}$ *xz constant*
- If $\alpha + \beta + \gamma = 1$, \mathbf{p} lies in plane defined by points.
- If (α, β, γ) all lie between 0 and 1, point is within triangle.
- Apply both constraints to get $\mathbf{p}(\alpha, \beta, \gamma) = \mathbf{a} + \beta(\mathbf{b} - \mathbf{a}) + \gamma(\mathbf{c} - \mathbf{a})$.
- Solve $\mathbf{p} = \mathbf{o} + \mathbf{d}t = \mathbf{a} + \beta(\mathbf{b} - \mathbf{a}) + \gamma(\mathbf{c} - \mathbf{a})$
- How many equations? How many unknowns? What are they?

Ray/Triangle Intersections: Solution

- How do we solve this?
- Rewriting $\mathbf{o} + \mathbf{d}t = \mathbf{a} + \beta(\mathbf{b}-\mathbf{a}) + \gamma(\mathbf{c}-\mathbf{a})$ in matrix form:

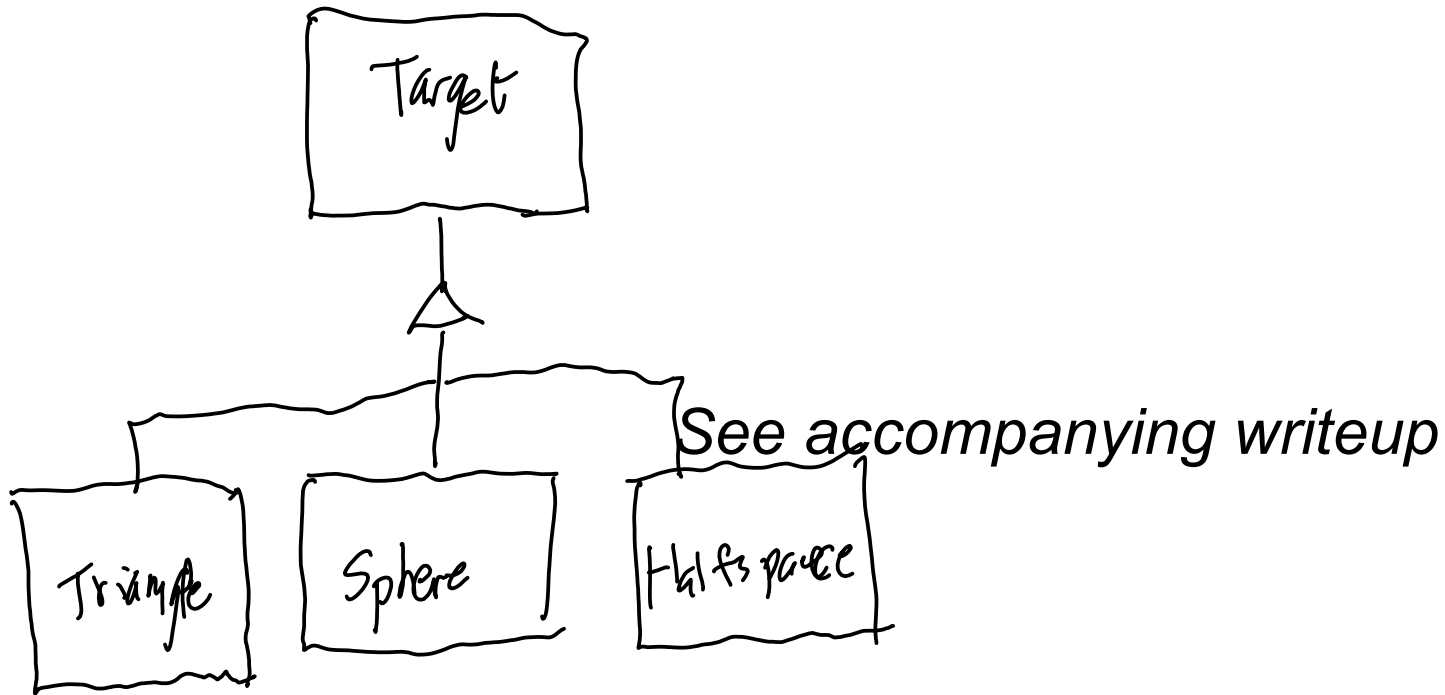
$$\begin{bmatrix} (\mathbf{b}-\mathbf{a}) & (\mathbf{c}-\mathbf{a}) & -\mathbf{d} \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \\ t \end{bmatrix} = \mathbf{M} \begin{bmatrix} \beta \\ \gamma \\ t \end{bmatrix} = [\mathbf{o}-\mathbf{a}]$$
$$\Rightarrow \begin{bmatrix} \beta \\ \gamma \\ t \end{bmatrix} = \mathbf{M}^{-1} [\mathbf{o}-\mathbf{a}]$$

- How do we interpret results? What if \mathbf{M} cannot be inverted? What ranges of β , γ , and t are important?
- Do we always need to solve for all three unknowns?

Generalizing Ray/Object Intersections

- This is an embarrassingly simple example of object orientation.
- Create a "superclass" `RtObject`. (Shirley calls it a "surface".)
- Add virtual methods `rayIntersects()`, `normal()`, etc. to it.
- Add modelling transforms to it.
- Create subclasses for each kind of object (`Plane`, `Sphere`, `Triangle`, etc.).
- Add specific methods to each subclass.
- Your raytracer works on a collection of `RtObjects`.
- All geometry-specific knowledge is in subclasses.

The Initial Target Hierarchy



Target Example: The Halfspace Class

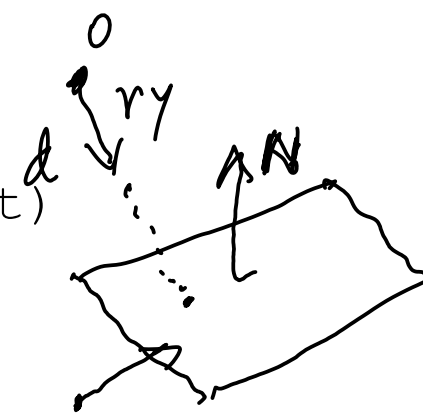
```

class Halfspace(Target):
    ...
    def ixFirst(hlfsp, ry):
        ry = ry.transform(hlfsp.tfScnToTgt)
        eqnAtO = ry.o[2] # z-component
        dDotN = ry.d.unit()[2]
        if dDotN == 0:
            if abs(eqnAtO) < EPSILON:
                p = ry.o
            else:
                return None
        else:
            p = ry.o - ry.d * eqnAtO / dDotN
        return Intersection(hlfsp,
                           p.transform(hlfsp.tfTgtToScn))
    def uN(hlfsp, p):
        return Vector3D(hlfsp.tfTgtToScn[0:2][2]).unit()

```

Handwritten annotations:

- intersection* (with an arrow pointing to the `Intersection` class in the return statement)
- object* (with an arrow pointing to the `ry.o` attribute)



Camera.raytrace(): A High-Level Ray Tracer

```
class Camera:
...
    def raytrace(cam, scn, w, h):
        img = Image(w, h)
        for i in range(w):
            for j in range(h):
                ry = cam.ray(w, h, i+0.5, j+0.5)
                img[i][j] = scn.trace(ry)
```

Handwritten annotations:

- cam* (with arrow pointing to `cam`)
- scene* (with arrow pointing to `scn`)
- image size* (with arrow pointing to `w, h`)
- in middle of pixel* (with arrow pointing to `i+0.5, j+0.5`)

- (**red** indicates methods we have yet to cover.)
- Note that the ray goes through the middle (0.5,0.5) of the pixel.
- We could combine the two function calls into one, but we don't, for good reason.

