

Working with Light



Working with Light for Computer Graphics

Physics and Optics:

- Light is emitted from a light source
 - e.g. the sun, a light bulb, computer monitor, cell phone, etc.
- That emitted light impacts various objects, and may be reflected or absorbed
 - This reflection/absorption modifies the light
 - e.g. creating color, brightness, dullness/shininess, highlights, etc.
- In addition, light may pass (transmit) through various materials and (in doing so) be bent, scattered, etc.
 - e.g. prism, stained glass windows, water, etc.

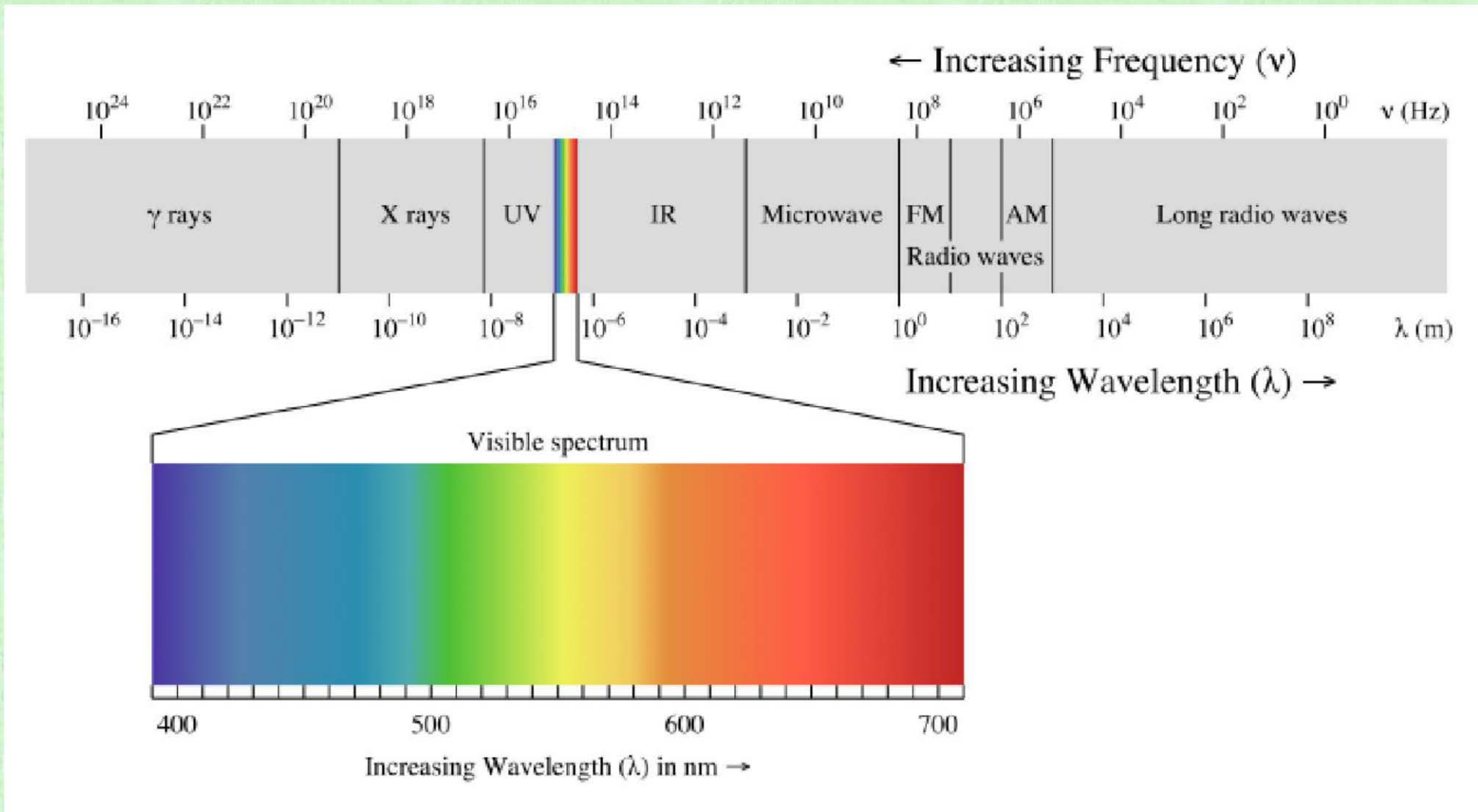
Human Perception:

- Eventually, some light may enter our eyes creating a signal
- Our brain creates an image based on the signals it gets from our eyes

Software/Hardware:

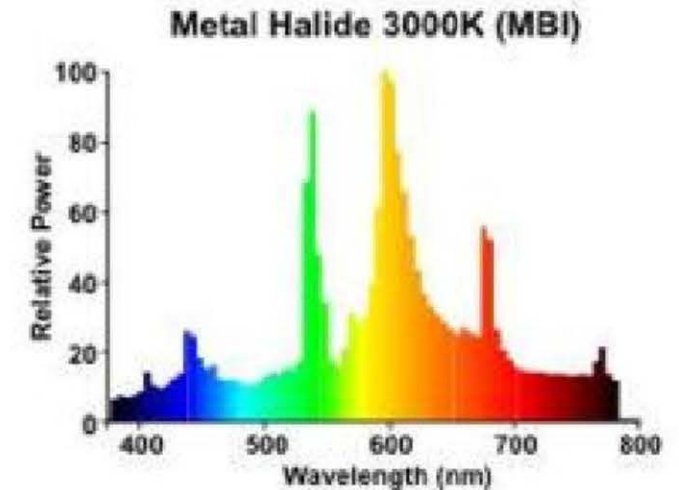
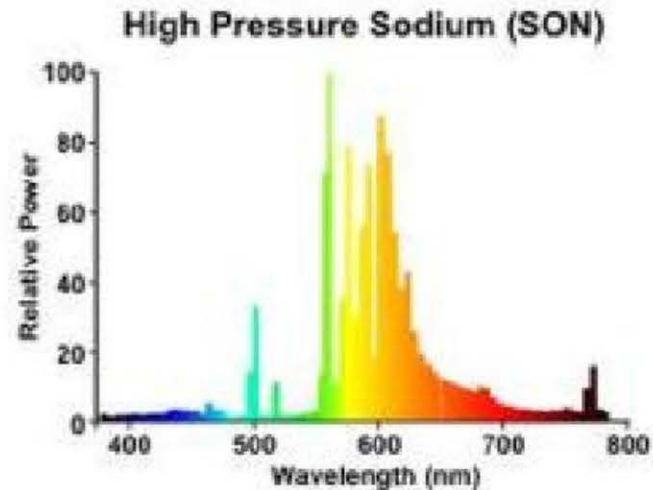
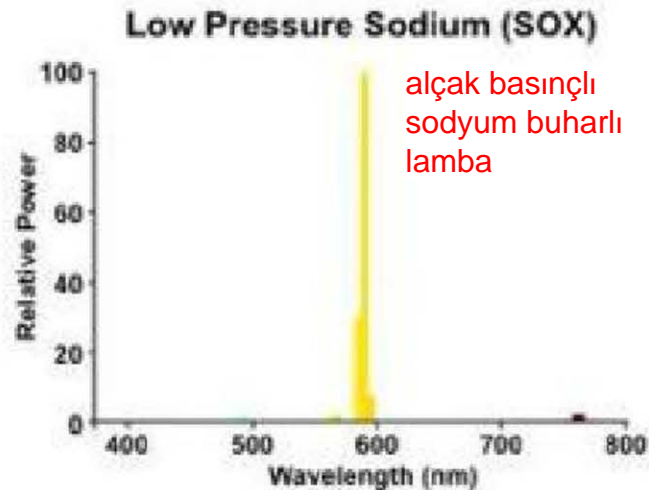
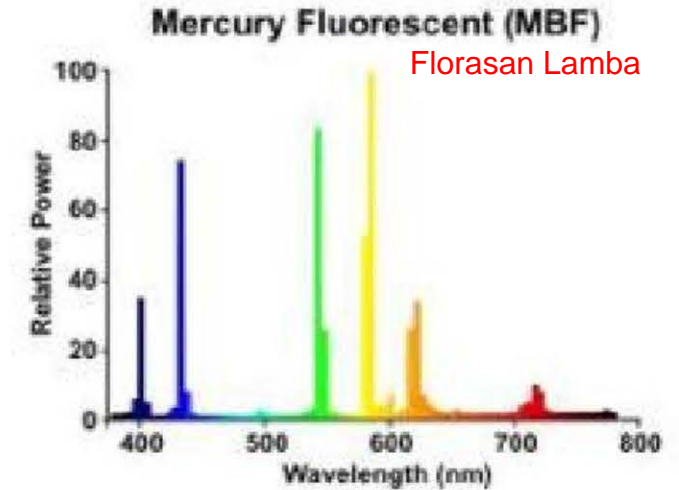
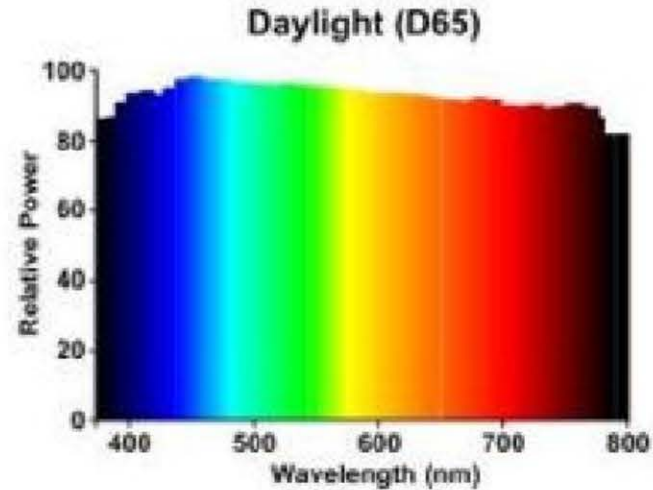
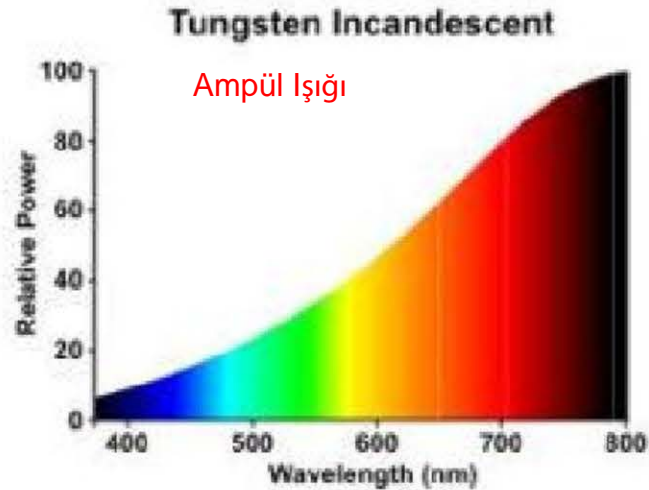
- Understanding the physics of light (i.e. optics) is important
- Understanding human perception allows for MANY optimizations/simplifications in both software/hardware
- The images we create ARE NOT intended to duplicate reality, only to fool humans into believing such

