Ray Tracing

Brian Curless
CSEP 557
Spring 2019

Reading

Required:

- Marschner and Shirley, Ch. 4, Section 13.1-13.2 (online handout)
- Triangle intersection (online handout)

Further reading:

- Shirley errata on syllabus page, needed if you work from his book instead of the handout, which has already been corrected.
- T. Whitted. An improved illumination model for shaded display. Communications of the ACM 23(6), 343-349, 1980.
- A. Glassner. An Introduction to Ray Tracing. Academic Press, 1989.
- K. Turkowski, "Properties of Surface Normal Transformations," Graphics Gems, 1990, pp. 539-547.

Geometric optics

Modern theories of light treat it as both a wave and a particle.

We will take a combined and somewhat simpler view of light – the view of **geometric optics**.

Here are the rules of geometric optics:

- ◆ Light is a flow of photons with wavelengths. We'll call these flows "light rays."
- ◆ Light rays travel in straight lines in free space.
- ◆ Light rays do not interfere with each other as they cross.
- ◆ Light rays obey the laws of reflection and refraction.
- ◆ Light rays travel from the light sources to the eye, but the physics is invariant under path reversal (reciprocity).
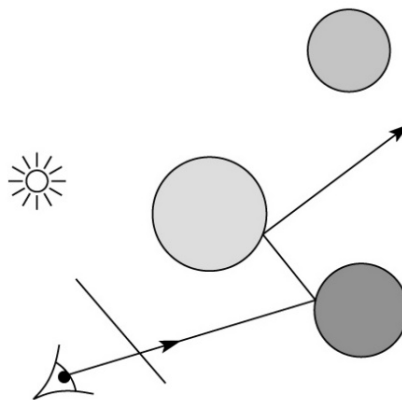
Eye vs. light ray tracing

Where does light begin?

At the light: light ray tracing (a.k.a., forward ray tracing or photon tracing)
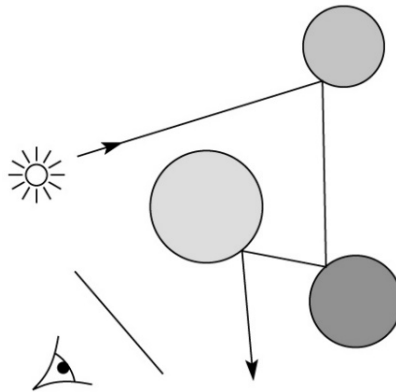
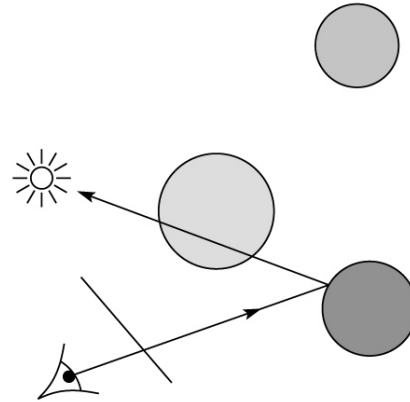At the eye: eye ray tracing (a.k.a., backward ray tracing)

We will generally follow rays from the eye into the scene.
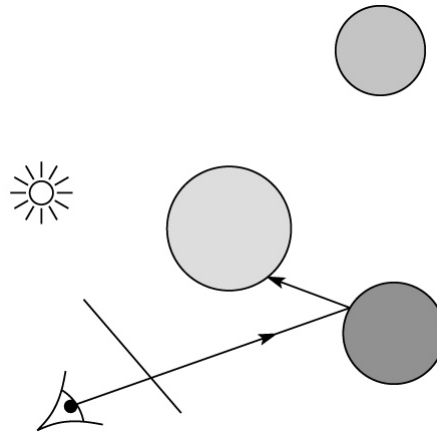
Precursors to ray tracing

Local illumination

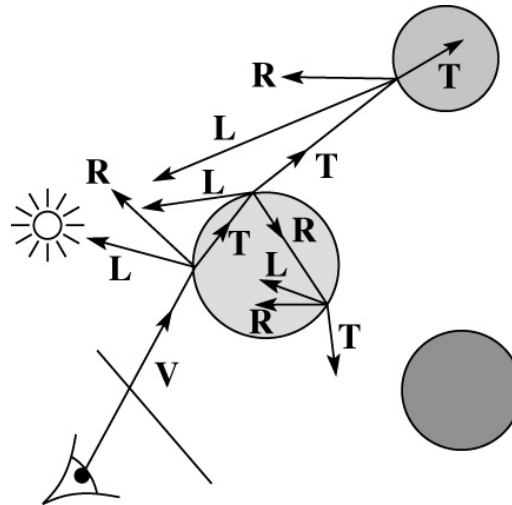♦ Cast one eye ray, then shade according to light

Appel (1968)

♦ Cast one eye ray + one ray to light

Whitted ray-tracing algorithm

In 1980, Turner Whitted introduced ray tracing to the graphics community.

- ◆ Combines eye ray tracing + rays to light
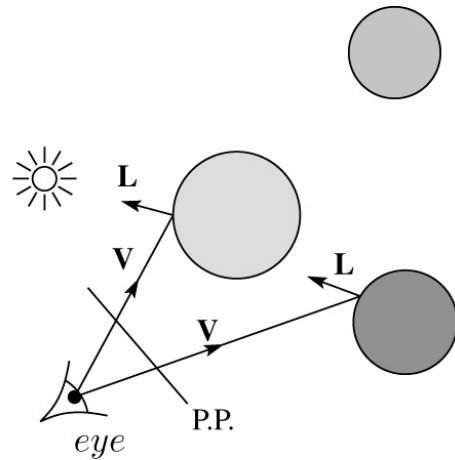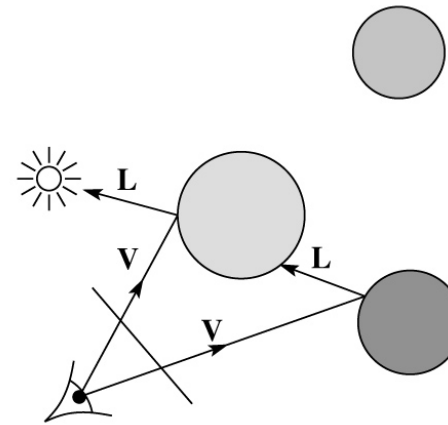- ◆ Recursively traces rays

Algorithm:

1. For each pixel, trace a **primary ray** in direction **V** to the first visible surface.

2. For each intersection, trace **secondary rays**:

- ◆ **Shadow rays** in directions L_i to light sources
- ◆ **Reflected ray** in direction **R**.
- ◆ **Refracted ray** or **transmitted ray** in direction **T**.
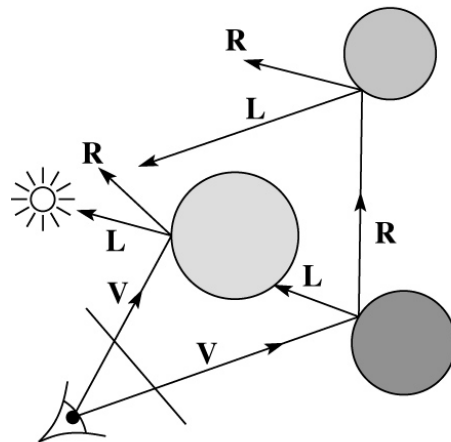
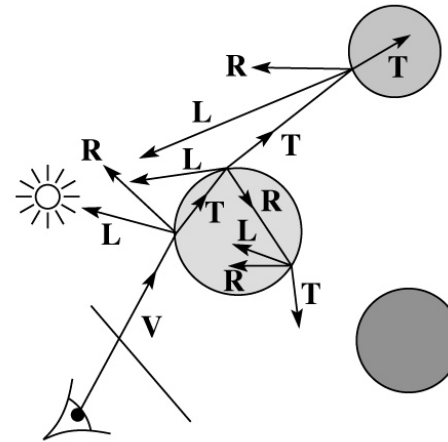# Whitted algorithm (cont'd)

Let's look at this in stages:

Primary rays

Shadow rays

Reflection rays

Refracted rays

Ray casting and local illumination

Now let's actually build the ray tracer in stages.
We'll start with ray casting and local illumination:
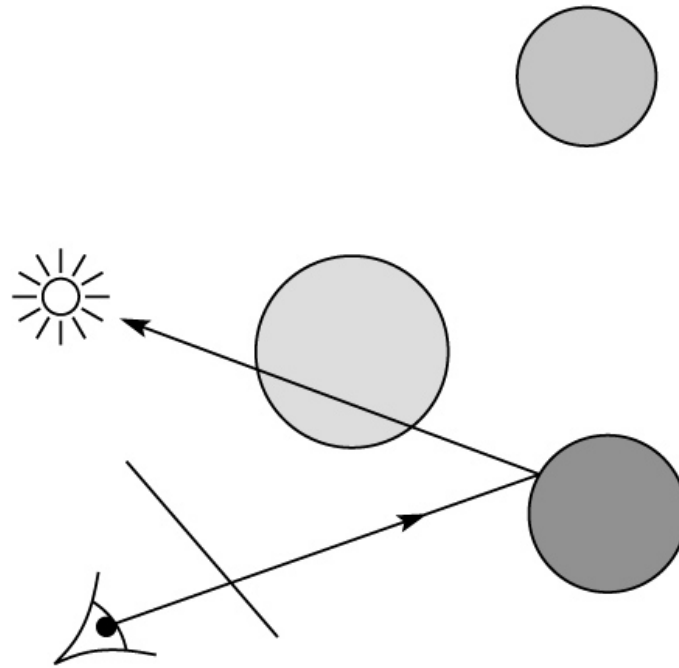
Direct illumination
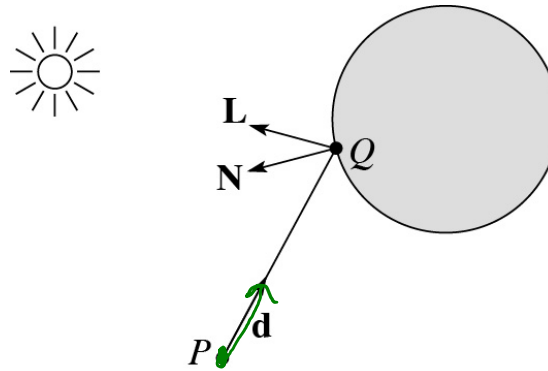
A ray is defined by an origin P and a unit direction $\mathbf{d}$ and is parameterized by $t > 0$:

$$\mathbf{r}(t) = P + t\,\mathbf{d}$$

Let $I(P, \mathbf{d})$ be the intensity seen along a ray. Then:

$$I(P, \mathbf{d}) = I_{\text{direct}}$$

where

- I_{direct} is computed from the Blinn-Phong model

Ray-tracing pseudocode

We build a ray traced image by casting rays through each of the pixels.