Scene.trace()

```
class Scene:
```

```
...
```

```
def trace(scn, ry):
```

```
    ix = scn.srf.ixFirst(ry)
```

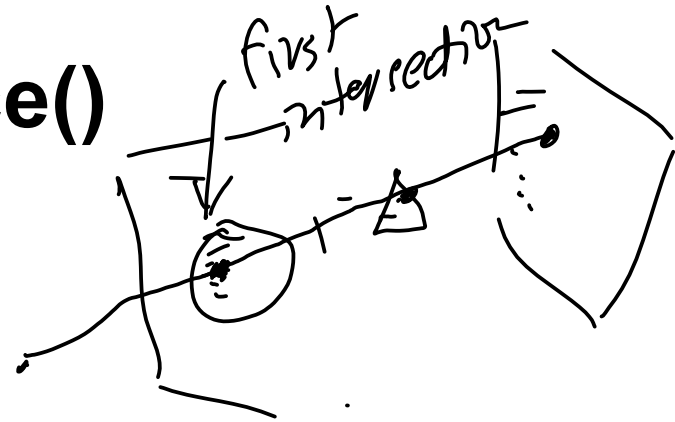
```
    if ix != None:
```

```
        uN = ix.uN()
```

```
        return ix.prim.mtl.illuminate(scn, ry, ix)
```

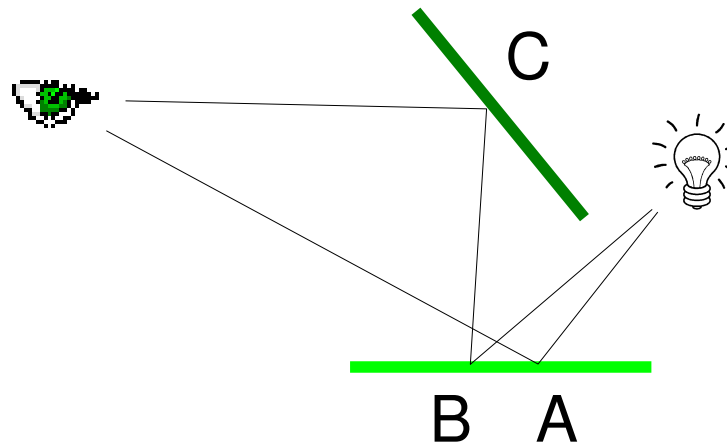
```
    else:
```

```
        return scn.background
```



- use float RGB
vectors
($0 \leq (R, G, B) \leq 1$)
usually

Lighting: Direct vs. Indirect

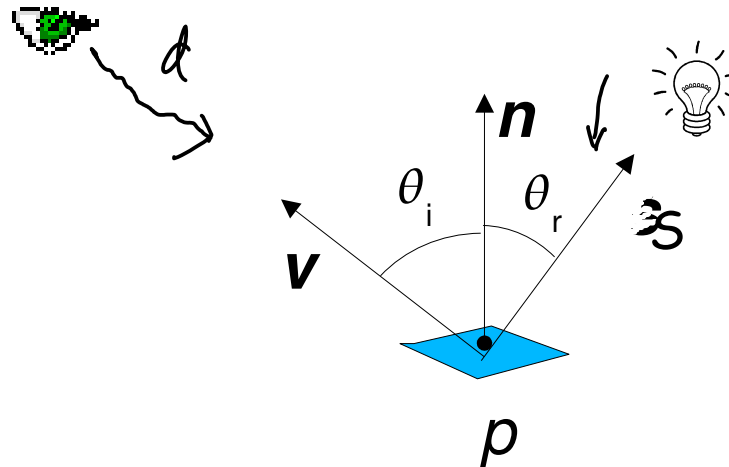


- *Direct* light bounces off a (non-mirror) surface *once* before it reaches the eye (e.g. bulb-A-eye).
- *Indirect* light bounces off more than one surface before it reaches the eye. (e.g. bulb-B-C-eye)
- The room you're in is probably lit mostly by indirect light.
- Direct light is easy to do with ray tracing.
- Indirect light is hard -- except for a few special cases -- so we use some hacks and wait to improve things later.

Lighting: Lambert's Law

- Lambert's Law defines a diffuse surface as one for which reflected light obeys the formula $L = E R \cos \theta$, where
 - E is the incident irradiance (power per unit area) intrinsic to the light source
 - R is the reflectance intrinsic to the surface
 - θ_i is the incident angle (between the normal and the light direction)
 - Note that there is no dependency on the reflected angle θ_r .
- Many surfaces in nature (and man-made) are diffuse, or almost so.
- This is the easiest illumination model (aka. *shader*) to implement.

Implementing Diffuse Surfaces



- For an intersection point p , we have
 - the eye vector v ($= -d/|d|$)
 - the surface normal n
 - a vector s in the direction of the light source
- So can we replace the cosine in Lambert's Law?

Shader.rdncDirect() and Diffuse.brdf()

The *Shader* class is for materials with a bidirectional reflectance distribution function (BRDF). The simplest of these is the Diffuse (i.e. Lambertian) shader:

```
class Shader(Material):
    ...
    def rdncDirect(shdr, ix, ryIn, lum):
        uV = -ryIn.d.unit()
        uS = lum.uS(ix.p) # points towards source
        return lum.irr(p, uN) * shdr.brdf(uS, uV)

class Diffuse(Shader):
    ...
    def brdf(dff, uS, uV):
        return dff.kd
```

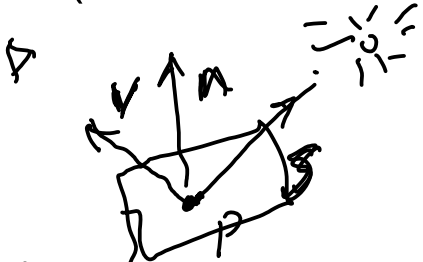
(x,y,z) → *radiance (physical term for light)*

lum.uS(ix.p) → *luminaire (light source)*

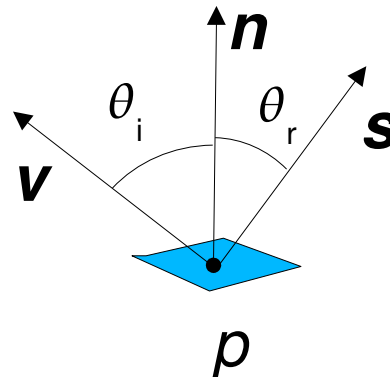
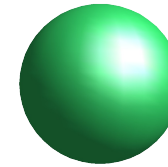
irr(p, uN) → *irradiance*

uN cos θ → *for diffuse*

(R, G, B)



Shadows



- What if there's an object between the light source and p ?
- Is there some tool we could use to find out if the path from p to the light source is blocked?