Electromagnetic Spectrum



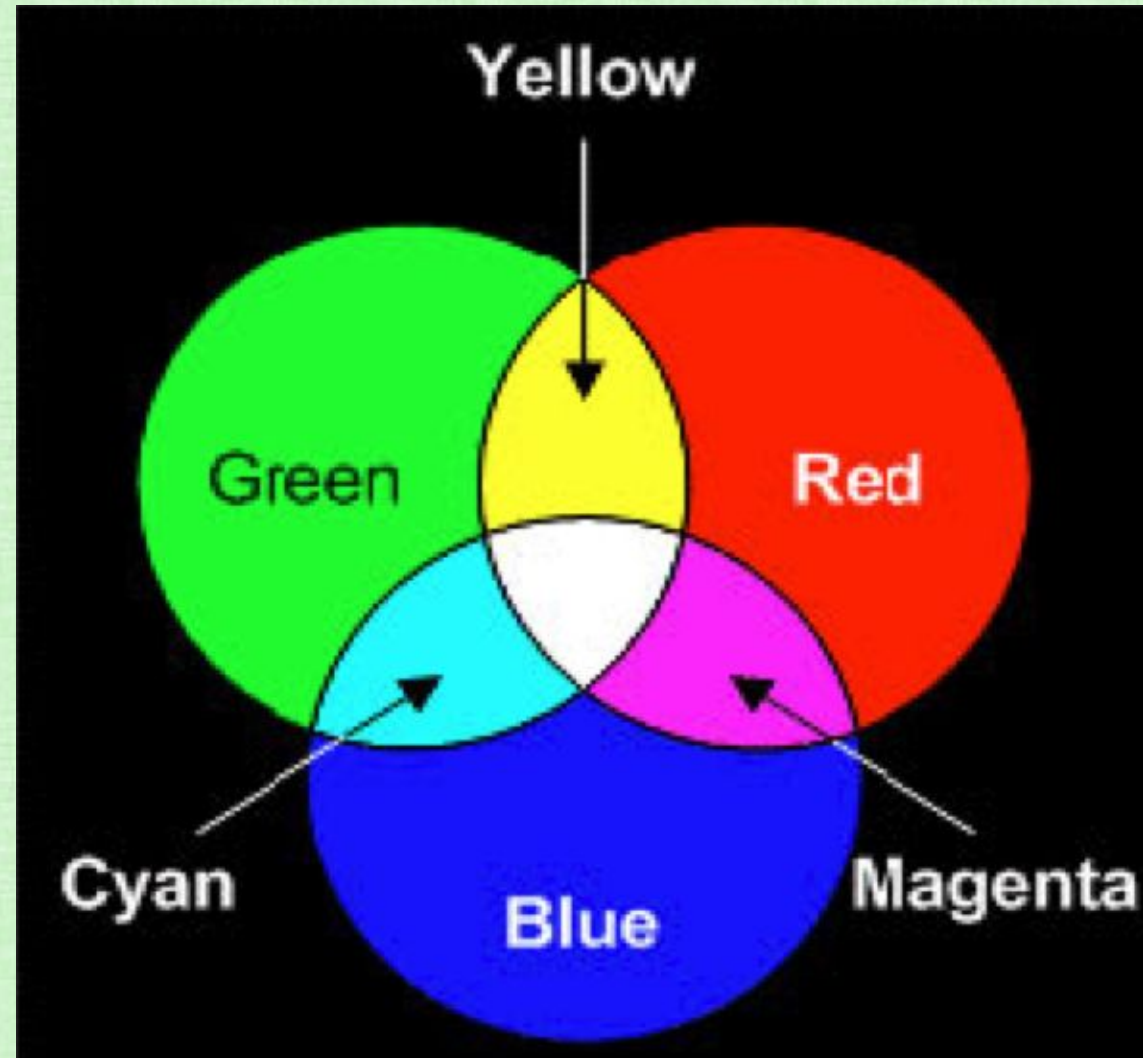
- The human eye can only see wavelengths between about 400 nm to 700 nm, so we focus on those

Relative Power Distribution of Lights



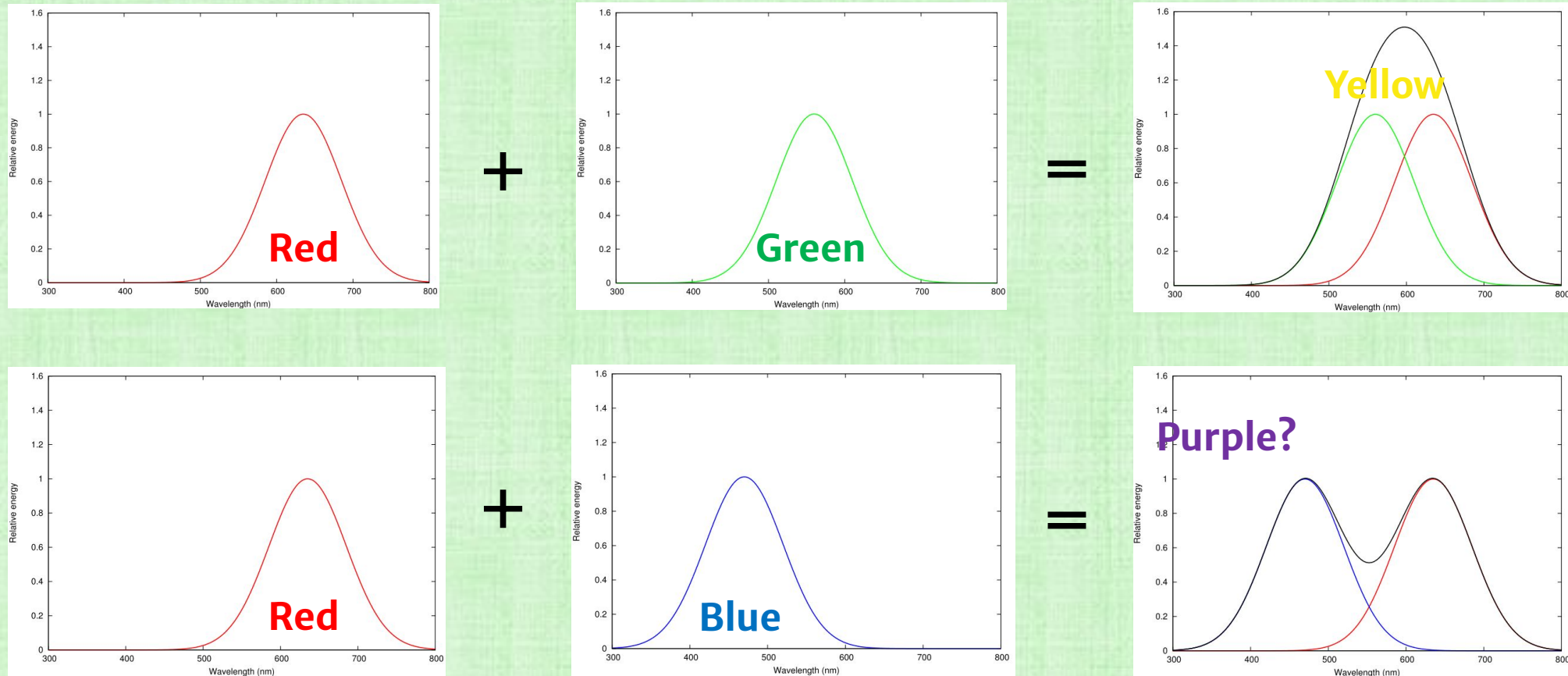
Adding Light Energy

The human eye perceives combinations of light energy as follows:

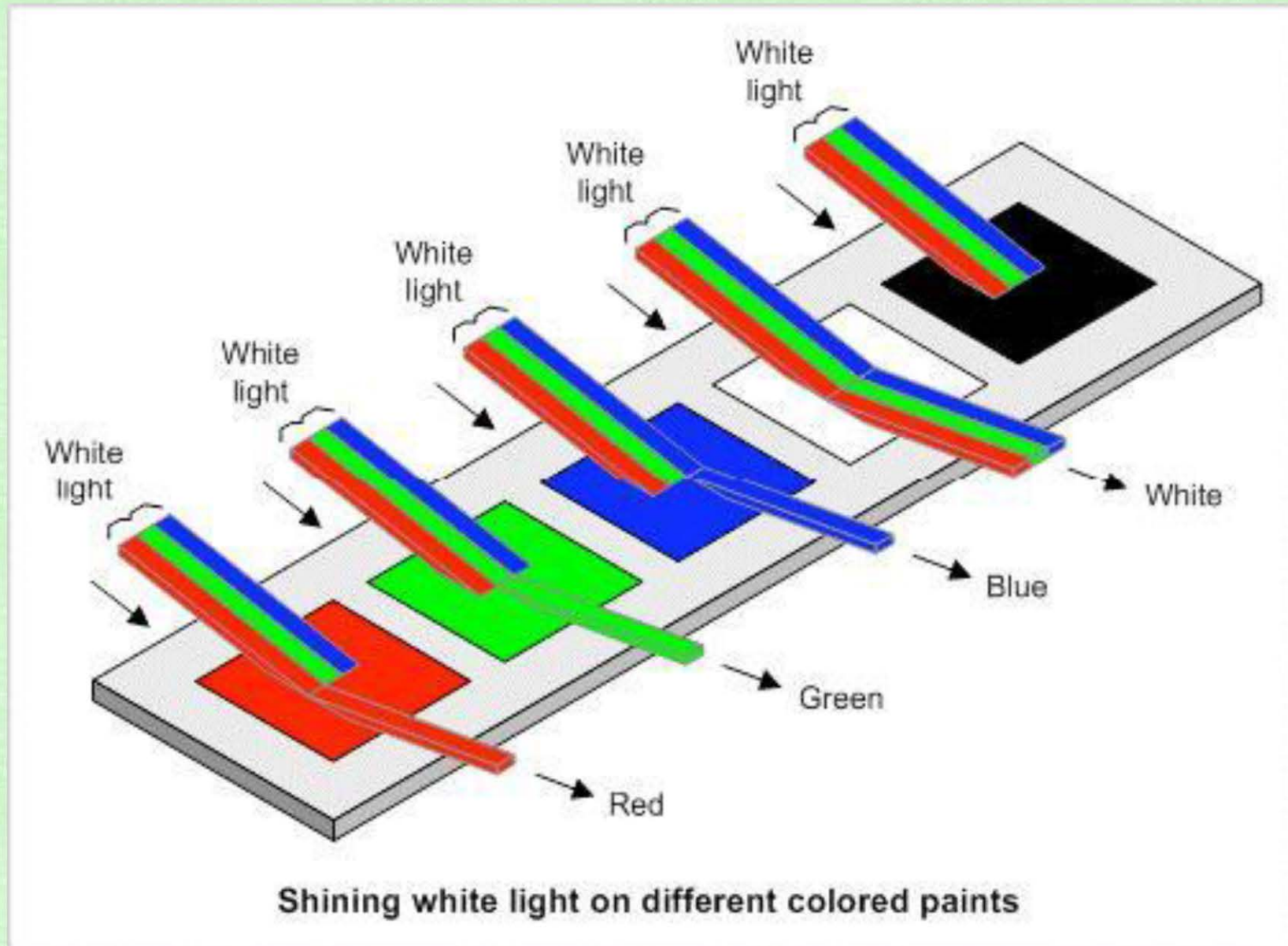


Adding Light Energy

- Energy adds (per wavelength) according to: $E(\lambda) = E_1(\lambda) + E_2(\lambda)$
- This leads to the following relative power distributions:



Absorbing & Reflecting Light Energy

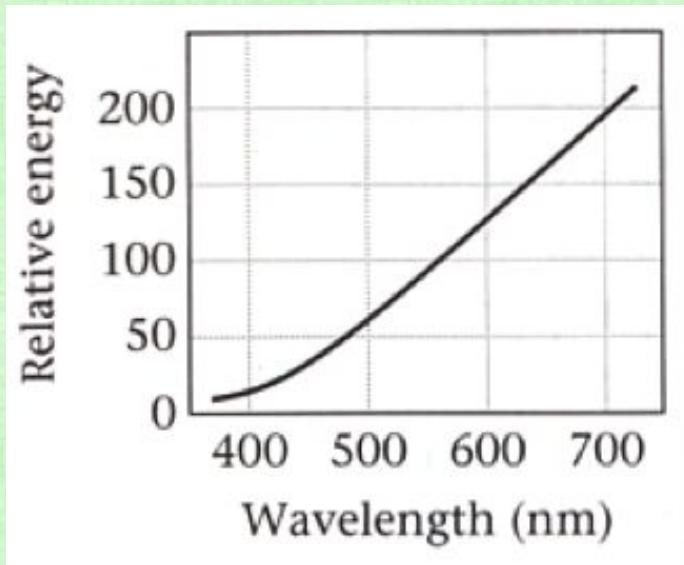


Absorbing & Reflecting Light Energy

Light energy (per wavelength) is either reflected or absorbed: $r(\lambda) + a(\lambda) = 1$
 $0 \leq r(\lambda), a(\lambda) \leq 1$

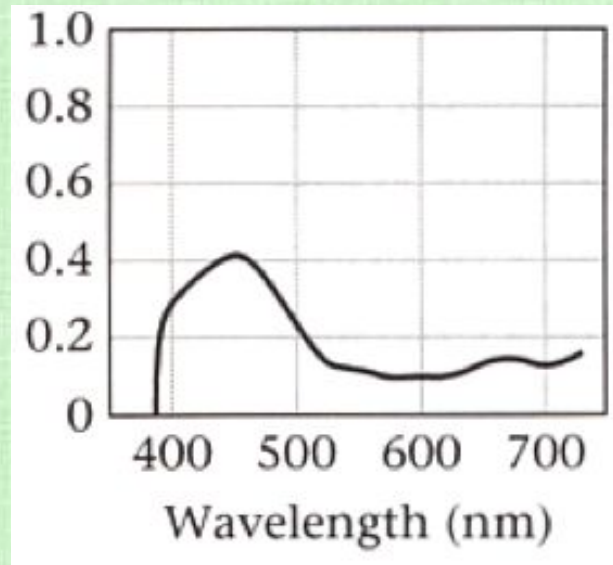
Reflected light energy (per wavelength) is computed via:

$$\textit{Reflected}(\lambda) = E(\lambda)r(\lambda) = E(\lambda)(1 - a(\lambda))$$



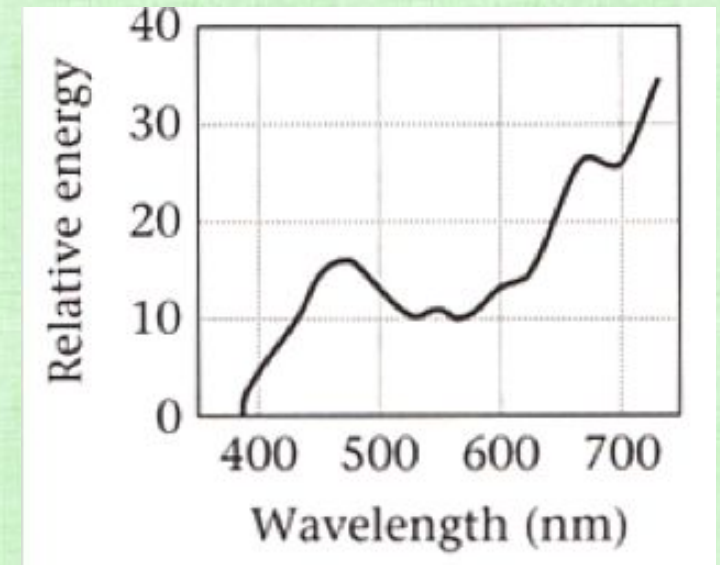
Incoming Energy

×



Material's Reflectance

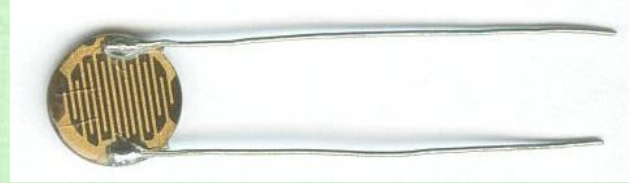
=



Reflected Energy

Sensor Absorption

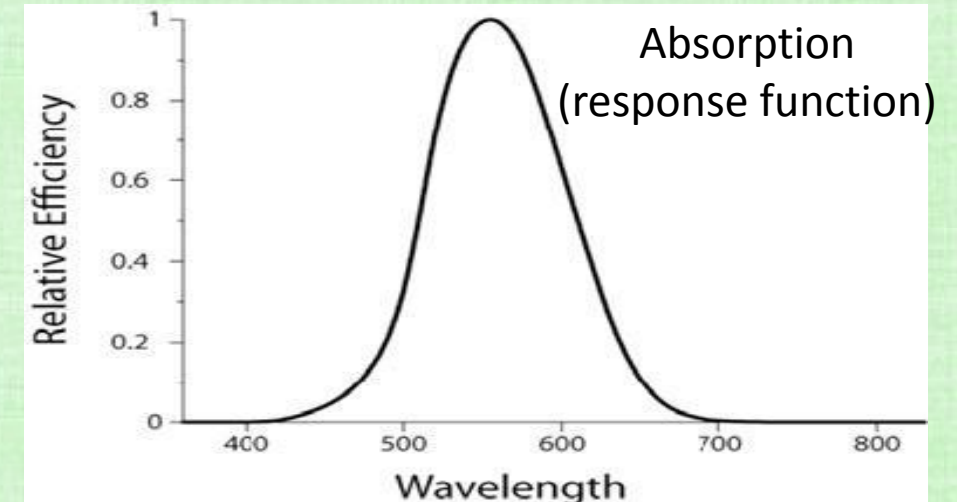
- Sensors absorb light (per unit time) and create a signal (per unit time)
- In order to be small (both biologically/mechanically), they are highly specialized



- Specialization leads to the (entire) sensor creating only **one signal** (per unit time)
- This conflates all the various wavelengths of light into a single signal (per unit time)

Signal power (energy per second):

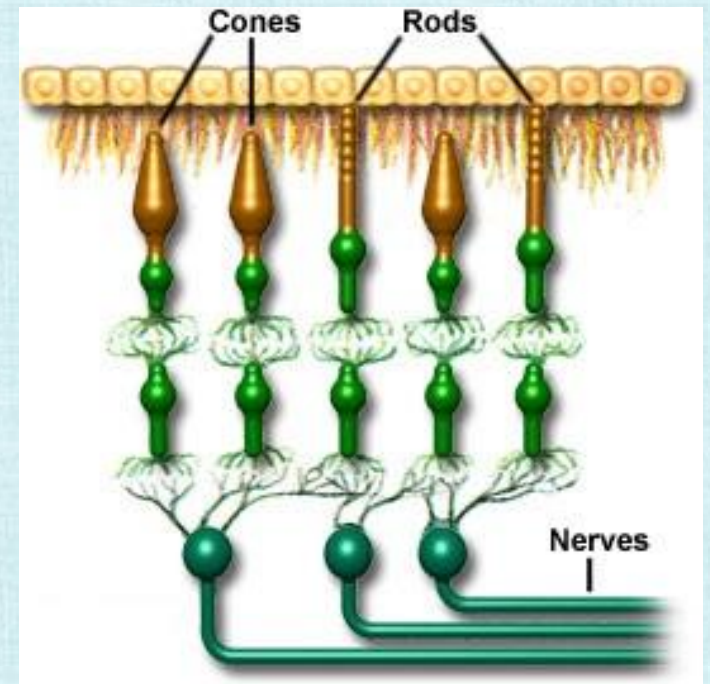
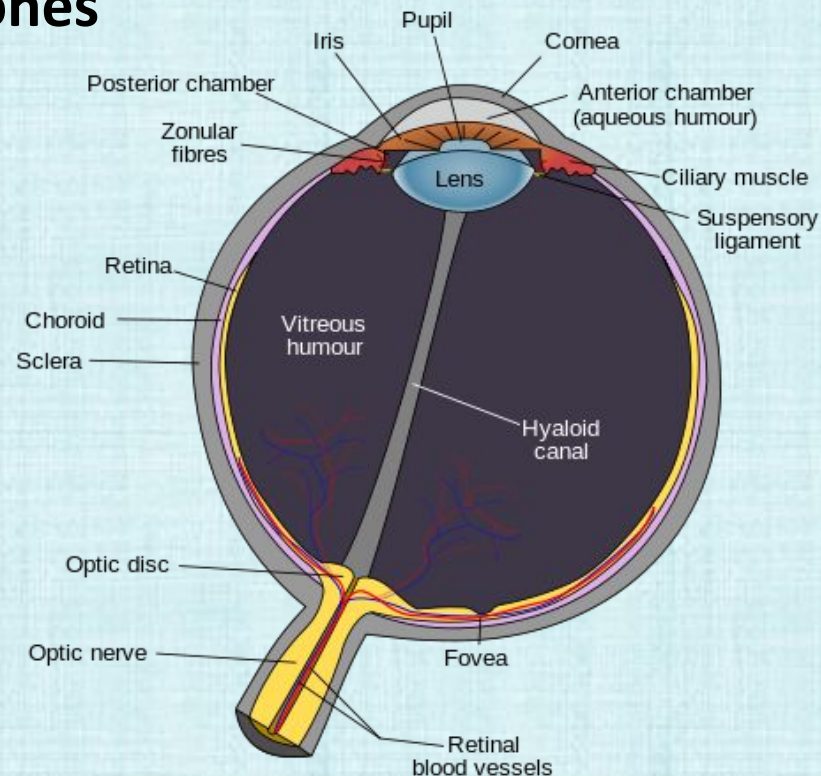
$$A = \int E(\lambda) a(\lambda) d\lambda$$