function *traceImage* (scene):

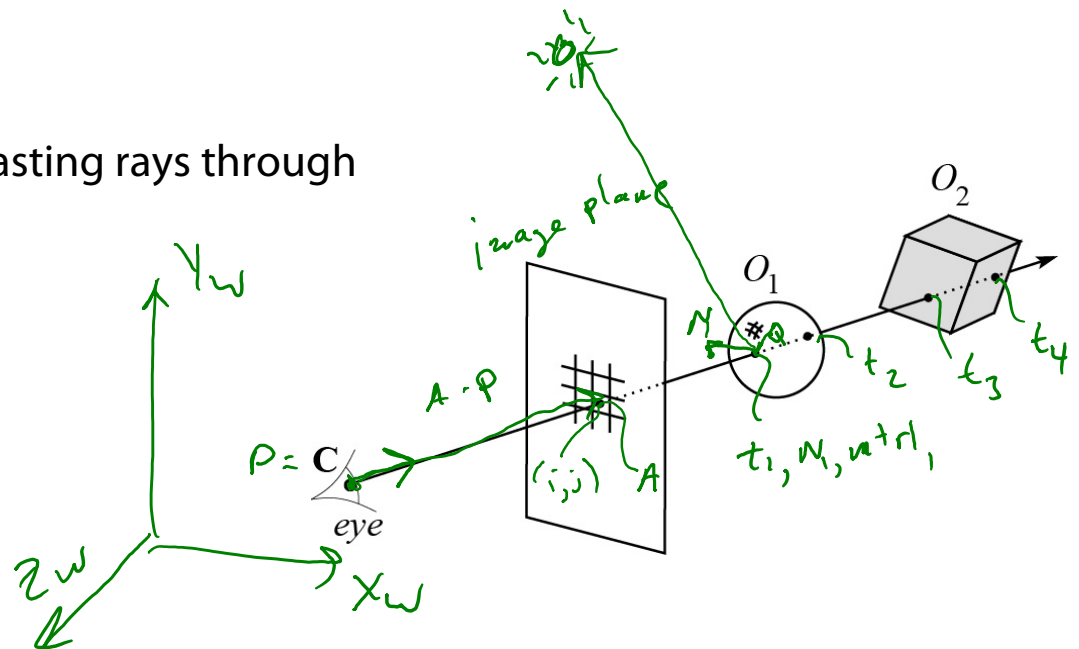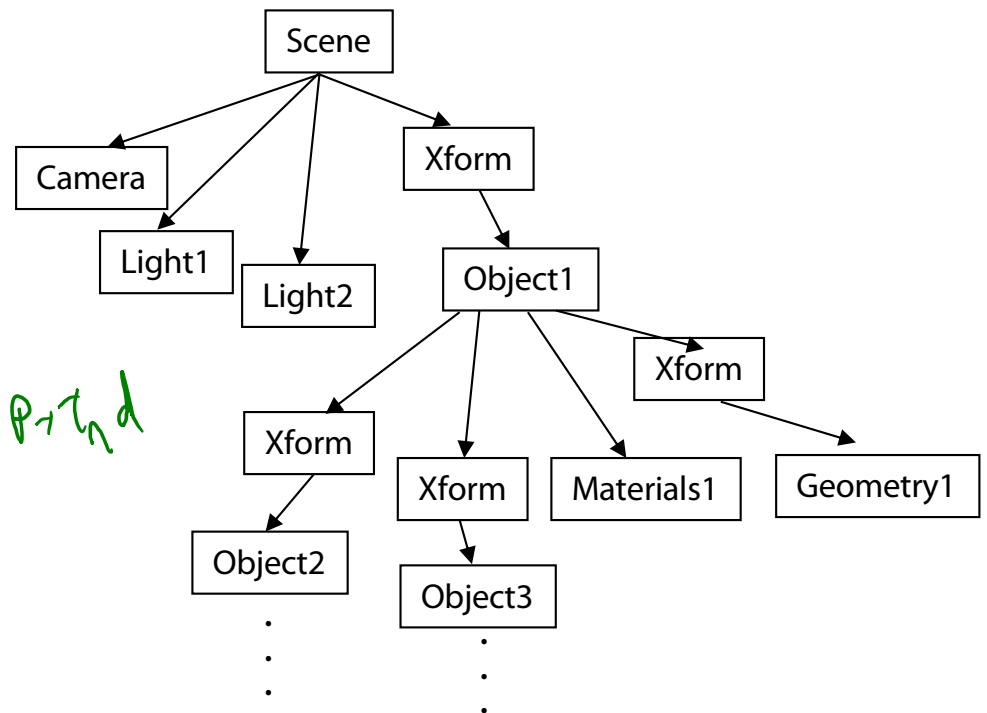   **for each** pixel (i, j) in image

        $A = pixelToWorld\ (i, j)$

        $P = \mathbf{C}$

        $\mathbf{d} = (A - P)/\| A - P \|$

        $I(\mathrm{i,j}) = traceRay\ (scene, P, \mathbf{d})$

    end for

end function

function *traceRay* (scene, P, $\mathbf{d}$):

    $(t_\cap, \mathbf{N}, \text{mtrl}) \leftarrow$ scene.*intersect* $(P, \mathbf{d})$

    $Q \leftarrow$ ray $(P, \mathbf{d})$ evaluated at $t_\cap$

    $I = shade\ (\ N, mtrl, d, scene, Q\ )$

    return I

end function

Handwritten annotations:

Y_W, Z_W, X_W

$P = \mathbf{C}$, eye

image plane

$A \cdot P$

(i,j), A

O_1, N, Q, $t_1, N, mtrl_1$, t_2, t_3, t_4

O_2

$Q = P + t_\cap d$

Scene

Camera

Light1

Light2

Xform

Object1

Xform

Object2

Xform

Object3

Xform

Xform

Materials1

Geometry1

Shading pseudocode

Next, we need to calculate the color returned by the *shade* function.

function *shade* (mtrl, scene, Q, $\mathbf{N}$, $\mathbf{d}$):

    $I \leftarrow$ mtrl.k_e

    for each light source Light **do**:

        atten = Light -> *distanceAttenuation* (Q)

        $\mathbf{L}$ = Light -> getDirection (Q)

        $I \leftarrow I +$ ambient + atten*(diffuse + specular)

    end for
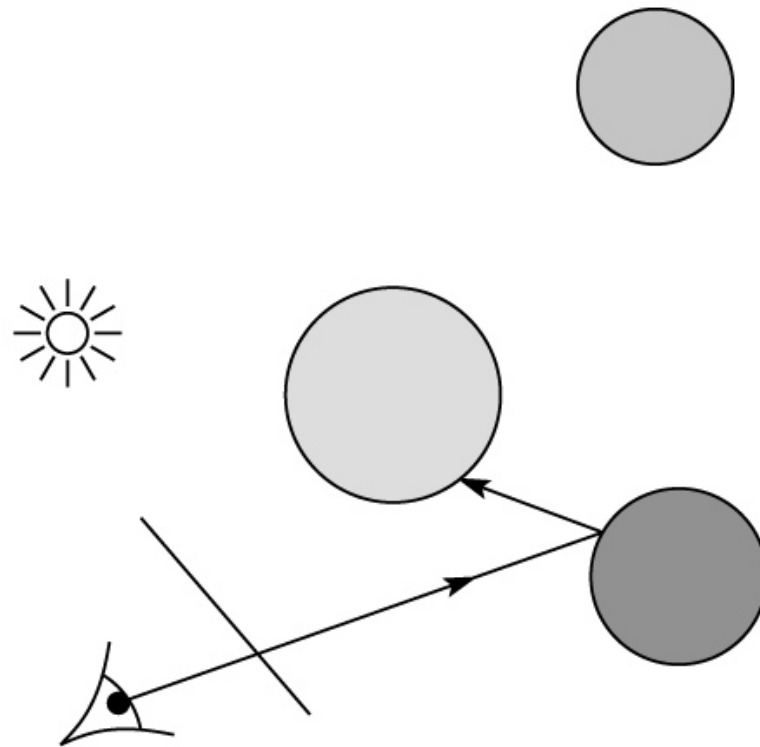
    return I

end function

$$r = \| E - Q \|$$

$$L = \frac{E - Q}{r}$$
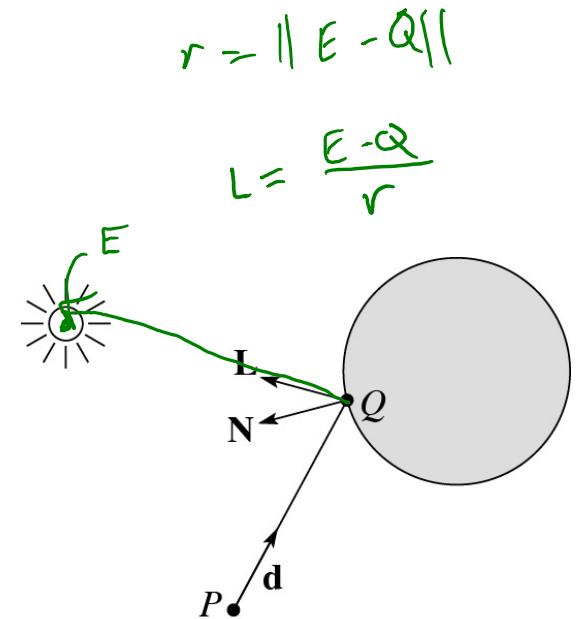
Ray casting with shadows

Now we'll add shadows by casting shadow rays:

Shading with shadows

To include shadows, we need to modify the shade function:

function *shade* (mtrl, scene, Q, **N**, **d**):

    I ← mtrl.k_e

    for each light source Light **do**:

        atten = Light -> *distanceAttenuation*(Q) *

            Light -> *shadowAttenuation*(Q , Scene)

        L = Light -> getDirection (Q)

        $I \leftarrow I$ + ambient + atten*(diffuse + specular)

    end for

    return I
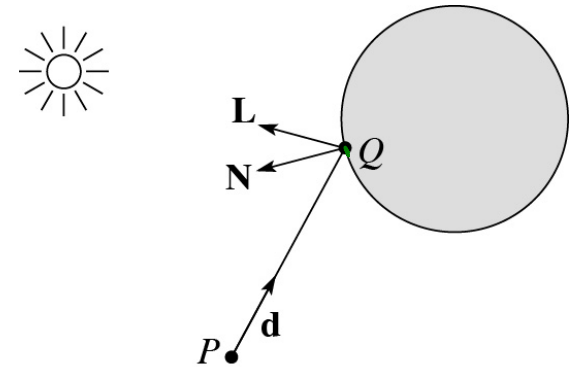
end function

Shadow attenuation

Computing a shadow can be as simple as checking to see if a ray makes it to the light source.