Material.illuminate():

```
class Material:
```

```
...
```

```
def illuminate(mtl, ix, ryIn, scn):
```

```
    L = mtl.rdncIndirect(ix, ryIn, scn)
```

```
    uV = -ryIn.d.unit()
```

```
    for lum in scn.lum:
```

```
        uS = lum.uS(p)
```

```
        if uN.dot(uS) > 0: (lum is above horizon)
```

```
            ryLum = Ray(ix.p, uS, ry.nRefr, ry.depth)
```

```
            ixLum = scn.tgt.ixFirst(ryLum) # better: ixAny?
```

```
            if ixLum == None (complete miss)
```

```
                or lum.isCloser(ixLum.p, ix.p):
```

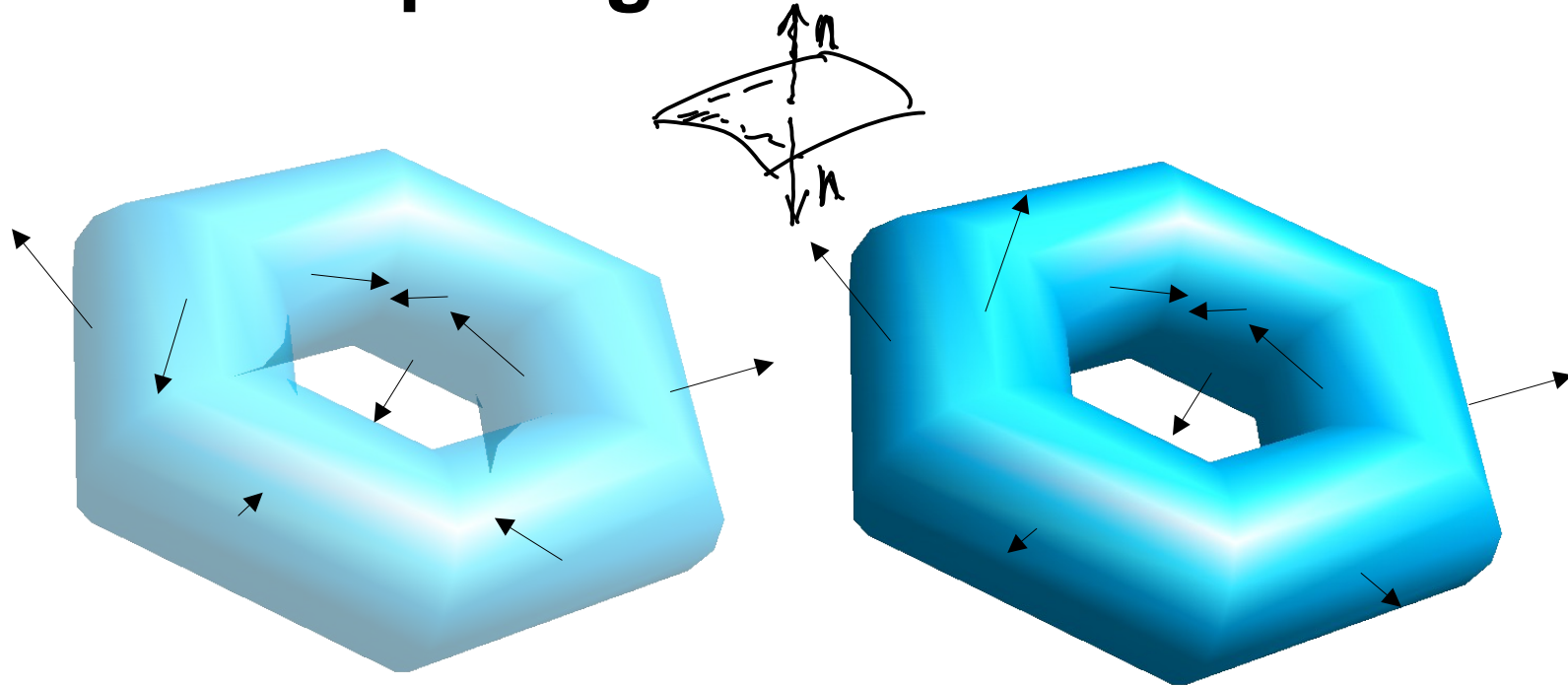
```
                L += mtl.rdncDirect(uV, ix.uN(), lum)
```

```
    return L
```



"Any" means you can look for any intersection, not just the closest ("First"). You can stop at the first intersection you find.