Not all wavelengths contribute equally to the final signal

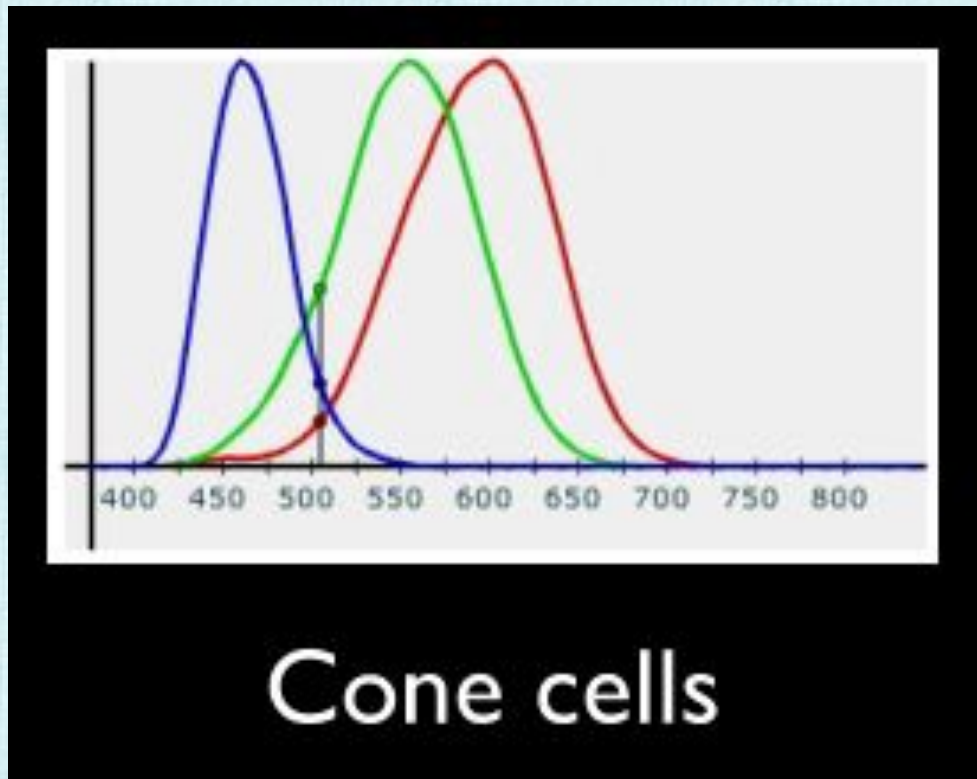
Sensors in the Human Eye

- The eye has 3 different types of cone sensors and 1 rod sensor
- Proteins in the cone/rod cells absorb photons changing the cell's membrane potential
- At night, cones are under-saturated (no/low/noisy signal), and rods produce most of the understandable signal
- During the day, the rod signals are over-saturated (all clamped to max), and we see primarily with cones

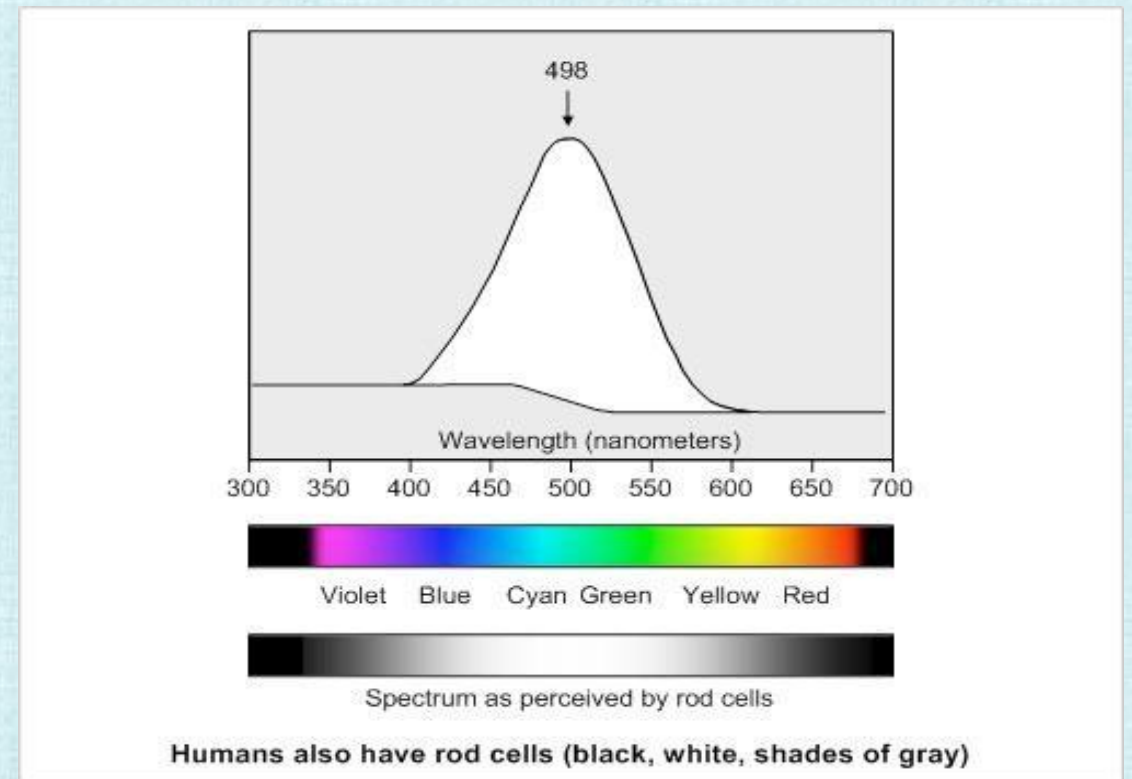


Response Functions for Cone/Rod Sensors

- The cone response functions vary based on the cone type (referred to as red/green/blue cones)
- The single rod sensor is interpreted as a black and white (or gray) light intensity



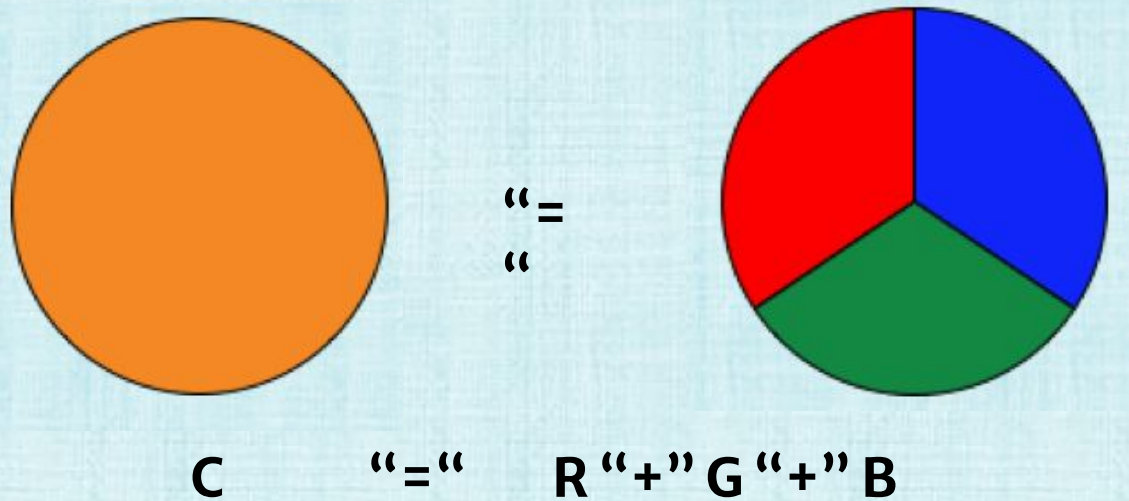
Note the similarity in red/green
(in regard to red/green colorblindness)



At night, the **single** signal from rods is interpreted as lacking color (or a shade of gray)