For a point light source:

function *PointLight* **::***shadowAttenuation* (scene, Q)

    $\mathbf{L}$ = getDirection(Q)

    ($t_\cap$, $\mathbf{N}$, mtrl) $\leftarrow$ scene.*intersect* (Q, $\mathbf{L}$)

    Compute t_{light}

    if ($t_\cap < t_{\text{light}}$) **then**:

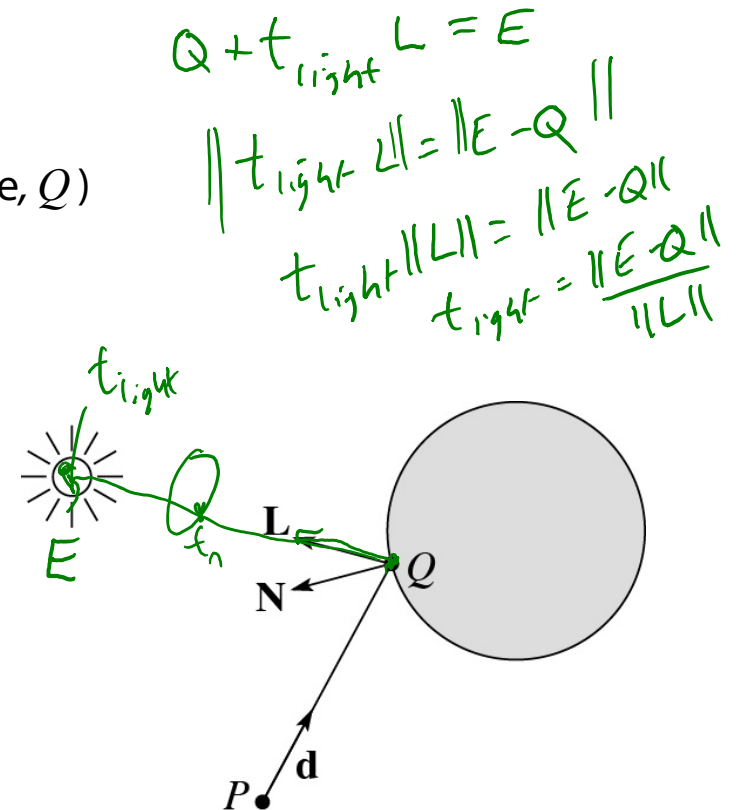   atten = (0, 0, 0)

    else

        atten = (1, 1, 1)

    end if

    return atten

end function

Note: we will later handle color-filtered shadowing, so this function needs to return a *color* value.

For a directional light, $t_{\text{light}} = \infty$.

$$Q + t_{\text{light}}\,\mathbf{L} = E$$

$$\| t_{\text{light}}\,\mathbf{L} \| = \| E - Q \|$$

$$t_{\text{light}} \| \mathbf{L} \| = \| E - Q \|$$

$$t_{\text{light}} = \frac{\| E - Q \|}{\| \mathbf{L} \|}$$

Shading in "Trace"

The Trace project uses a version of the Blinn-Phong shading equation we derived in class, with two modifications:

- ◆ Distance attenuation is clamped to be at most 1:

$$A_j^{dist} = \min\left\{1, \frac{1}{a_j r_j^{\ 2} + b_j r_j + c_j}\right\}$$

- ◆ Shadow attenuation A^{shadow} is included and is RGB-valued.
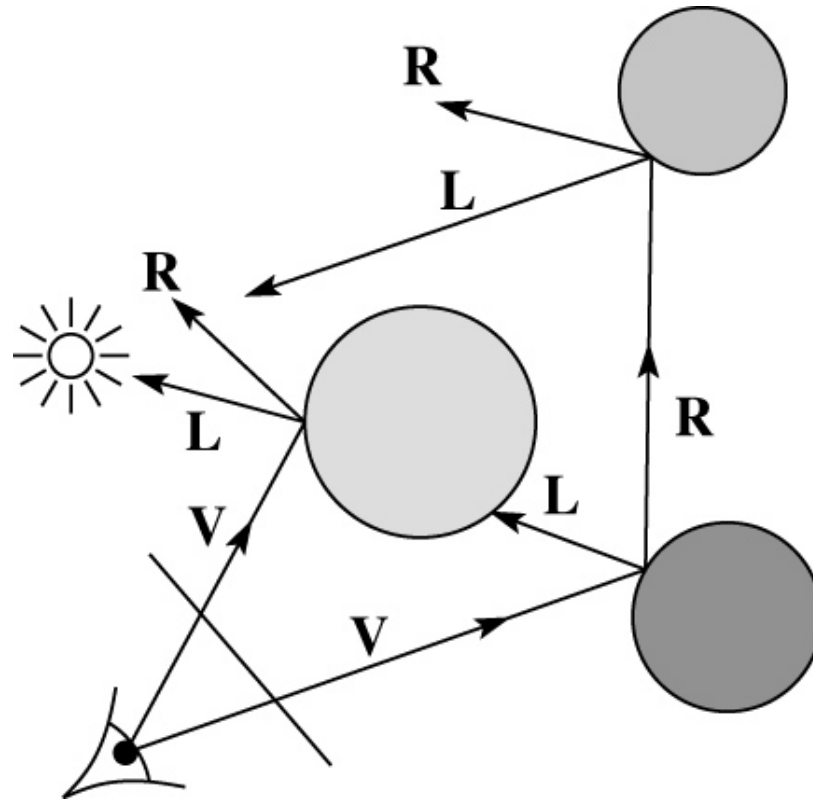
Here's what the shading equation should look like:

$$I = k_e + \sum_j k_d I_{La,j} + A_j^{shadow} A_j^{dist} I_{L,j} B_j \left[k_d \left(\mathbf{N} \cdot \mathbf{L}_j \right) + k_s \left(\mathbf{N} \cdot \mathbf{H}_j \right)_+^{n_s} \right]$$

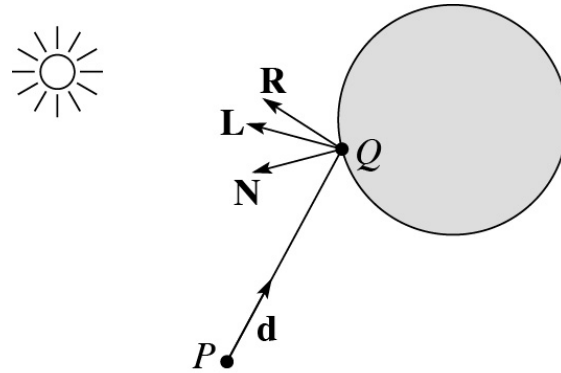**This is the shading equation to use in the Trace project!**

Recursive ray tracing with reflection

Now we'll add reflection:

Shading with reflection

Let $I(P, \mathbf{d})$ be the intensity seen along a ray. Then:

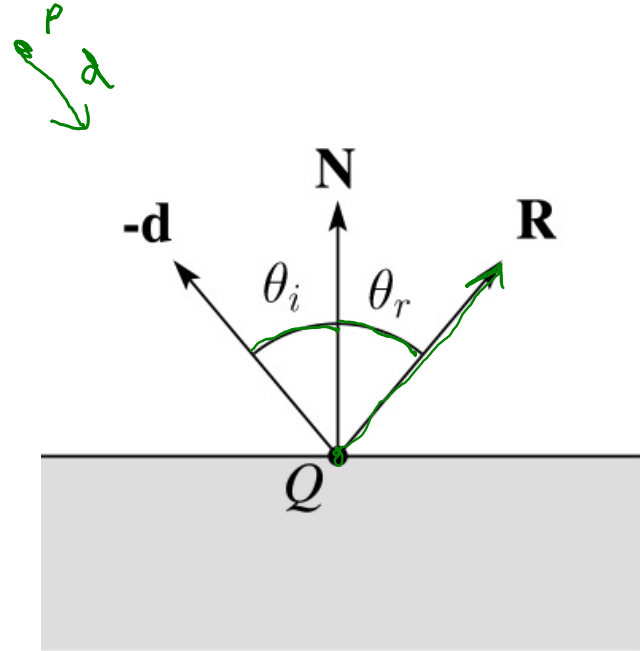$$I\,(P, \mathbf{d}) = I_{\text{direct}} + I_{\text{reflected}}$$

where

- I_{direct} is computed from the Blinn-Phong model, plus shadow attenuation
- $I_{\text{reflected}} = k_s\, I(Q, \mathbf{R})$

Remember that is a color value.

k_s
λ

(Sometimes another variable, k_r, is used instead of k_s to allow for separate control of specular light reflection vs specular interreflection. We will not do this.)
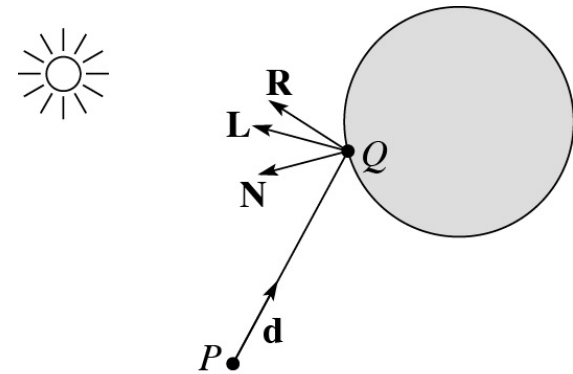
Reflection

Law of reflection:

$$\theta_i = \theta_r$$

R is co-planar with **d** and **N**.