Computing Surface Normals

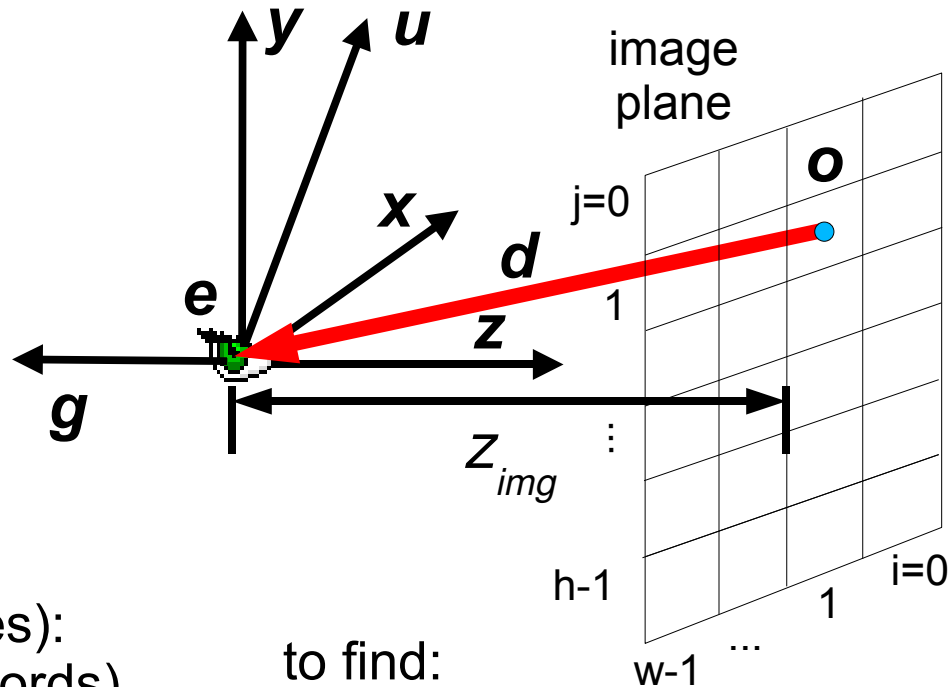
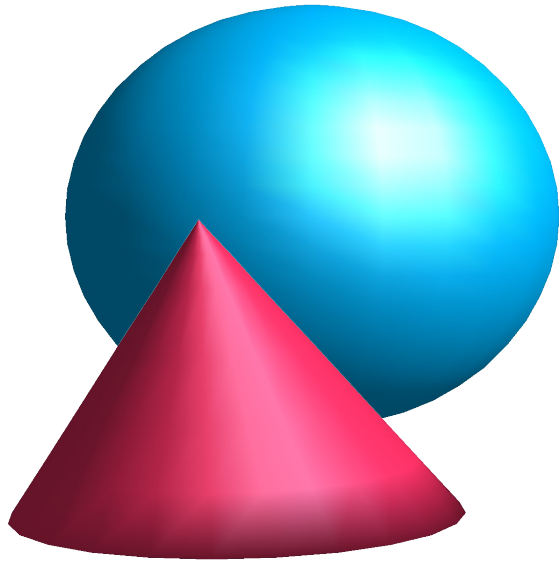


ambiguously-facing

outward-facing

- Normals have a sign ambiguity: both \mathbf{n} and $-\mathbf{n}$ are perpendicular to the surface. (This is true for all surfaces.)
- If we adopt the convention that all surfaces are closed and \mathbf{n} always faces *outward*, our computations get easier.

Viewing Geometry



given (in scene coordinates):

- \mathbf{e} : eye position (world coords)
 - \mathbf{g} : "gaze" direction
 - \mathbf{u} : up vector
- and also
- w and h : image dimensions
 - i and j : pixel indices

to find:

- z_{img} : distance of image plane from origin
- \mathbf{o} : ray origin (in the pixel)
- \mathbf{d} : ray direction

Note the reversal of image plane and eye point compared to OpenGL.
There's a reason for this we'll see in the "distribution raytracing" unit.

Viewing Coordinate Bases

These are the same (symbology notwithstanding) as they were in CptS 442/542:

$$\hat{\mathbf{z}} = -\frac{\mathbf{g}}{|\mathbf{g}|}$$

$$\hat{\mathbf{x}} = \frac{\mathbf{u} \times \hat{\mathbf{z}}}{|\mathbf{u} \times \hat{\mathbf{z}}|}$$

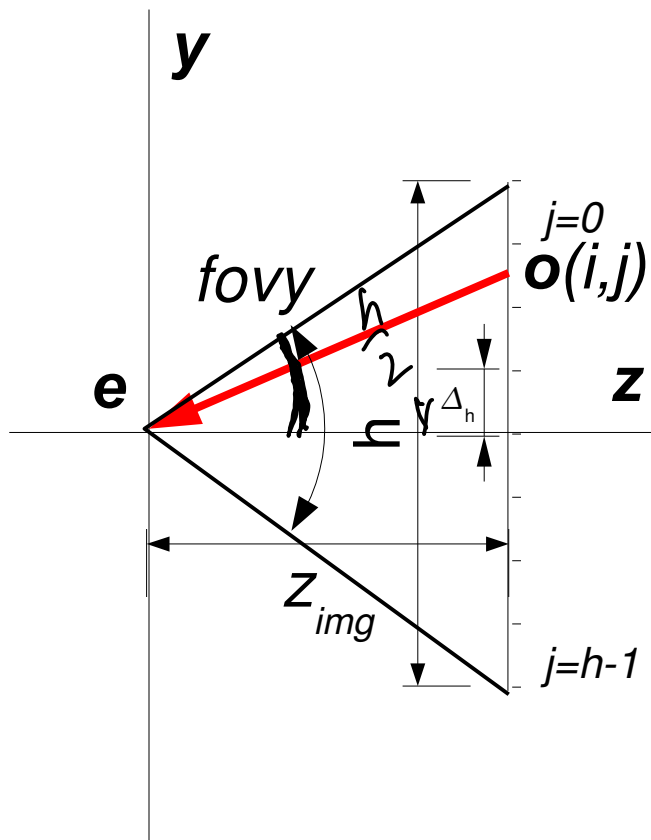
$$\hat{\mathbf{y}} = \hat{\mathbf{z}} \times \hat{\mathbf{x}}$$

Combined with \mathbf{e} , we have the transforms between world and viewing coordinates.

Is this a right-handed or left-handed coordinate system?

The Ray Direction I

Looking sideways at the viewplane...



We want to convert image coordinate j to a y lying between $-h/2$ and $h/2$. Likewise for i to an x lying between $-w/2$ and $w/2$.

First, we need z_{img} .

By trig, we have:

$$\tan\left(\frac{fovy}{2}\right) = \frac{\frac{h}{2}}{z_{img}} \rightarrow z_{img} = \frac{h}{2 \tan\left(\frac{fovy}{2}\right)}$$

and likewise for $fovx$ and w .

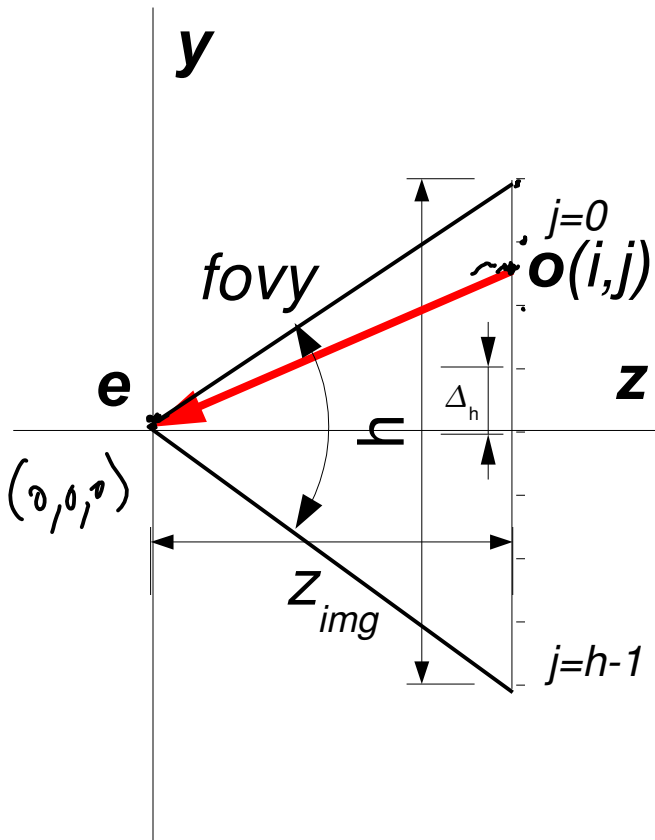
The Ray Direction II

So the formula for $\mathbf{o}(i,j)$ in camera coordinates is:

$$\mathbf{o}(i, j) = \begin{bmatrix} \frac{w}{2} - i \\ \frac{h}{2} - j \\ z_{img} \end{bmatrix}$$

and since \mathbf{e} is the camera coordinate origin:

$$\mathbf{d}(i, j) = \mathbf{e} - \mathbf{o}(i, j) = -\mathbf{o}(i, j)$$



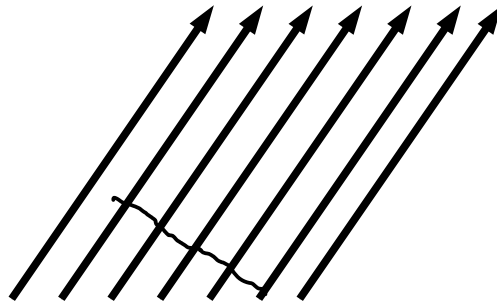
The Ray Direction III

Having $\mathbf{d}(i,j)$ in camera coordinates, we need to turn it back into world coordinate^s, in which our objects are defined, so we multiply each component of \mathbf{d} by its corresponding vector (this is actually a matrix transform):

$$\mathbf{d}'(i, j) = \begin{bmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \end{bmatrix} \mathbf{d} = -\left(\frac{w}{2} - i\right) \hat{\mathbf{x}} - \left(\frac{h}{2} - j\right) \hat{\mathbf{y}} - (z_{img}) \hat{\mathbf{z}}$$

and we use \mathbf{e} (in its original world coordinates) as the ray origin.

Directional Light Sources

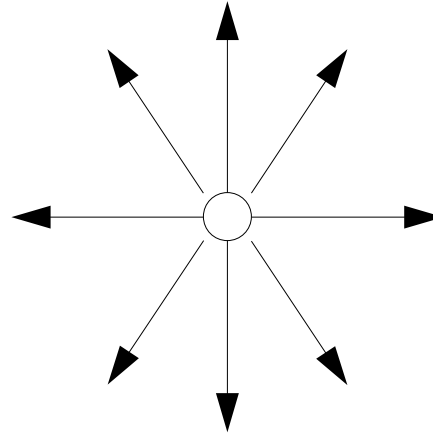


parameterized by

- direction \mathbf{l}
- irradiance E_{\perp}

both of which are constant

Point Light Sources



parameterized by

- position \mathbf{P}
- power Φ

irradiance is

power per unit area

$$E(\mathbf{p}) = \frac{\Phi}{4\pi|\mathbf{p} - \mathbf{P}|^2}$$

direction is

$$\mathbf{l}(\mathbf{p}) = \frac{\mathbf{P} - \mathbf{p}}{|\mathbf{P} - \mathbf{p}|}$$

Ambient Light Sources

?

parameterized by

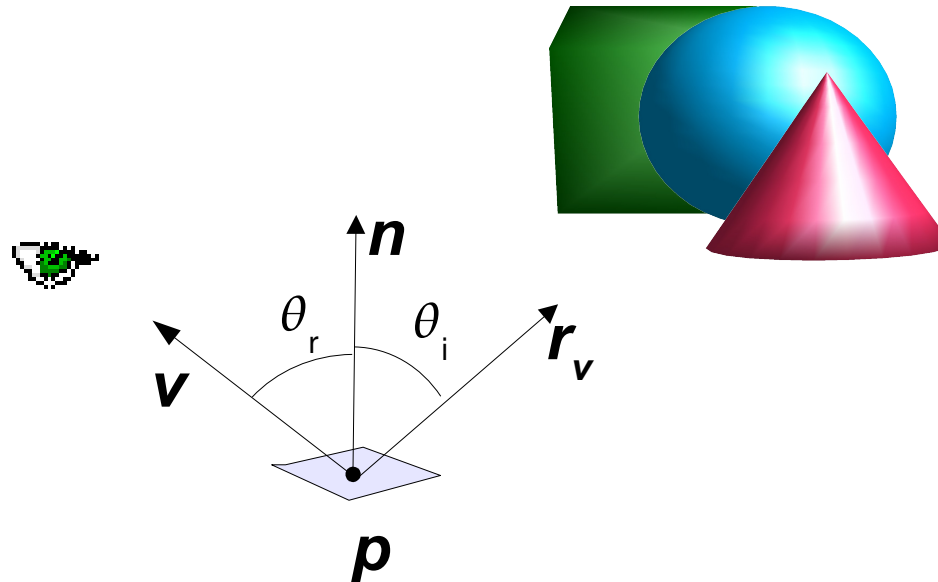
- ambient light L_a

Each object then gets a (multichannel) "ambient color".

This is an egregious hack!

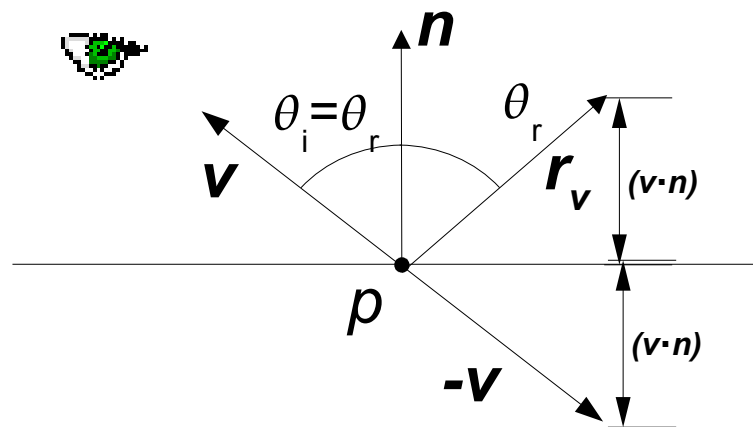
Non-Diffuse Materials

Materials: Smooth Metal



If the intersection p is smooth metal, we want a mirror effect, so we cast a mirror ray starting at p in the reflected direction r_v and incorporate the color we get for that ray into the color at p . We allow for less than 100% reflectivity -- no mirror is perfect!