Trichromatic Theory

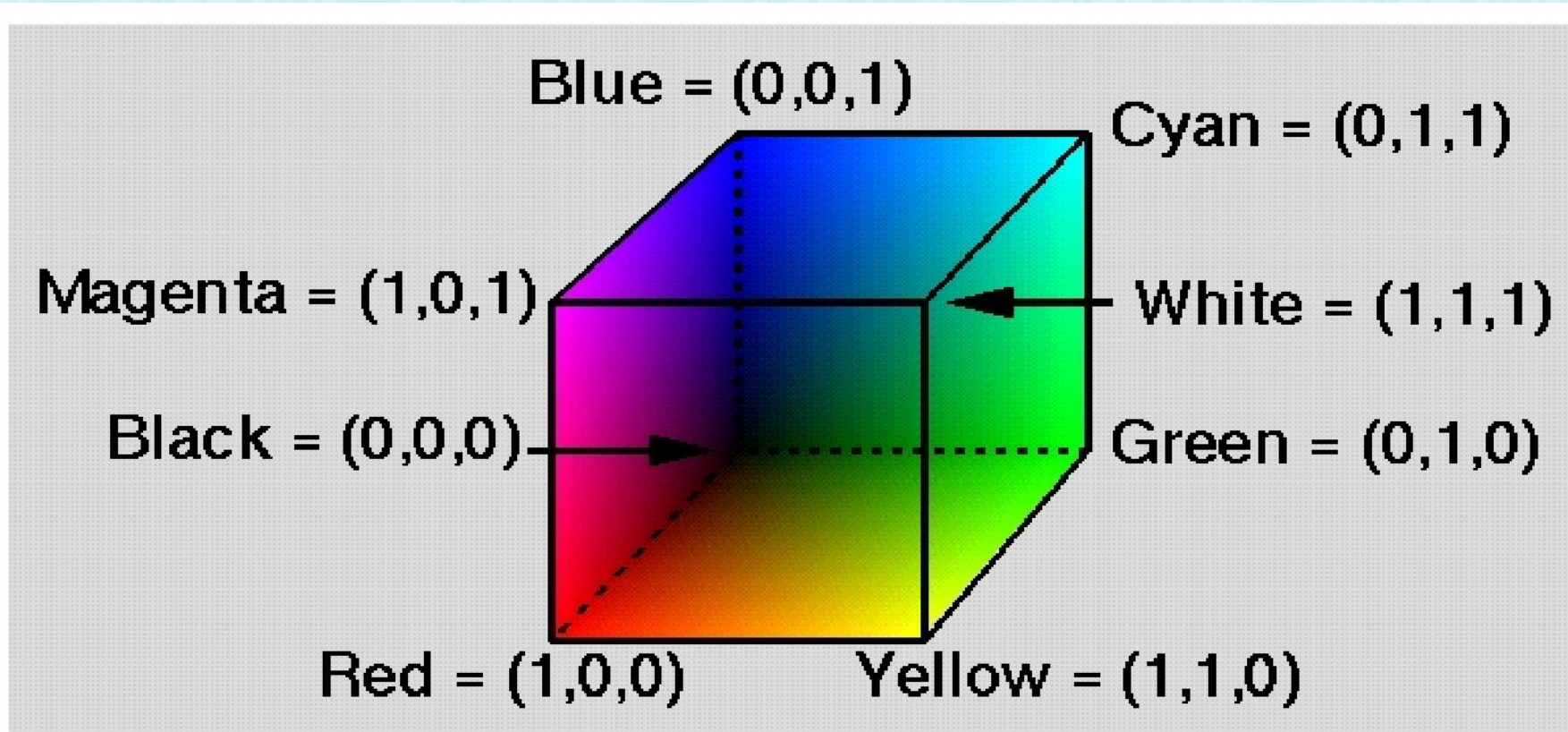
- Given any human perceived “color”:
 - Can adjust the brightness of 3 single wavelength lasers (e.g. R = 700 nm, G = 546 nm, B = 435 nm) to fool a human observer into “mistakenly” thinking that the laser combination matches that “color” (i.e. as a single wavelength)
- This is doable because each of the three cones can only send one signal (i.e., a 3 dimensional basis)



- Since the eye only perceives 3 signals (ignoring rods), only 3 signals are required for images, cameras, printers, displays, etc. (human perceived color is a 3 dimensional space!)
- Image formats store values in the R, G, and B channels

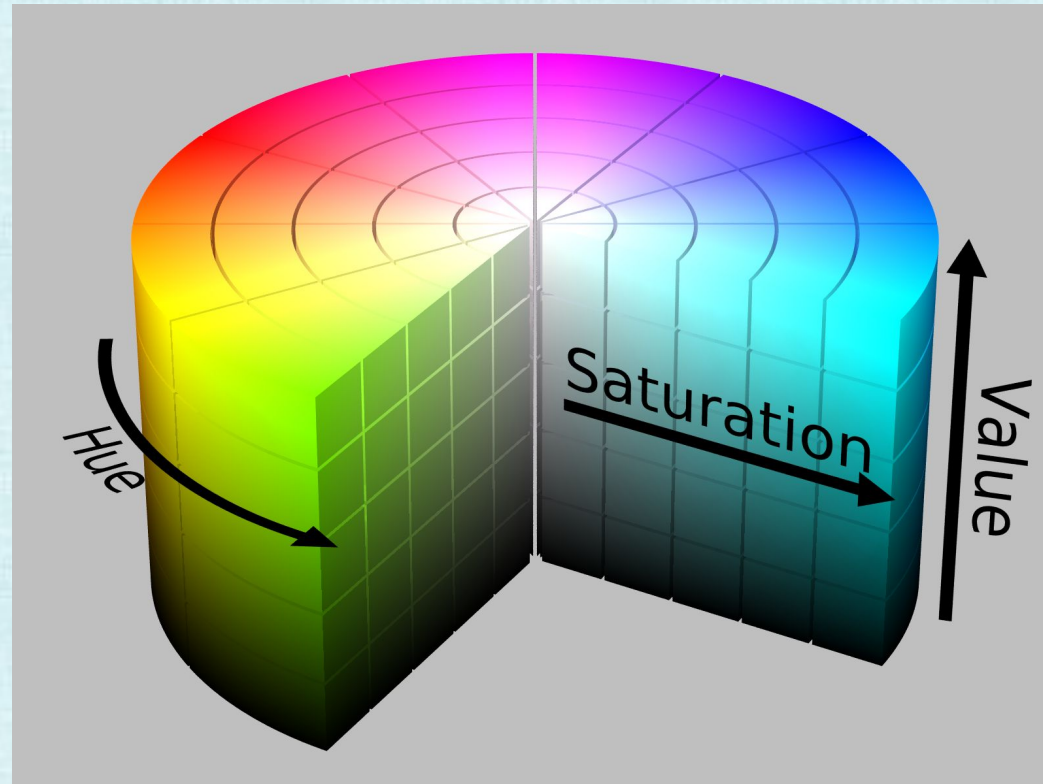
3D Color Space

- Map each primary color (Red, Green, Blue) to the unit distance along the x, y, z axes
- Black at (0,0,0), white at (1,1,1)
- The resulting RGB Color Cube represents all possible colors



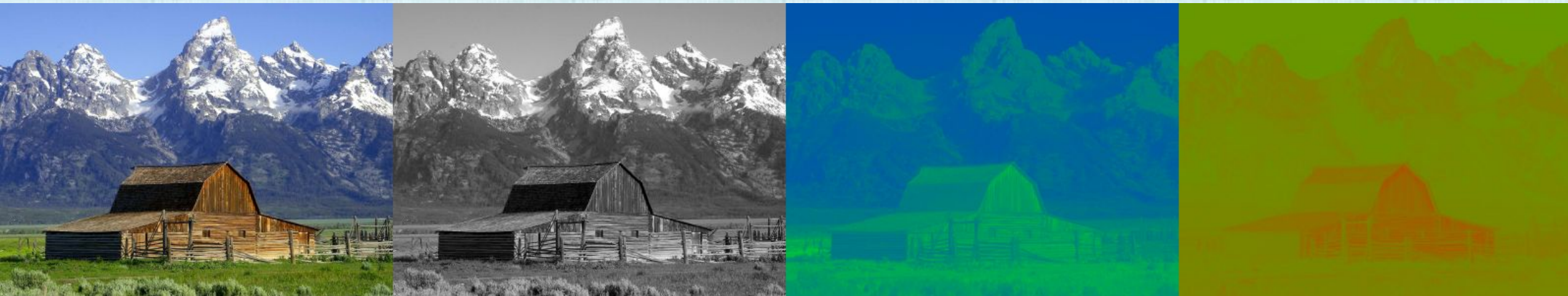
Cylindrical HSV Color Space

- A better 3D color space for user interfaces is based on Hue, Saturation, and Value (HSV)
- Hue: rainbow of colors (“wavelength”)
- Saturation: intensity for a particular color
- Value: lightness or darkness of a particular color



Luminance and Chrominance (YUV)

- Another 3D color space uses 1 luminance (Y) and 2 chrominance (UV) channels
- Black and White televisions used Y only, which perceptually holds the most spatial details
- Thus, can compress more aggressively in U & V than in Y



Original

Y

U

V

Interchangeability of Color Spaces

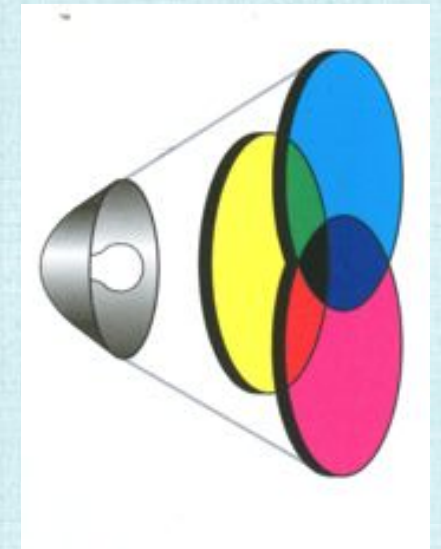
- One can map back and forth between any two 3D color spaces via matrix multiplication (using an appropriate matrix and its inverse)

- For example:
$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} .299 & .587 & .114 \\ -.14713 & -.28886 & .436 \\ .615 & -.51499 & -.10001 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

- Aside: note how important the Green channel is for the details in Y, as well as how unimportant the Blue channel is for those detail

Additive vs. Subtractive Color Spaces

- **Additive Color Space:**
 - Superimposed colored lights (e.g. phone display)
 - Add spectra (wavelength by wavelength)
 - $R + G + B = \text{white}$
- **Subtractive Color Space:**
 - Sequence of color filters (e.g. ink pigments or paint)
 - Multiply by all absorption coefficients (wavelength by wavelength)
 - $R + G + B = \text{black}$



Printers (CMYK)

- Printers use a subtractive color model
- Cyan, Magenta, Yellow (CMY) are the three primary colors of the subtractive color model
- The ink partially or entirely masks/filters/absorbs colors on a white background, reducing the light that would otherwise be reflected
- Equal mixtures of C, M, Y (ideally) produce all shades of gray
- However, in practice, ink mixtures do not give perfect grays
- In addition, it's difficult to get perfect alignment of the 3 inks
- Instead, most fine details are printed with the Key color (= K = black)
- This also reduces ink bleeding and the time for ink to dry (besides saving money on colored ink)

C
M
Y
K



Limited Spatial Resolution

- Sensors are have a non-zero size (or surface area)
- Thus, the number of signals per square inch is limited, based on how closely sensors can be packed together