Ray-tracing pseudocode, revisited

function *traceRay* (scene, P, **d**):

    $(t_\cap, \mathbf{N}, \text{mtrl}) \leftarrow$ scene.*intersect* $(P, \mathbf{d})$

    $Q \leftarrow$ ray $(P, \mathbf{d})$ evaluated at $t_\cap$

    $I = shade$ (scene, mtrl, Q, $\mathbf{N}$, $\mathbf{d}$)

    $\mathbf{R} = reflectDirection$ ($\mathbf{N}$, $\mathbf{d}$)

    $I \leftarrow I + \text{mtrl}.k_s * traceRay(\text{scene}, Q, \mathbf{R})$

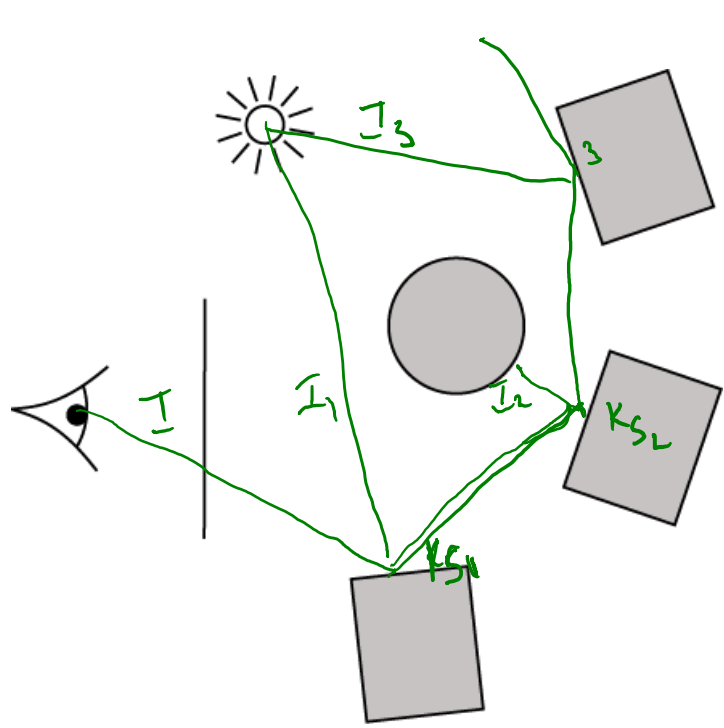   **return** I

end function

Terminating recursion

Q: How do you bottom out of recursive ray tracing?

Possibles:

do this ⟶ # of bounces = max-depth ⟹ end recursion
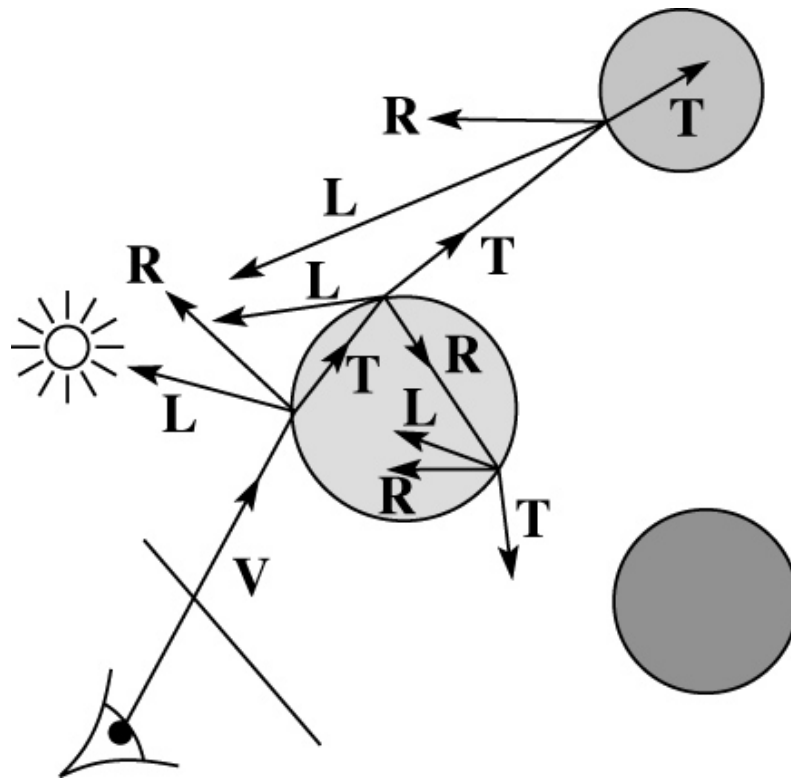
$$I = I_1 + K_{S_1}(I_2 + K_{S_2}(I_3 + K_{S_3}(\ldots$$

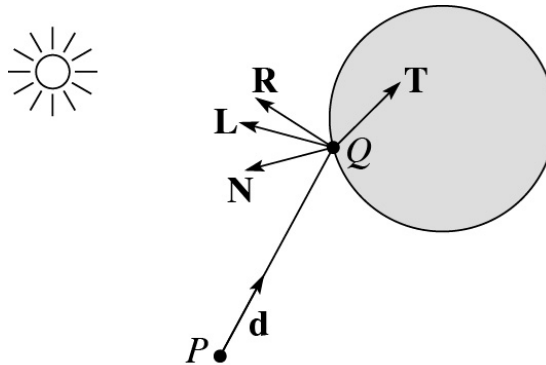$$= I_1 + K_{S_1}I_2 + K_{S_1}K_{S_2}I_3 + K_{S_1}K_{S_2}K_{S_3}I_4$$

$\prod K_{S_i} <$ thresh

adaptive
or
early ray termination

Whitted ray tracing

Finally, we'll add refraction, giving us the Whitted ray tracing model:

Shading with reflection and refraction

Let $I(P, \mathbf{d})$ be the intensity seen along a ray. Then:

$$I(P, \mathbf{d}) = I_{\text{direct}} + I_{\text{reflected}} + I_{\text{transmitted}}$$

where

- I_{direct} is computed from the Blinn-Phong model, plus shadow attenuation
- $I_{\text{reflected}} = k_s\, I(Q, \mathbf{R})$
- $I_{\text{transmitted}} = k_t\, I(Q, \mathbf{T})$

Typically, we set $k_t = 1 - k_s$ (or (0,0,0), if opaque, where k_t is a color value).

[Generally, for ideal specular surfaces, k_s and k_t are determined by "Fresnel reflection," which depends on angle of incidence and changes the polarization of the light. This is discussed in Marschner's textbook and can be implemented for extra credit.]
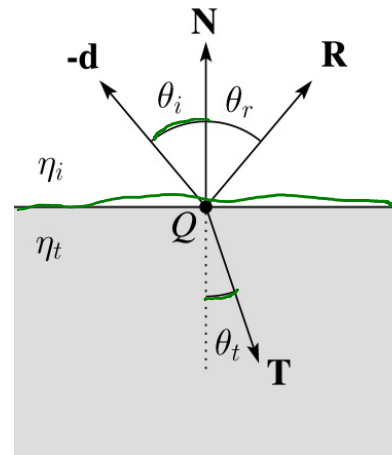
Refraction

Snell's law of refraction:

$$\eta_i \sin\theta_i = \eta_t \sin\theta_t$$

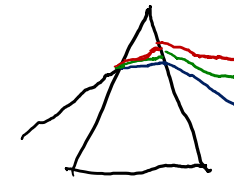where η_i, η_t are **indices of refraction**.

In all cases, **R** and **T** are co-planar with **d** and **N**.

The index of refraction is material dependent.

It can also vary with wavelength, an effect called **dispersion** that explains the colorful light rainbows from prisms. (We will generally assume no dispersion.)

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

Medium	Index of refraction
Vaccum	1
Air	1.0003
Water	1.33
Fused quartz	1.46
Glass, crown	1.52
Glass, dense flint	1.66
Diamond	2.42
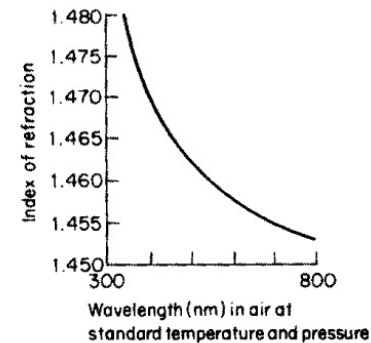
Index of refraction variation for fused quartz

23

Total Internal Reflection

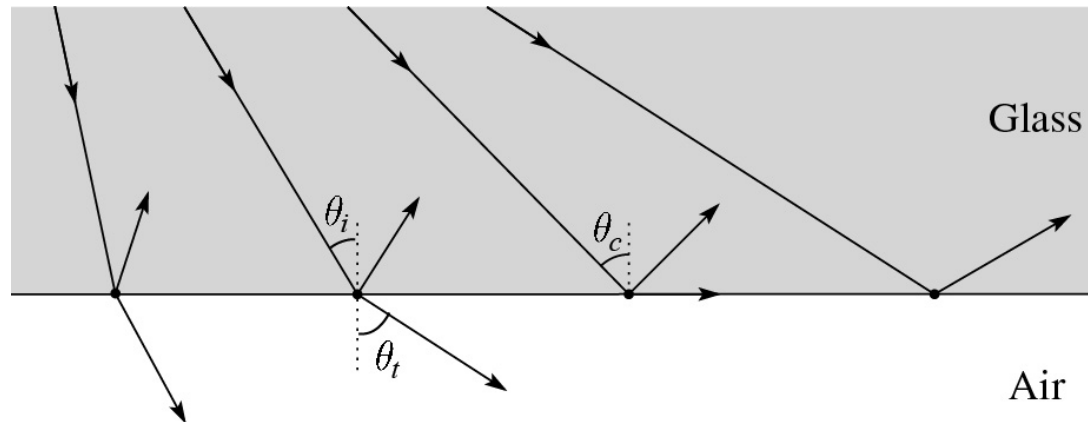The equation for the angle of refraction can be computed from Snell's law:

$$\eta_i \sin\theta_i = \eta_t \sin\theta_t$$

$$\sin\theta_t = \frac{\eta_i \sin\theta_i}{\eta_t}$$

$$\theta_t = \sin^{-1}\left[\frac{\eta_i \sin\theta_i}{\eta_t}\right]$$

What "bad thing" can happen when $\eta_i > \eta_t$?