The Reflection Vector



The reflection vector is given by:

$$r_v = -v + 2(v \cdot n)n$$

But we can also use the ray direction d ($\propto -v$) instead of v :

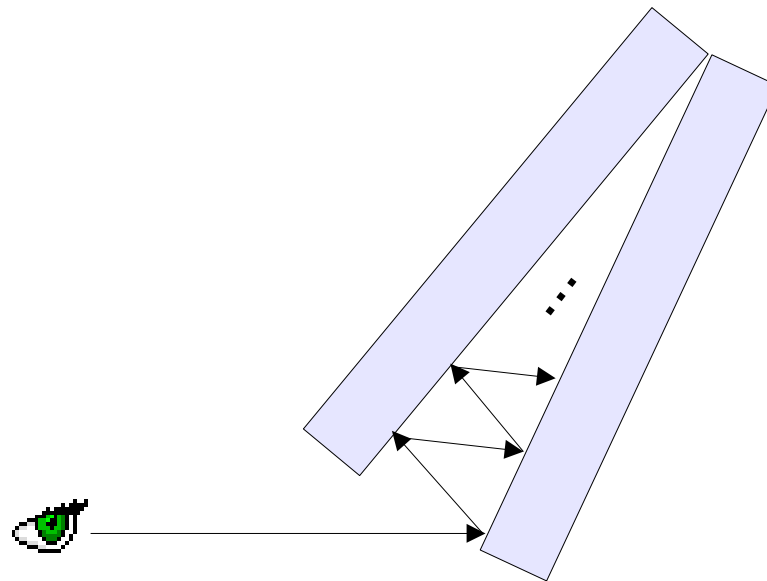
$$r_v = d - 2(d \cdot n)n$$

Reflector.rdnclndirect()

```
class Reflector(Material):  
    ...  
    def rdnclndirect(mtl, ix, ryIn, scn):  
        uV = -ryIn.d.unit()  
        ry = Ray(p, uV.reflect(ix.uN()),  
                 ryIn.nRefr, depth+1)  
        return mtl.kr * scn.trace(ry)
```

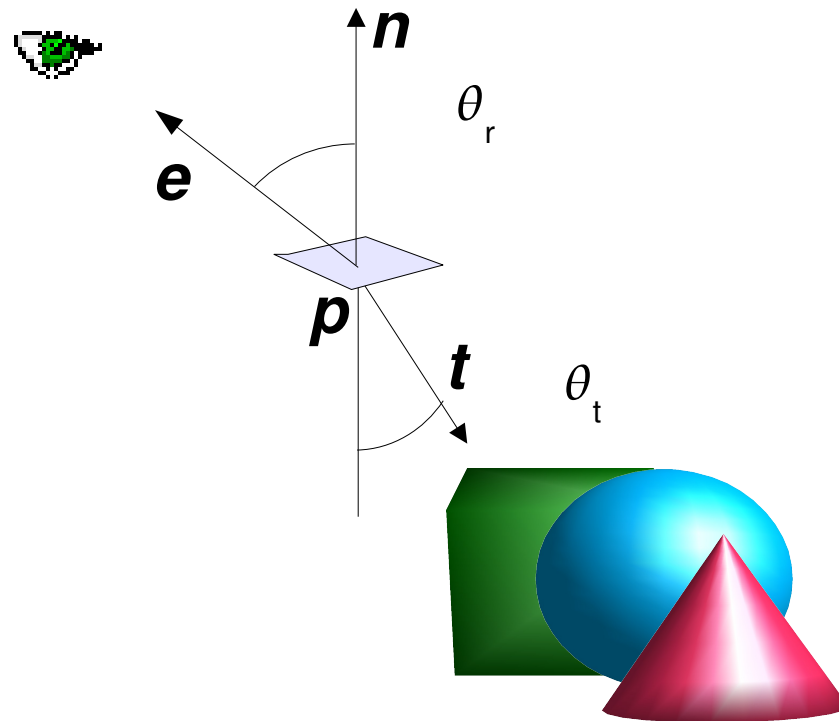
What depth is For

Suppose we have two perfect mirrors placed like this:



The depth parameter is one way to limit the number of reflections. There are more accurate possibilities, but this is the easiest. Note, however, that this is one of the shortcomings of raytracing.

Materials: Dielectrics



If the surface at intersection p is dielectric ((semi-)transparent and light-refracting, like glass or water), we want to show refraction, so we cast a ray starting at p in the transmitted refracted direction t and incorporate the color we get for that ray into the color at p . This may be combined with reflection.

Snell's Law

$$n_t \sin \theta_t = n_i \sin \theta_i$$

Where n_t and n_i are indices of refraction of the two media.

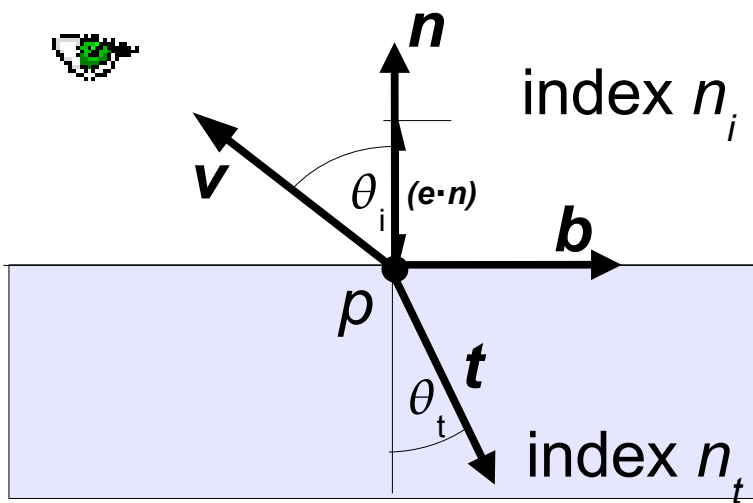
Rewrite as

$$\sin \theta_t = \frac{n_i}{n_t} \sin \theta_i \equiv \eta \sqrt{1 - (\mathbf{v} \cdot \mathbf{n})^2}$$

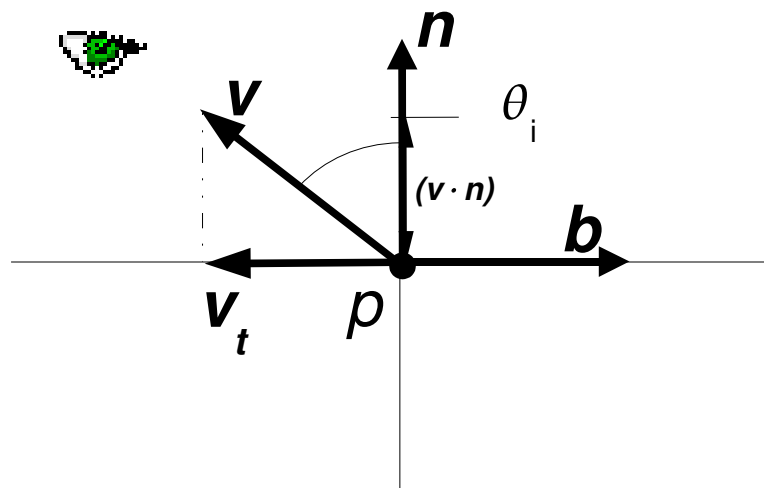
If we had the tangent (unit) vector \mathbf{b} , we could write:

$$\mathbf{t} = \sin \theta_t \mathbf{b} - \cos \theta_t \mathbf{n}$$

$$\Rightarrow \mathbf{t} = \eta \sqrt{1 - (\mathbf{v} \cdot \mathbf{n})^2} \mathbf{b} - \sqrt{1 - \left(\eta \sqrt{1 - (\mathbf{v} \cdot \mathbf{n})^2} \right)^2} \mathbf{n}$$