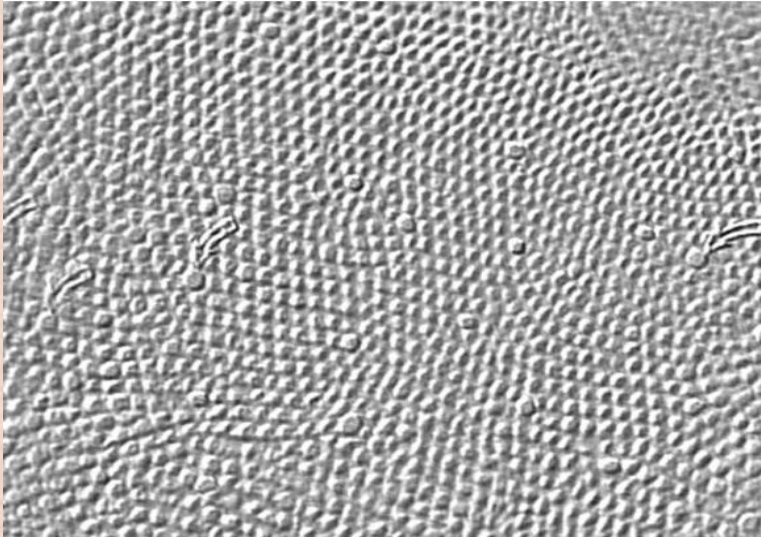
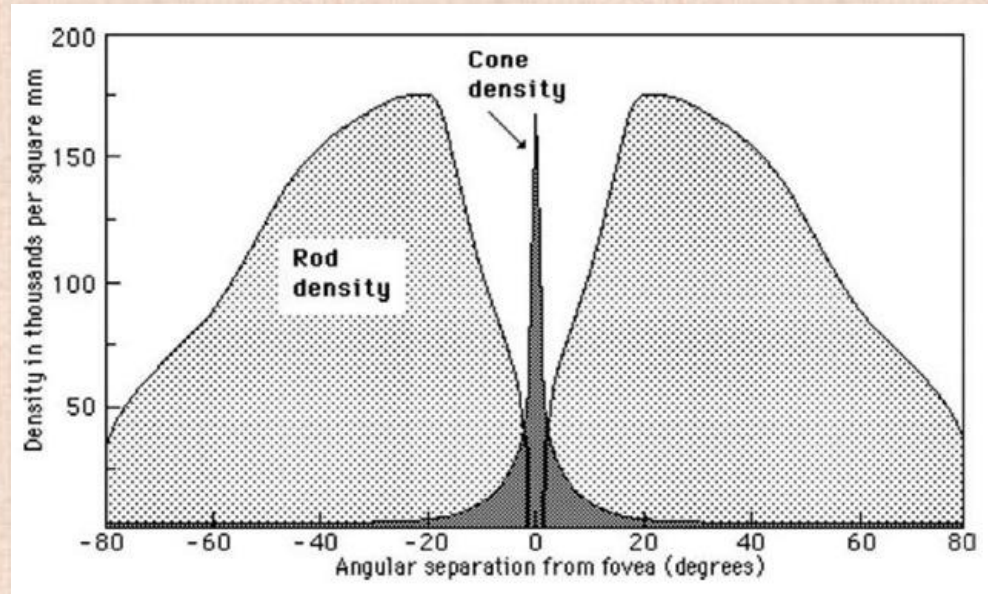


Fig. 13. Tangential section through the human fovea. Larger cones (arrows) are blue cones.



- Cones are densely packed near the center of the retina (the fovea), giving maximum detail for whatever the eye is looking directly at
- Rods have almost zero density at the fovea, which is why astronomers look out of the “side” of their eye

Distance Matters

- Closer/farther away objects project to larger/smaller areas on the cones, meaning more/less cones receive light signals from it
- Thus, closer objects can be seen in higher spatial detail than farther away objects

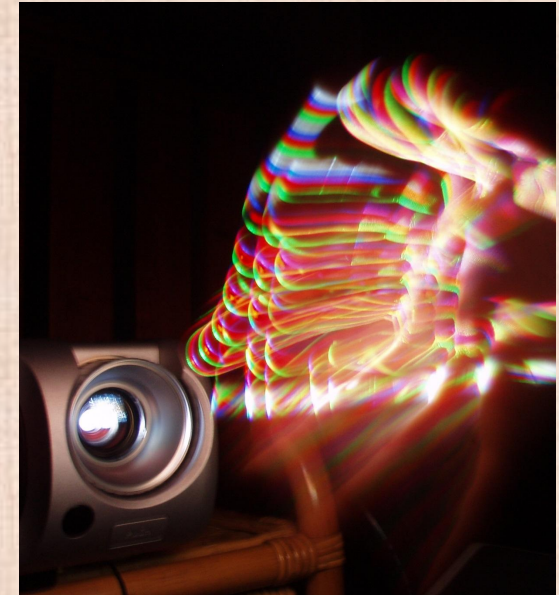
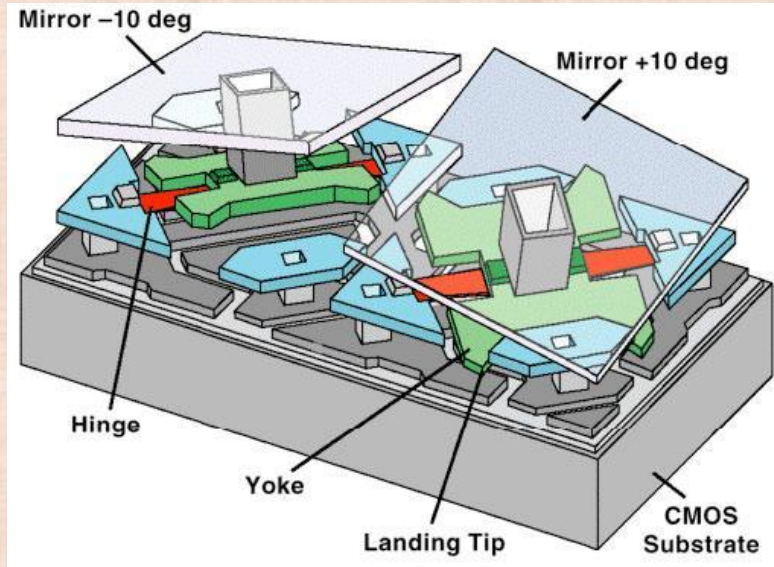


Resolution: 2048x1080
Size: 13.7m diagonal
4.29 dots per inch (dpi)

- Lower resolution is acceptable for a cinema screen, since viewers sit much farther away from it as compared to a cell phone - with 300+ pixels per inch (ppi)
- The number of cones per (image) feature is comparable between cinema screens and cell phones, given the differing distance of the observer

Projectors

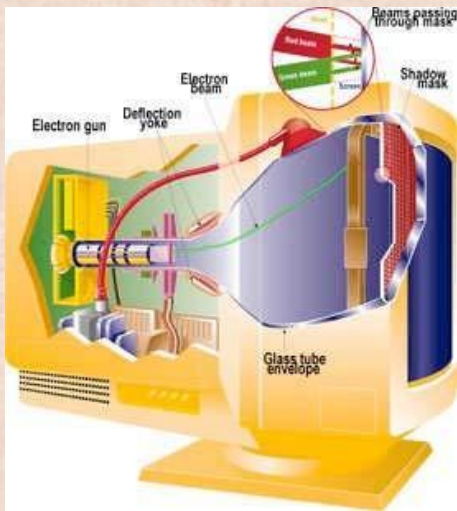
- Making large displays for far away viewers is difficult; thus, projectors are very important



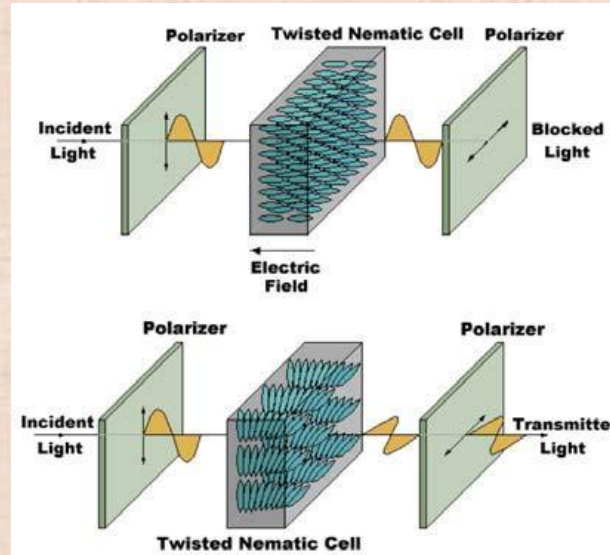
- A Digital Micro-Mirror Device (DMD) is the core component in Digital Light Processing (DLP) projectors
- Each mirror corresponds to one pixel, and has two states; it can either reflect the light into or out of the “pupil” of the projector
- Rapidly toggling a mirror between these two states produces brighter/dimmer light, controlled by the ratio of on-time to off-time

Display Technology

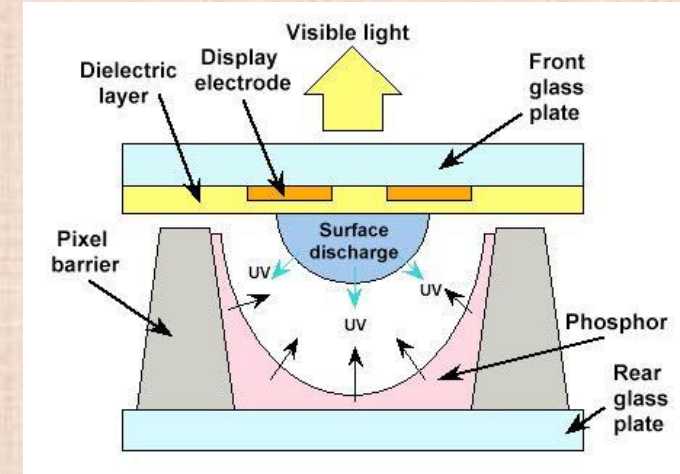
- The closer one sits to a display, the more cones per feature and thus more detail one can see
- Significant efforts have been spent on improving display (spatial) resolution



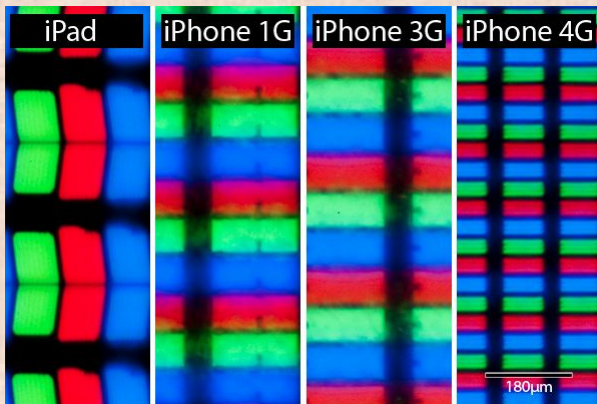
cathode ray tube (CRT)