When θ_t is exactly 90°, we say that θ_i has achieved the "critical angle" θ_c.

For $\theta_i > \theta_c$, *no rays are transmitted*, and only reflection occurs, a phenomenon known as "total internal reflection" or TIR.

Marschner's notation

Marschner uses different symbols. Here is the translation between them:

$$\mathbf{r} = \mathbf{R}$$
$$\mathbf{t} = \mathbf{T}$$
$$\phi = \theta_t$$
$$\theta = \theta_r = \theta_i$$
$$n = \eta_i$$
$$n_t = \eta_t$$

Ray-tracing pseudocode, revisited

function *traceRay* (scene, P, **d**):

$(t_\cap, \mathbf{N}, \text{mtrl}) \leftarrow$ scene.*intersect* $(P, \mathbf{d})$

$Q \leftarrow$ ray $(P, \mathbf{d})$ evaluated at $t_\cap$

$I = shade$ (scene, mtrl, Q, $\mathbf{N}$, $\mathbf{d}$)

$\mathbf{R} = reflectDirection$ $(\mathbf{N}, \mathbf{d})$

$I \leftarrow I + \text{mtrl}.k_s * traceRay$ (scene, Q, $\mathbf{R}$)

if ray is entering object **then**

$\eta_i = $ index_of_air (=1.0003)

$\eta_t = $ mtrl.index

else

$\eta_i = $ mtrl.index

$\eta_t = $ index_of_air (=1.0003)

if (*notTIR*(d, N, M_i, M_t)) **then**

$\mathbf{T} = refractDirection$ ( )

$I \leftarrow I + \text{mtrl}.k_t * traceRay$ (scene, Q, $\mathbf{T}$)

end if

return I

end function

$N \cdot (-d) > 0$

$\downarrow$
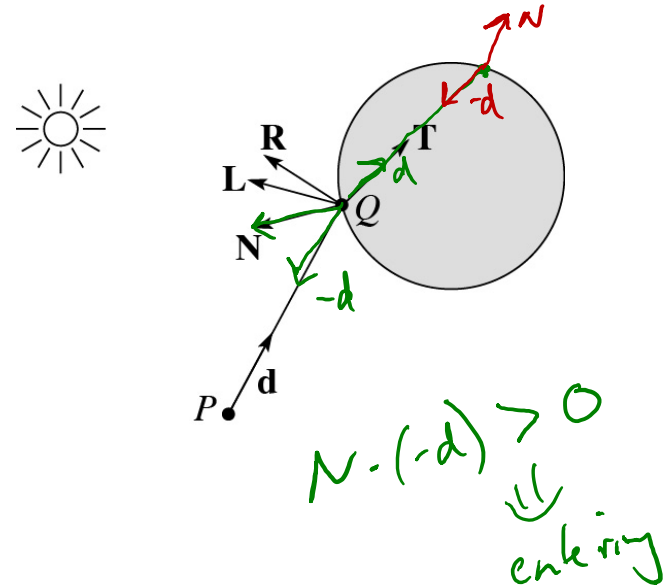
entering

Q: How do we decide if a ray is entering the object?

26

Terminating recursion, incl. refraction

do this

$\downarrow$

max depth cut-off

Q: *Now* how do you bottom out of recursive ray tracing?

$\prod_i K_{s_i} \text{ or } t_i < thresh$

$\downarrow$

end
recursion

Extra
credit

$Ks4$

$Kk3$

Kk

Ks
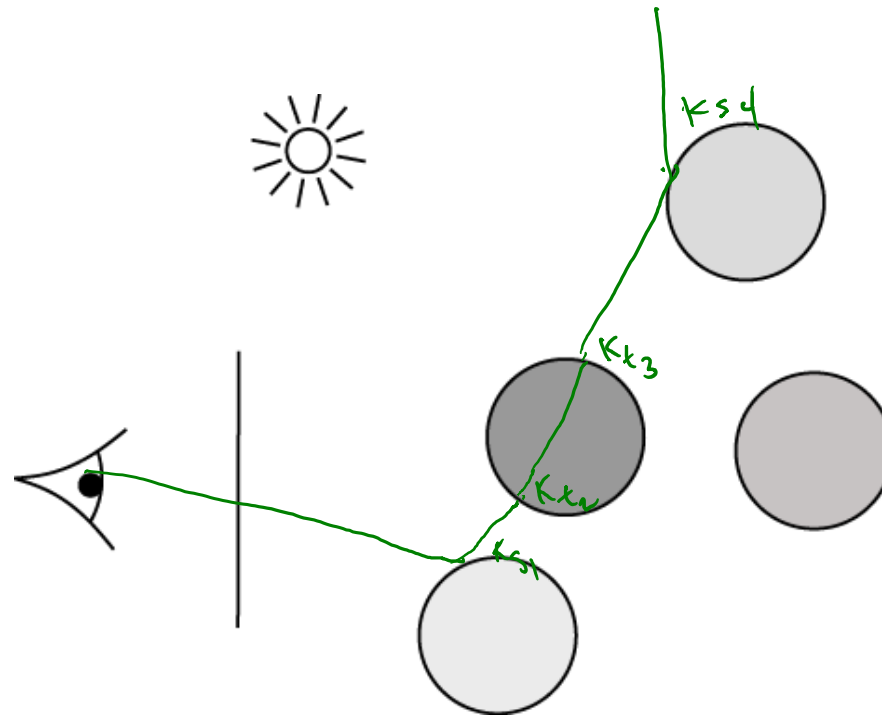
# Shadow attenuation (cont'd)

Q: What if there are transparent objects along a path to the light source?

Do this

Πk_{ti}
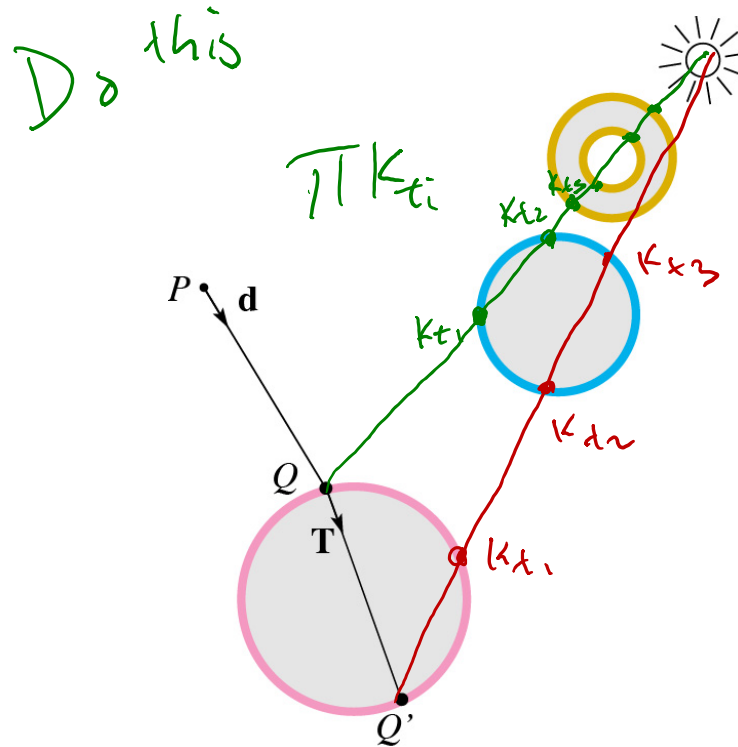
We'll take the view that the color is really only at the surface, like a glass object with a colored transparency coating on it. In this case, we multiply in the transparency constant, k_t, every time an object is entered or exited, possibly more than once for the same object.
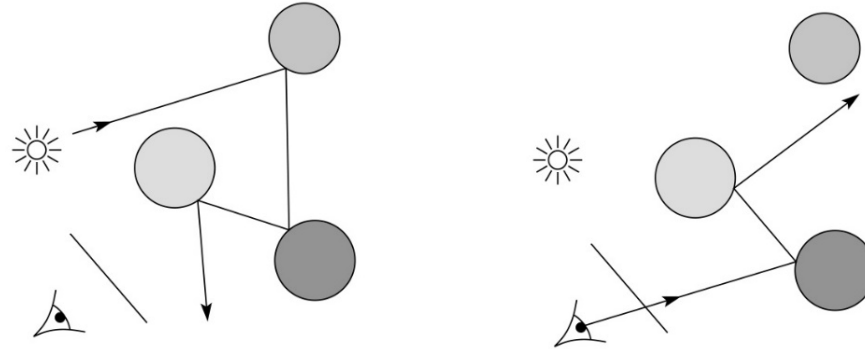
Shadow attenuation (cont'd)