Computing the Tangent Vector



\mathbf{v} has both a normal (parallel to \mathbf{n}) and a tangential (perpendicular to \mathbf{n}) component. We extract the tangential component

$$\mathbf{v}_t = \mathbf{v} - (\mathbf{n} \cdot \mathbf{v}) \mathbf{n}$$

\mathbf{b} is this vector, negated and normalized:

$$\begin{aligned} \mathbf{b} &= \frac{-\mathbf{v}_t}{|\mathbf{v}_t|} = \frac{(\mathbf{n} \cdot \mathbf{v}) \mathbf{n} - \mathbf{v}}{\sqrt{(\mathbf{v} - (\mathbf{n} \cdot \mathbf{v}) \mathbf{n}) \cdot (\mathbf{v} - (\mathbf{n} \cdot \mathbf{v}) \mathbf{n})}} = \frac{(\mathbf{n} \cdot \mathbf{v}) \mathbf{n} - \mathbf{v}}{\sqrt{(\mathbf{v} \cdot \mathbf{v}) - 2(\mathbf{n} \cdot \mathbf{v})^2 + (\mathbf{n} \cdot \mathbf{v})^2 (\mathbf{n} \cdot \mathbf{n})}} \\ &= \frac{(\mathbf{n} \cdot \mathbf{v}) \mathbf{n} - \mathbf{v}}{\sqrt{1 - (\mathbf{n} \cdot \mathbf{v})^2}} = \frac{\cos \theta_i \mathbf{n} - \mathbf{v}}{\sin \theta_i} \end{aligned}$$

The Transmitted Ray Direction

Substituting

$$\mathbf{b} = \frac{(\mathbf{n} \cdot \mathbf{v}) \mathbf{n} - \mathbf{v}}{\sqrt{1 - (\mathbf{n} \cdot \mathbf{v})^2}}$$

into

$$\mathbf{t} = \eta \sqrt{1 - (\mathbf{n} \cdot \mathbf{v})^2} \mathbf{b} - \sqrt{1 - \left(\eta \sqrt{1 - (\mathbf{n} \cdot \mathbf{v})^2} \right)^2} \mathbf{n}$$

we get (whew!)

$$\mathbf{t} = \eta \left((\mathbf{n} \cdot \mathbf{v}) \mathbf{n} - \mathbf{v} \right) - \sqrt{1 - \left(\eta \sqrt{1 - (\mathbf{n} \cdot \mathbf{v})^2} \right)^2} \mathbf{n}$$

What could go wrong?

Indices of Refraction

<u>Medium</u>	<u>Index of Refraction</u>
Methylene Iodide	1.74
Glass, dense flint	1.66
Carbon bisulfide	1.63
Sodium chloride	1.53
Glass, crown	1.52
Fused Quartz	1.46
Ethyl alcohol	1.36
Water	1.33
Air (1 atm, 20° C)	1.0003
Vacuum	1.00

The Fresnel Term

When light reflects off a dielectric, its reflectivity has an angular dependence.

$$R(\mathbf{v}, \mathbf{s}, \eta) = \frac{1}{2} \frac{(g - c)^2}{(g + c)^2} \left\{ 1 + \frac{(c(g + c) - 1)^2}{(c(g - c) - 1)^2} \right\}$$

where

$$\begin{aligned} c &\equiv \mathbf{v} \cdot \mathbf{h} \\ h &\equiv \widehat{\mathbf{v} + \mathbf{s}} \\ g^2 &\equiv \eta^2 + c^2 - 1 \end{aligned}$$

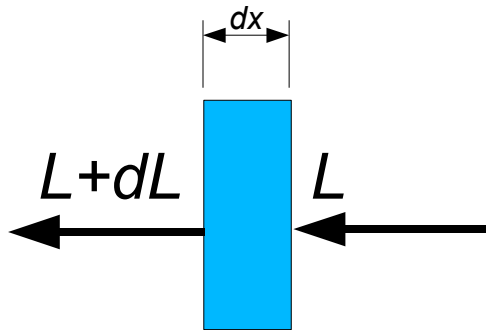
This is time-consuming. A reasonable approximation (Shirley) is

$$R(\mathbf{v}, \mathbf{n}, \eta) = R_0 + (1 - R_0)(1 - (\mathbf{v} \cdot \mathbf{n}))^5 \quad R_0 \equiv \left(\frac{\eta^{-1} - 1}{\eta^{-1} + 1} \right)^2$$

Dielectric.rdnclndirect()

```
class Dielectric(Material):
    ...
    def rdnclndirect(dlct, ix, ryIn, scn):
        uV = -ryIn.d.unit()
        ryRefl = Ray(ix.p, uV.reflect(ix.uN), ryIn.nRefr,
                     ryIn.depth+1)
        ryRefr = Ray(ix.p, uV.refract(ix.uN, dlct.nRefr),
                     dlct.nRefr, ryIn.depth+1))
        f = dlct.fresnel(uV, ix.uN, ryIn.nRefr)
        if f == 1: # total internal reflection
            rdnc = scn.trace(ryRefl)
        else:
            rdnc = (f * scn.trace(ryRefr)
                    + (1 - f) * scn.trace(ryRefl))
        return dlct.kAbs * rdnc
```