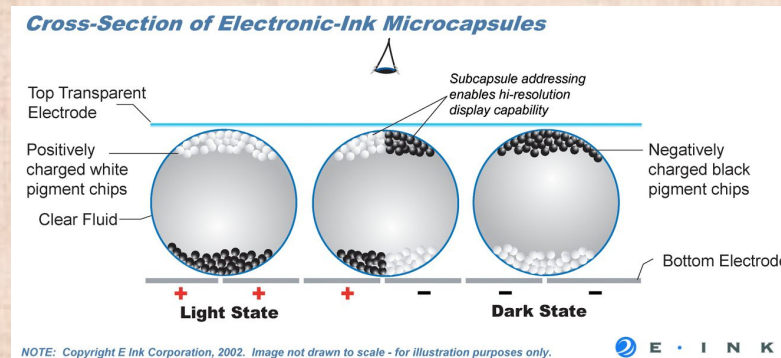
liquid crystal display (LCD)



plasma



iPhone/iPad LCD

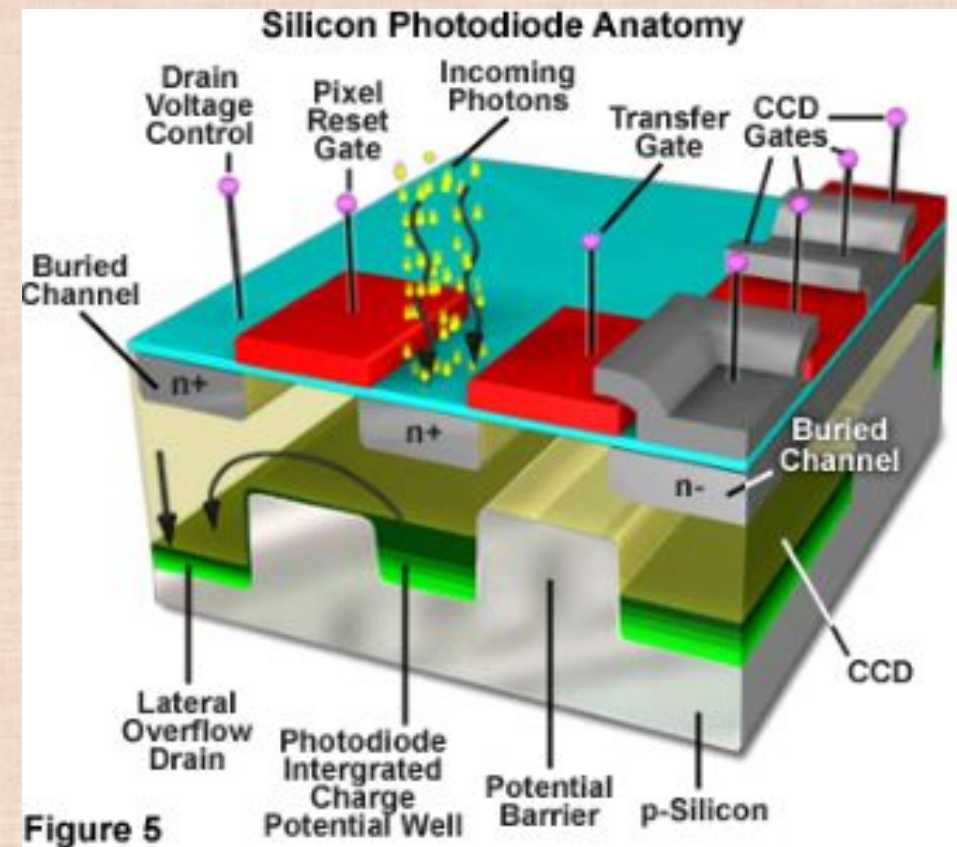
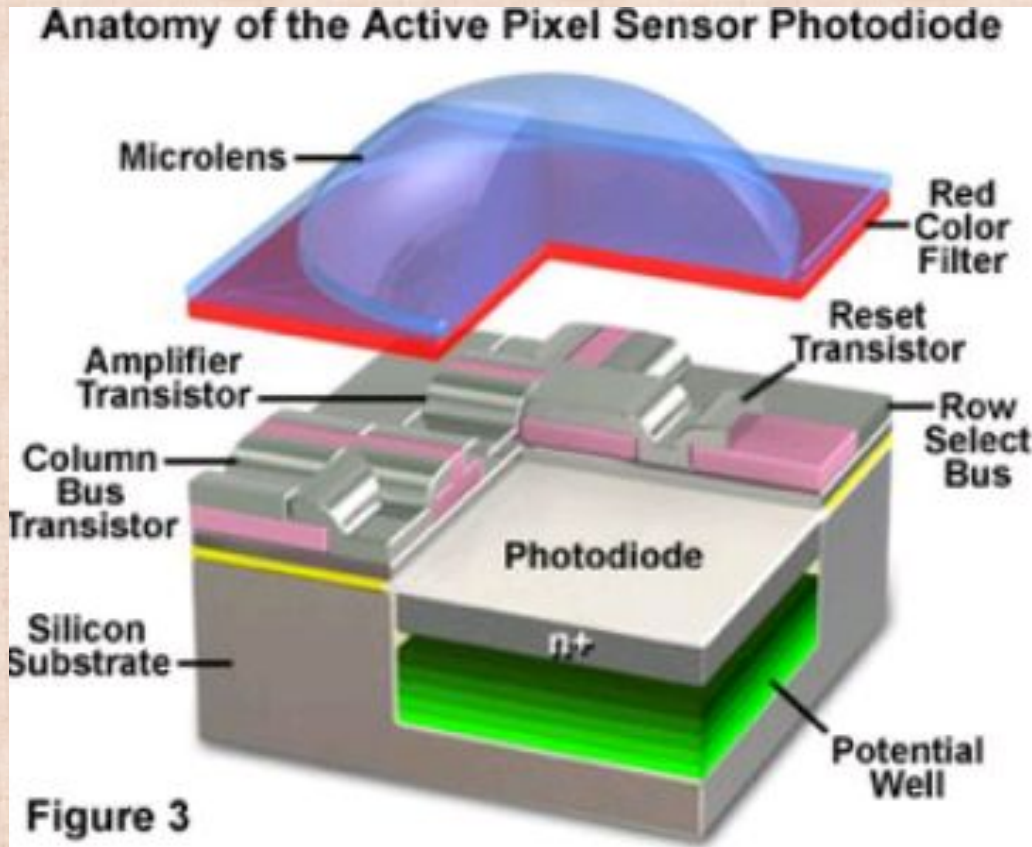


Electronic ink, ebook readers



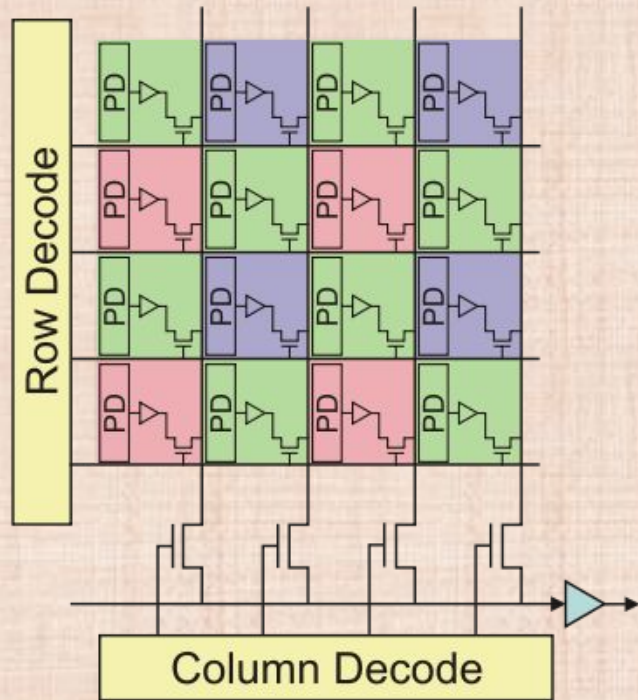
Camera Resolution

- Camera sensors use the photoelectric effect to generate an electron when hit by a photon (with some efficiency/probability)
- They are quite complex and, like cones, take up physical space
- This limits the resolution of what they can capture (just like for cones in the eye)

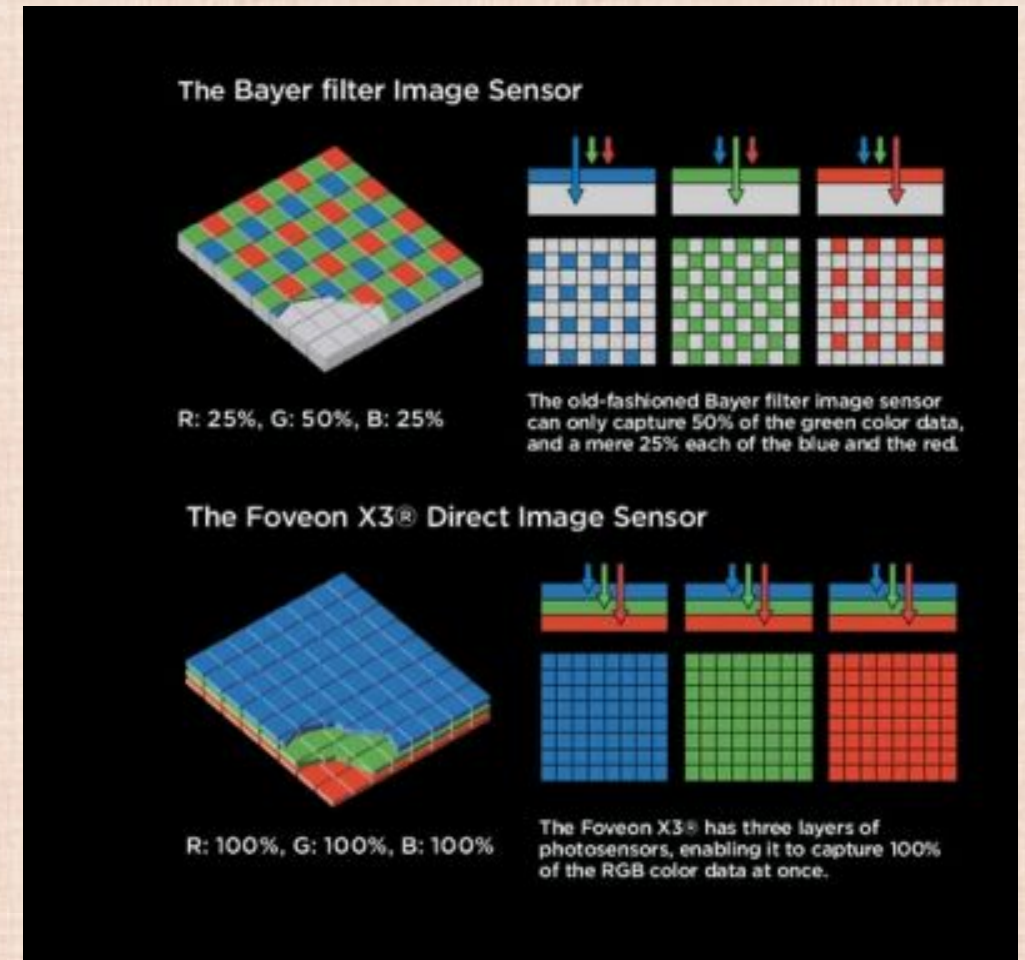


Camera Resolution

- Each camera sensor records incoming light energy per second (power)
- Each captures only one signal (per unit time) for its entire 2D spatial area
- Color filters are used to limit incident light to a particular color (so the same sensor can be used for every color)