Another model would be to treat the glass as solidly
colored in the interior. Marschner's textbook
describes a the resulting volumetric attenuation based
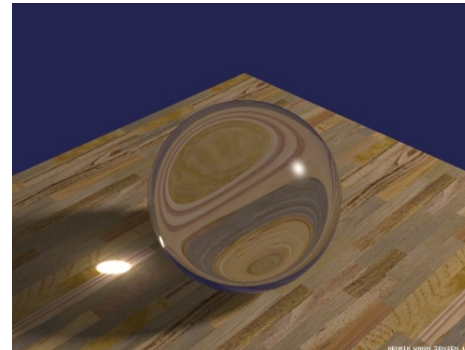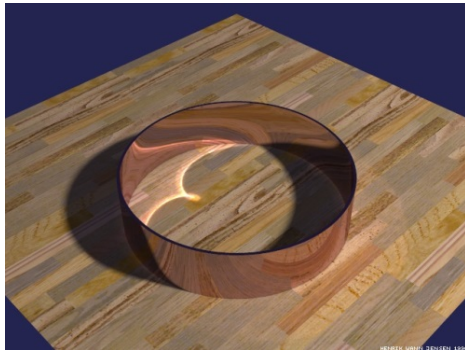on Beer's Law, which you can implement for extra
credit.
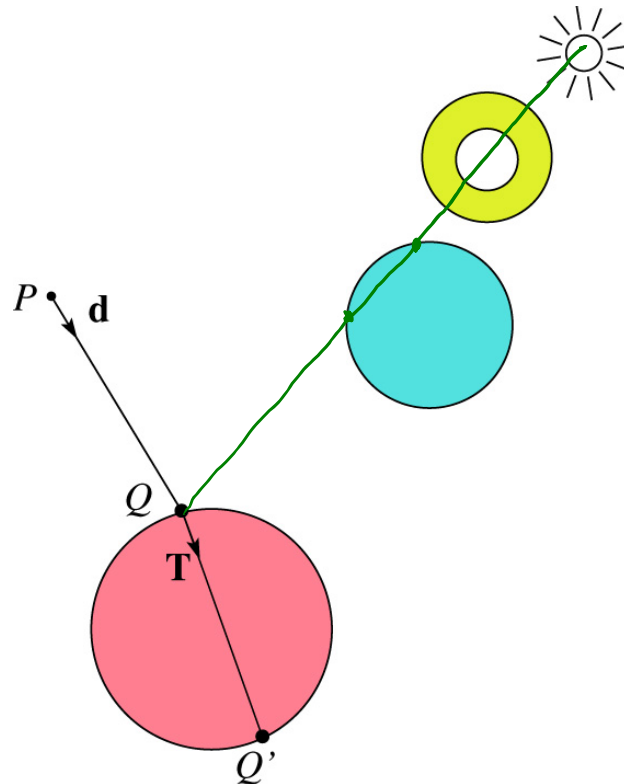
Photon mapping

Combine light ray tracing (photon tracing) and eye ray tracing:

…to get **photon mapping**.

Renderings by Henrik Wann Jensen:

http://graphics.ucsd.edu/~henrik/
images/caustics.html

Normals and shading, reflection, and refraction when inside

When a ray is inside an object and intersects the object's surface on the way out, the normal will be pointing **away** from the ray (i.e., the normal always points to the outside by default).

You must **negate** the normal before doing any of the **shading, reflection, and refraction** that follows.

Finally, when shading a point inside of an object, apply k_t to the ambient component, since that "ambient light" had to pass through the object to get there in the first place.
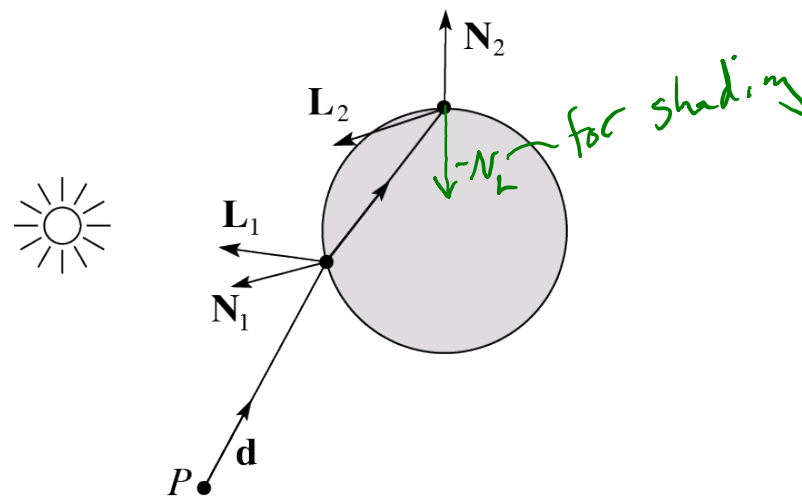
Intersecting rays with spheres

Now we've done everything except figure out what that "scene.*intersect* $(P, \mathbf{d})$" function does.

Mostly, it calls each object to find out the t-value at which the ray intersects the object. Let's start with intersecting spheres…
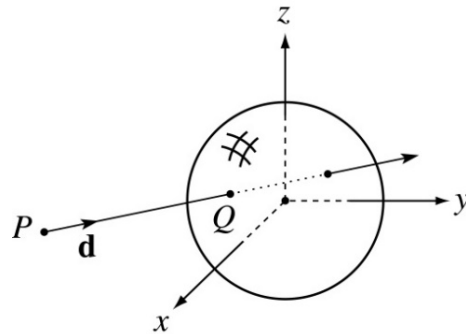
Given:

- The coordinates of a point along a ray passing through P in the direction $\mathbf{d}$ are:

$$x = P_x + td_x$$

$$y = P_y + td_y$$

$$z = P_z + td_z$$

- A sphere S of radius r centered at the origin defined by the equation:

$$x^2 + y^2 + z^2 = r^2$$

Find: The t at which the ray intersects S.

32

Intersecting rays with spheres

Solution by substitution:

$$x^2 + y^2 + z^2 - r^2 = 0$$

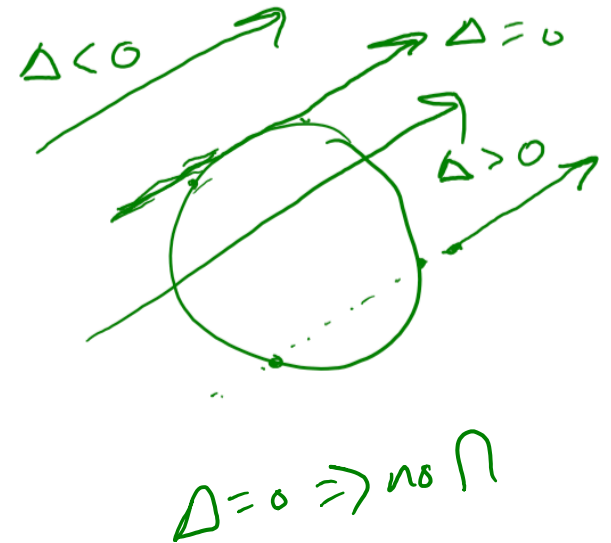$$(P_x + td_x)^2 + (P_y + td_y)^2 + (P_z + td_z)^2 - r^2 = 0$$

$$at^2 + bt + c = 0$$

where

$$a = d_x^2 + d_y^2 + d_z^2$$

$$b = 2(P_x d_x + P_y d_y + P_z d_z)$$

$$c = P_x^2 + P_y^2 + P_z^2 - r^2$$

$\Delta < 0$ $\Delta = 0$

$\Delta > 0$

$\Delta = 0 \Rightarrow$ no $\cap$

Q: What are the solutions of the quadratic equation in t and what do they mean?

$$t = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$\Delta = b^2 - 4ac$$

discriminant

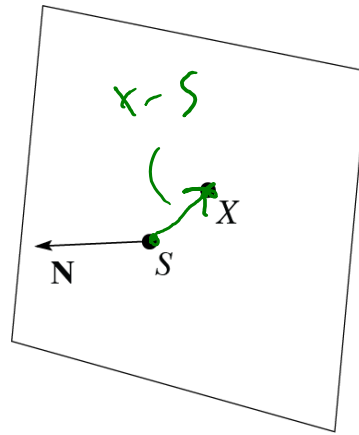$$\Delta < 0 \Rightarrow 2 \text{ complex conjugate roots}$$

Q: What is the normal to the sphere at a point (x, y, z) on the sphere?

$$\tilde{N} = (x, y, z) / \| (x, y, z) \|$$
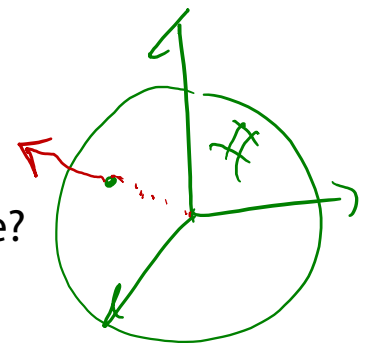
Note: the Trace project only requires you to handle a sphere of radius $r = 0.5$. This sphere may be arbitrarily transformed when placed in the scene, but everything works out in the end. More later!

Ray-plane intersection

Next, we will considering intersecting a ray with a plane.

To do this, we first need to define the plane equation.

Given a point S on a plane with normal $\mathbf{N}$, how would we determine if another point X is on the plane?

(Hint: start by forming the vector $X - S$.)

$$N \cdot (x - s) = 0$$

$$N \cdot x - N \cdot s = 0$$
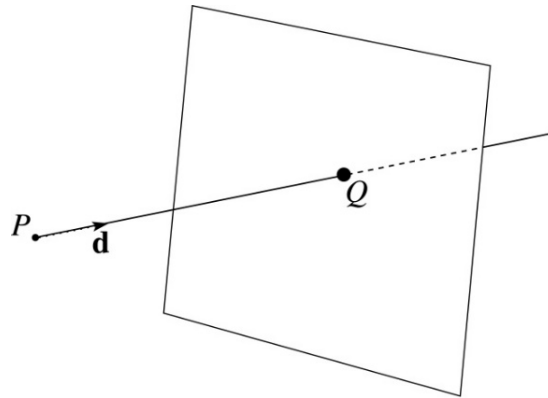
$$N \cdot x = \underbrace{N \cdot s}_{k}$$

$$N \cdot x = k$$

This is the plane equation!