Attenuating Media



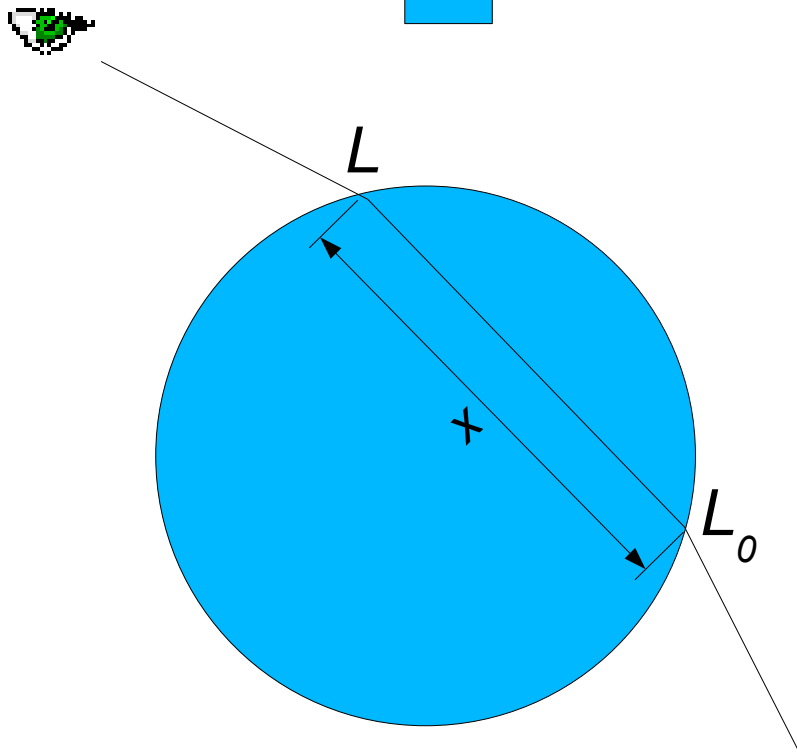
Beer's Law: Light passing through an absorbing medium is diminished by a fractional amount proportional to the thickness of the medium.

$$dL = -C L dx$$

$$\Rightarrow \frac{dL}{L} = -C dx \Rightarrow L = L_0 e^{-Cx}$$

So e^{-Cx} is the *attenuation factor*.

This is the same idea as fog in OpenGL. How would you incorporate this into the schema?

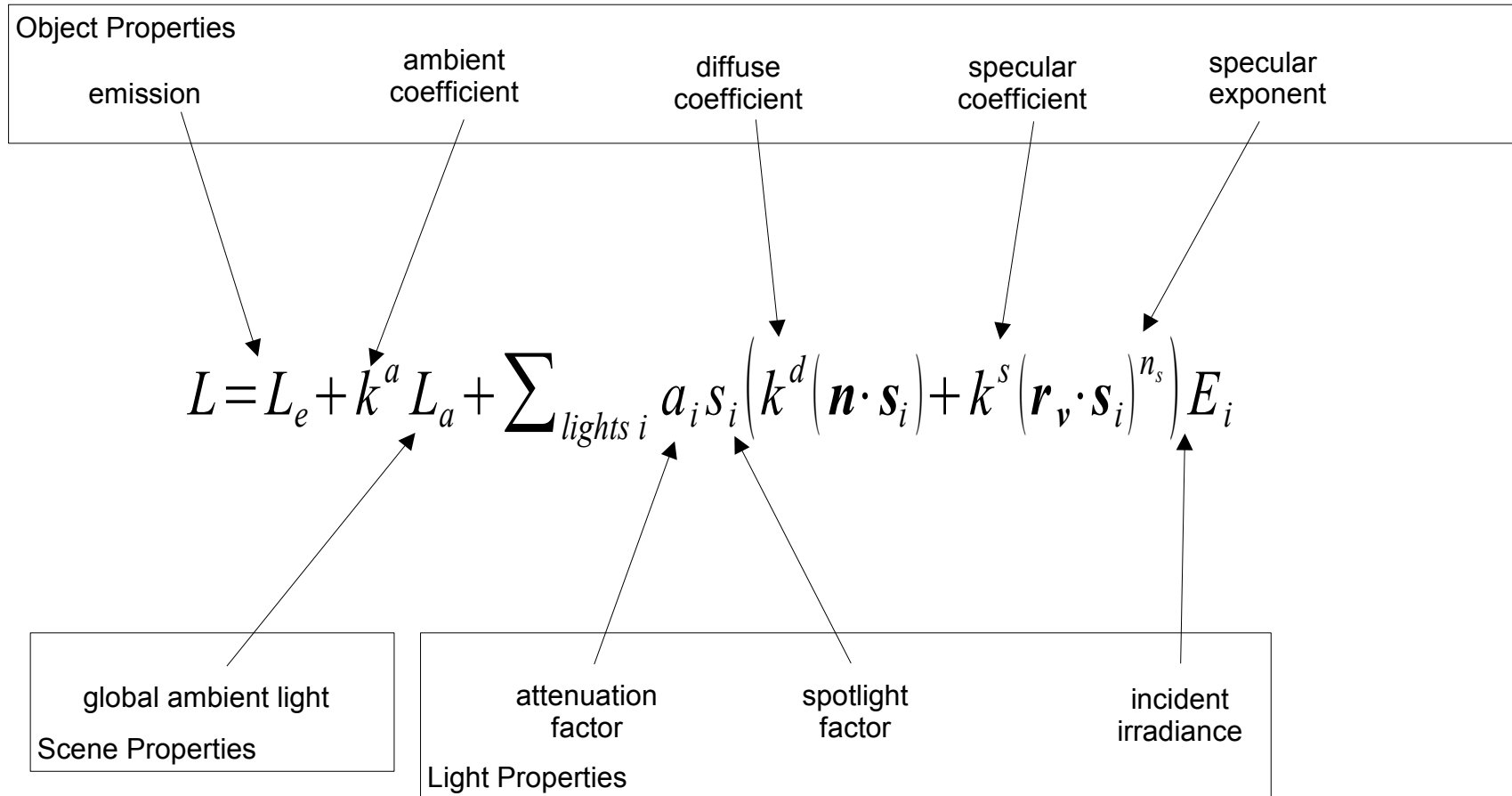


Materials: Polished Surfaces

This is an alternative shader developed by Shirley:

$$L = R \left(1 - \left(1 - (\mathbf{n} \cdot \mathbf{v}) \right)^5 \right) \left(1 - \left(1 - (\mathbf{n} \cdot \mathbf{s}) \right)^5 \right) E$$

A Conventional Ray Tracing Shader



Negative dot products are ignored.

DiffusePhongAmbient.rdncDirect()

This is the traditional reflectance model used in OpenGL. Note that we must divide the irradiance by the incident cosine to recover the original (non-physically plausible) formula.

```
class DiffusePhongAmbient(Material):
    ...
    def rdncDirect(dpa, ry, ix, lum):
        uV = -ry.d.unit()
        uN = ix.uN()
        uS = lum.uS(ix.p) # points towards source
        uRv = uV.reflect(uN)
        # officially, "irradiance" includes S.N factor
        irr = lum.irr(p, uN)
        # but the specular term (non-physically) ignores it
        irrSpec = irr / uS.dot(uN) # so we remove it
        return dpa.kd * irr \
            + dpa.ks * (uS.dot(uRv)**dpa.expo) * irrSpec \
            + dpa.ka * lum.irrAmb
```


So What's Wrong with Ray Tracing (so far)?

- It takes a long time.
- We're always using rays to sample.
- We don't handle interreflecting diffuse surfaces.
- Our light sources are limited to point or unidirectional.