Ray-plane intersection (cont'd)

Now consider a ray intersecting a plane, where the plane equation is:

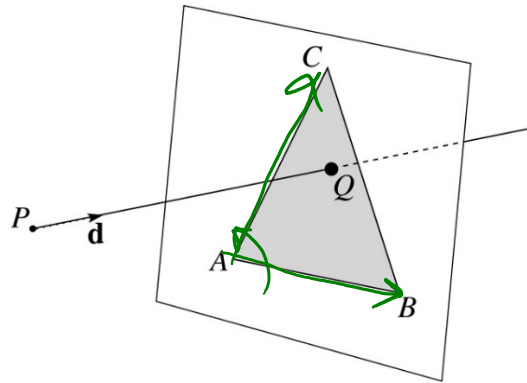$$\mathbf{N} \cdot X = k$$

We can solve for the intersection parameter (and thus the point) by substituting X with the ray $P + t\,\mathbf{d}$:

$$N \cdot (P + t\,d) = k$$

$$N \cdot P + N \cdot (t\,d) = k$$

$$N \cdot P + t\,N \cdot d = k$$

$$t\,(N \cdot d) = k - N \cdot P$$

$$t = \frac{k - N \cdot P}{N \cdot d}$$

$$N \cdot d = 0$$
$$\Rightarrow \text{ray} \parallel \text{plane}$$
$$\Rightarrow \text{no } \cap$$

Ray-triangle intersection

To intersect with a triangle, we first solve for the equation of its supporting plane.

How might we compute the (un-normalized) normal?

$$N = (B-A) \times (C-A)$$
$$\hat{N} = N/\|N\|$$

if Area$_\triangle = 0$ i.e. $N=0$ or unnormalized $N=0$

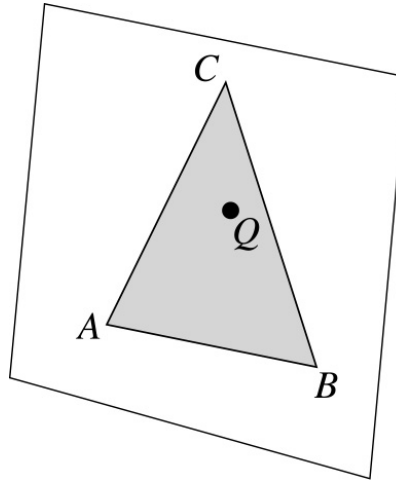Given this normal, how would we compute k?

$$\hat{N} \cdot X = k = \hat{N} \cdot A = \hat{N} \cdot B = \hat{N} \cdot C$$

Using these coefficients, we can intersect the ray with the triangle to solve for Q.

Now, we need to decide if Q is inside or outside of the triangle…

3D inside-outside test

One way to do this "inside-outside test," is to see if Q lies on the left side of each edge as we move counterclockwise around the triangle.

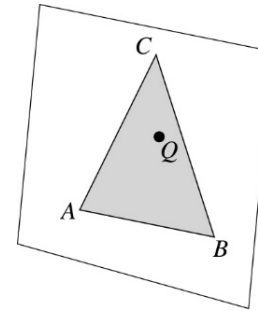How might we use cross and products to do this?

See HW + handout ⁿonline

Barycentric coordinates

As we'll see in a moment, it is often useful to represent Q as an **affine combination** of A, B, and C:

$$Q = \alpha A + \beta B + \gamma C$$
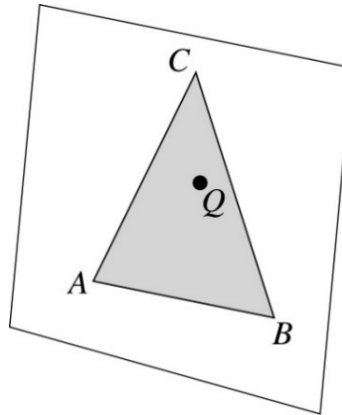
where:

$$\alpha + \beta + \gamma = 1$$

We call α, β, and γ, the **barycentric coordinates** of Q with respect to A, B, and C.

Computing barycentric coordinates

Given a point Q that is inside of triangle ABC, we can solve for Q's barycentric coordinates in a simple way:

$$\alpha = \frac{\text{Area}(QBC)}{\text{Area}(ABC)} \qquad \beta = \frac{\text{Area}(AQC)}{\text{Area}(ABC)} \qquad \gamma = \frac{\text{Area}(ABQ)}{\text{Area}(ABC)}$$
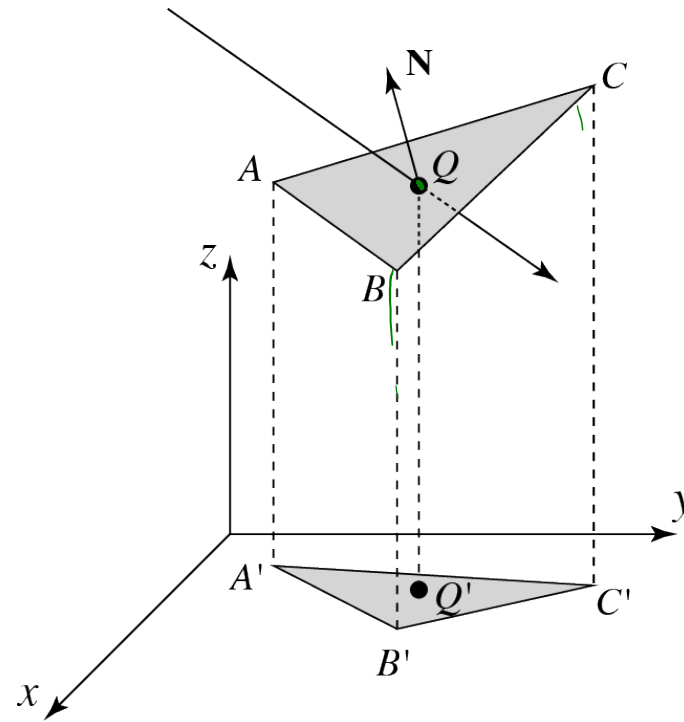

How can cross products help here?

See HW + handout

In the end, these calculations can be performed in the 2D projection as well!

Improvement: project down to 2D first

Without loss of generality, we can make this
determination after projecting down a
dimension:

$$Q = \alpha A + \beta B + \gamma C$$

$$M = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$MQ = M(\alpha A + \beta B + \gamma C)$$
$$= \alpha MA + \beta MB + \gamma MC$$
$$Q' = \alpha A' + \beta B' + \gamma C'$$

If Q' is inside of $A'B'C'$, then Q is inside of ABC.

Why is this projection desirable? *faster*

Which axis should you "project away"? (Hint: *axis w/ largest*
consider the triangle normal.) *| normal component |*

Interpolating vertex properties

The barycentric coordinates can also be used to interpolate vertex properties such as:

- material properties
- texture coordinates
- normals

For example:

$$k_d(Q) = \alpha k_d(A) + \beta k_d(B) + \gamma k_d(C)$$

Phong interpolated normals

Recall the idea of interpolating normal from the shading lecture, now updated to allow reflection and refraction.

Here's how it works:

1. Compute normals at the vertices.
2. Interpolate normals and normalize.
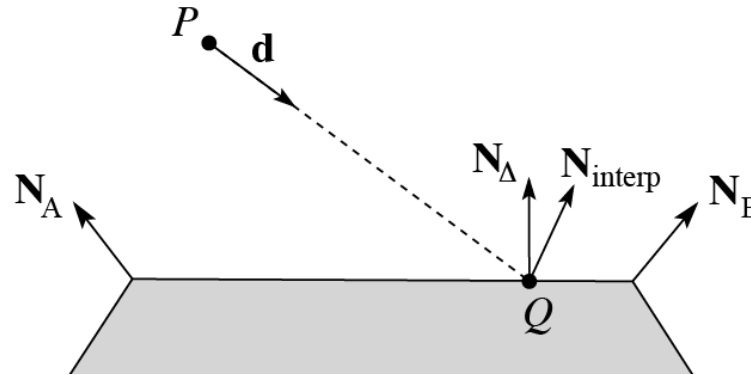3. Shade, reflect, and refract using the interpolated normals.

Interpolate normals

α, β, γ
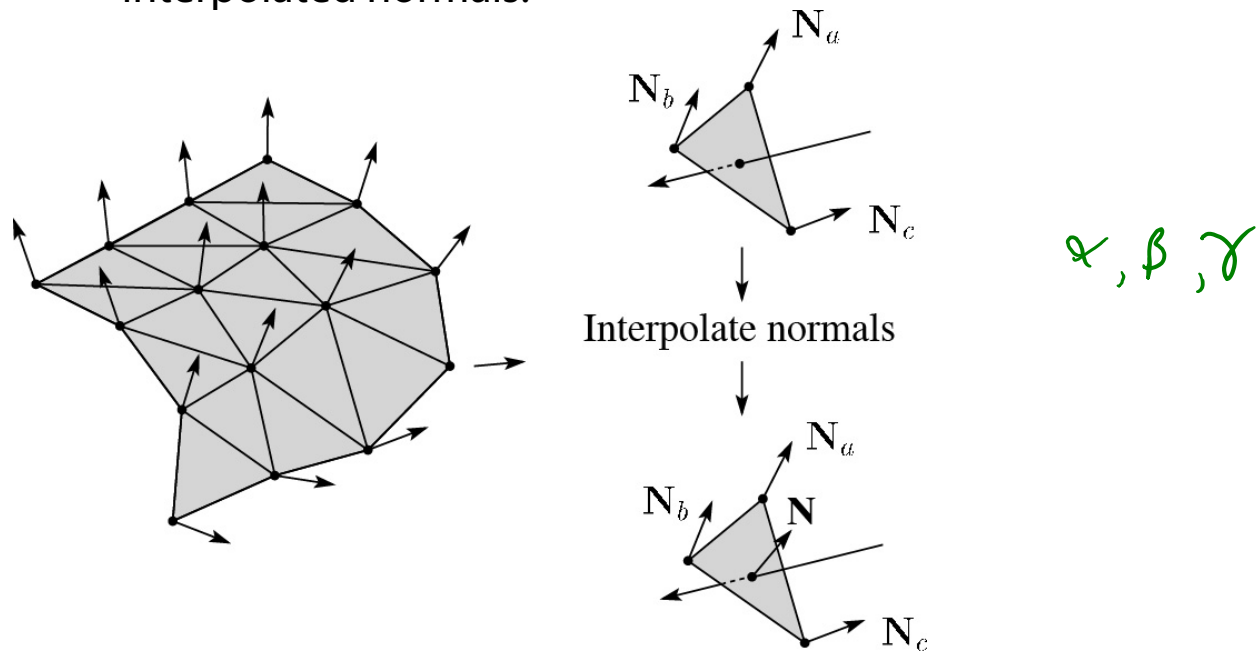
Q: How do we interpolate N_a, N_b, N_c to get N?
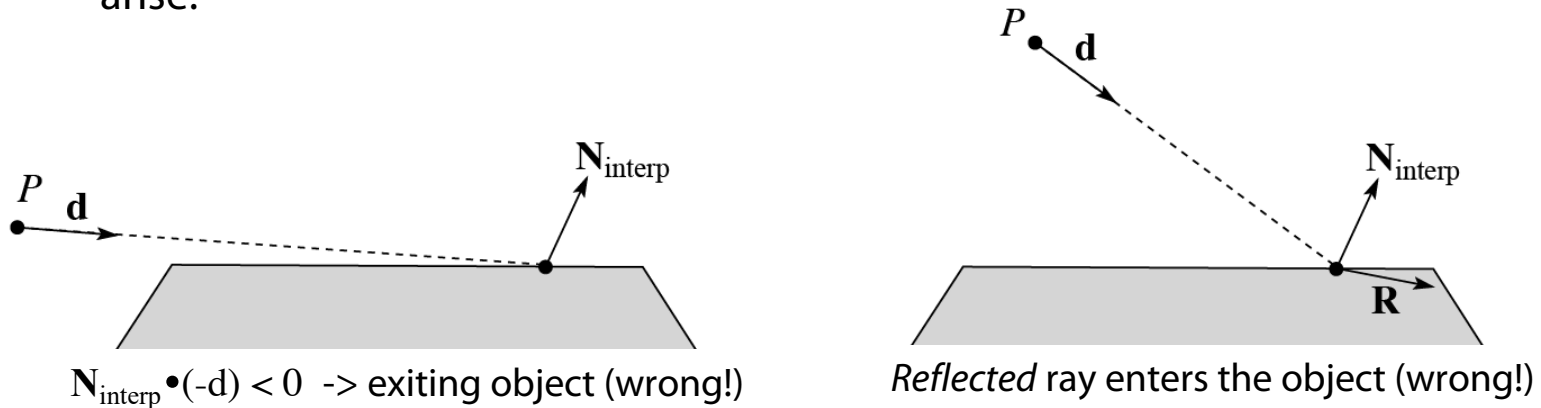
$$N = \alpha N_a + \beta N_b + \gamma N_c$$

$$\hat{N} = N / \|N\|$$

Interpolated normal in a ray tracer

As before, we will use the interpolated normal for shading, but a problem can arise when using this normal for other ray tracing purposes. Consider:

We see that the interpolated normal $\mathbf{N}_{\text{interp}}$ is of course different from the true geometric (triangle) normal $\mathbf{N}_\Delta$. Here are a couple problems that can arise:

$\mathbf{N}_{\text{interp}} \bullet (\text{-d}) < 0 \; \rightarrow$ exiting object (wrong!)

Reflected ray enters the object (wrong!)

A similar problem can arise for refraction, in which the refracted direction **T** is exiting the object, which is also wrong. There is no "right" answer for handling the discrepancy between normal and actual geometry.

43

Interpolated normal in a ray tracer, cont'd

We could play it "safe" and always use N_Δ, but then we will not get nice curved reflections and refractions even when these rays are valid. For Trace, do the following…

Determining when entering/exiting object:

- Use the geometric/true normal (N_Δ) when deciding whether you are entering/exiting an object.

Shading:

- Use N_{interp} for shading.

Reflection:

1. Start by using N_{interp} to compute reflection direction R.
2. If R is (incorrectly) entering the object, then re-compute R using N_Δ.

**Refraction**:

1. Start by using N_{interp} to check for Total Internal Reflection (TIR).
2. If TIR, then do not cast a refracted ray.
3. Else, use N_{interp} to compute refraction direction T.
4. If T is (incorrectly) exiting the object, then use N_Δ to re-test for TIR and (if no TIR) to re-compute T.

Epsilons

Due to finite precision arithmetic, we do not
always get the exact intersection at a surface.

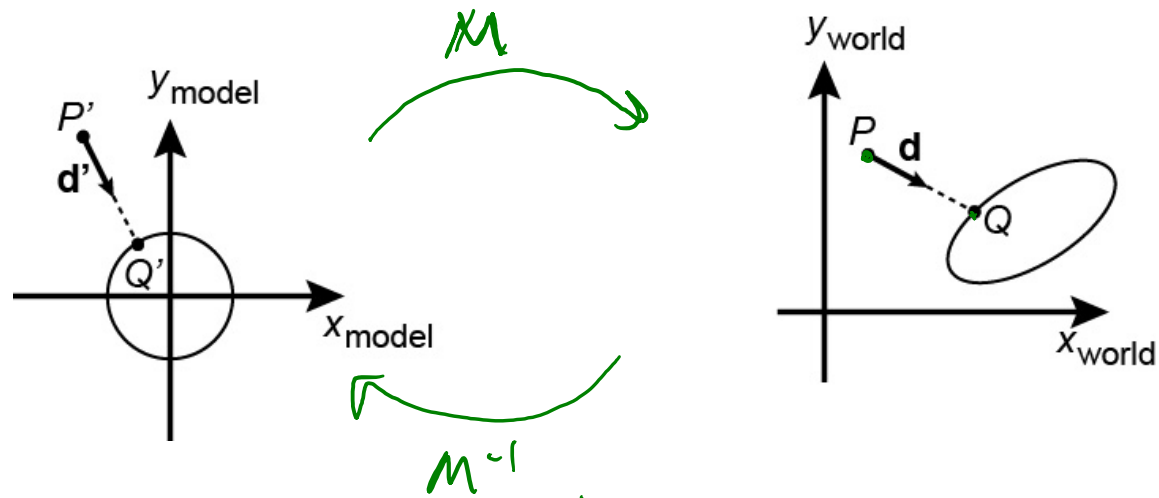**Q**: What kinds of problems might this cause?
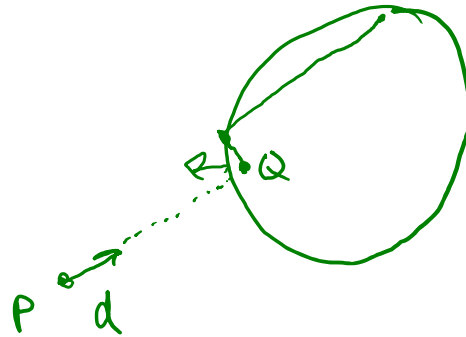
Q: How might we resolve this?

$$t < RAY\_EPSILON \Rightarrow no\ intersect$$

Intersecting with xformed geometry

In general, objects will be placed using transformations. What if the object being intersected were transformed by a matrix M?

Apply M^{-1} to the ray first and intersect in object (local) coordinates! **Note**: *do not normalize* **d**'!

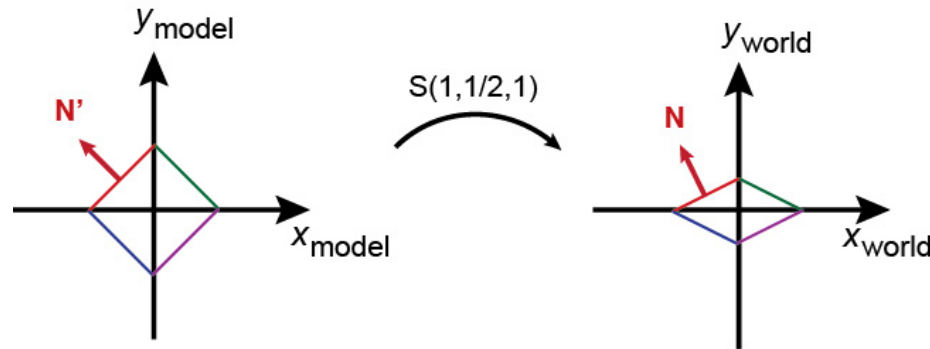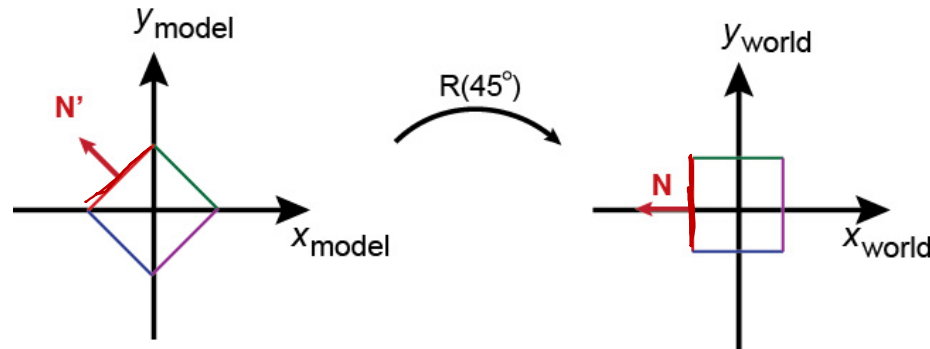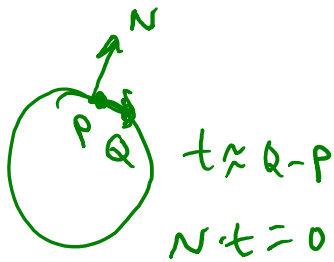$$M^{-1}\left(Q = P + t_n d \right)$$

$$M^{-1}Q = M^{-1}P + M^{-1}(t_n d)$$

$$M^{-1}Q = M^{-1}P + t_n M^{-1}d$$

$$Q' = P' + t_n d'$$

Intersecting with xformed geometry

$$R^T = R^{-1}$$

The intersected normal is in object (local) coordinates.
How do we transform it to world coordinates?

$$R^{-T} = R$$

$$t \approx Q - P$$

$$N \cdot t = 0$$

$$S = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$S^{-T} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$M = \begin{bmatrix} A & t \\ 0 & 1 \end{bmatrix}$$

$$N = A^{-T} N'$$

Summary

What to take home from this lecture:

- The meanings of all the boldfaced terms.
- Enough to implement basic recursive ray tracing.
- How reflection and transmission directions are computed.
- How ray-object intersection tests are performed on spheres, planes, and triangles
- How barycentric coordinates within triangles are computed
- How ray epsilons are used.