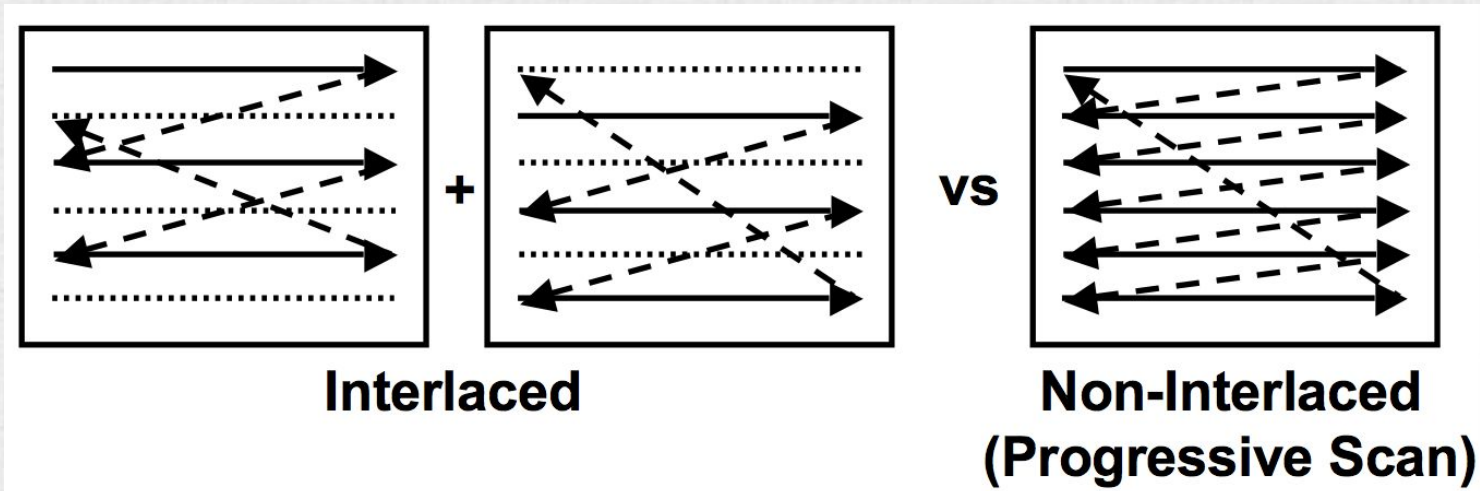


Note the doubled number of Green sensors, due to that color's importance in capturing spatial details



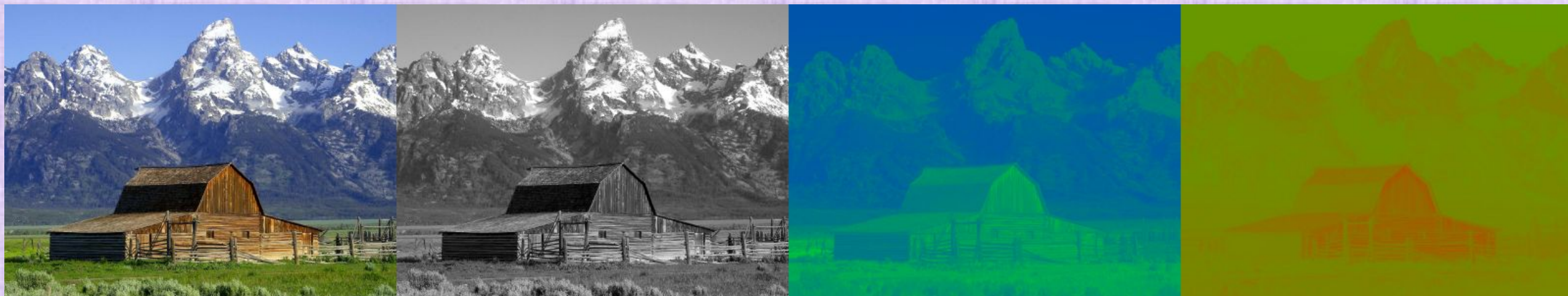
Aside: Temporal Resolution

- For moving images (animations), 16 Hz (or more) is required for humans to **not** interpret them as a series of still images
 - Movies are recorded at 24 images per second
 - TV broadcasts have 30 images per second
- Flicker fusion threshold: frequency at which an intermittent light stimulus appears to be steady to the observer
- Even though motion seems continuous at 24-30 images per second, the brightness may still seem to flicker
 - Movies are thus refreshed at 48 or 72 Hz (each image is projected 2 or 3 times)
 - Computer monitors refresh at 60-80 Hz (or more) independent of what is being displayed
 - TVs (used to) use interlacing to approximate 60 Hz, showing half of each frame at a time



Brightness (Luminance)

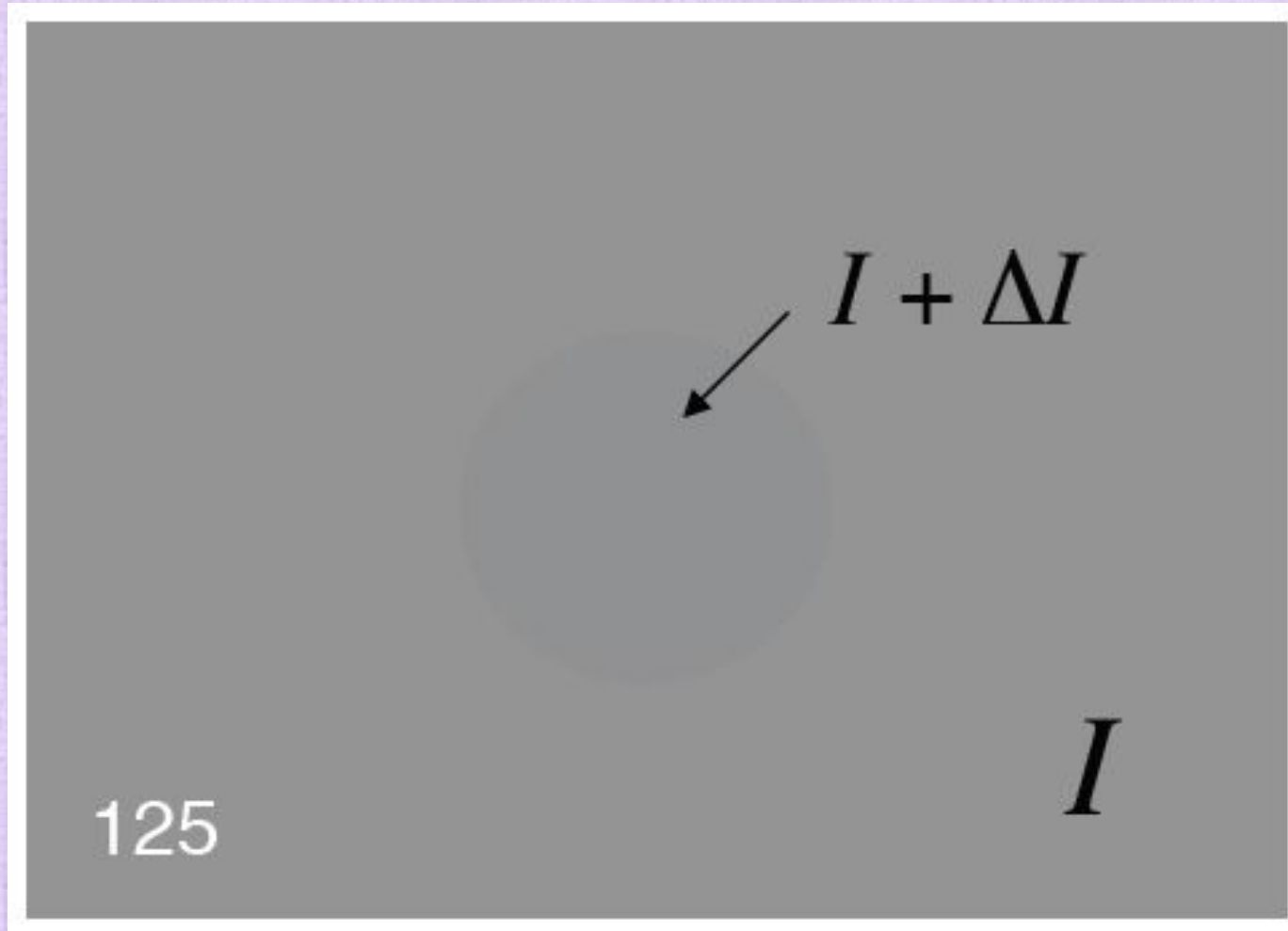
- The human eye is much more sensitive to spatial variations in brightness (so-called gray scale) than to spatial variations in color
- The 3 images on the right add together to give the image on the left
- Notice which images on the right has the most spatial details



Original Image = Greyscale + Color 1 + Color 2

Brightness Discrimination Experiment

- Changing the brightness (intensity) of the circle by 1 to 2% makes it just noticeable to most observers



Discretizing Brightness

- Since our eye can see small brightness changes, we need many levels for brightness
- Otherwise, changing brightness by the smallest amount would look discontinuous
- Thus, we typically use 256 levels for brightness
- That is, we store each color (R, G, B) with an integer ranging from 0 to 255
- High Dynamic Range (HDR) image formats use an even larger range than 0-255

32 levels



64 levels



128 levels



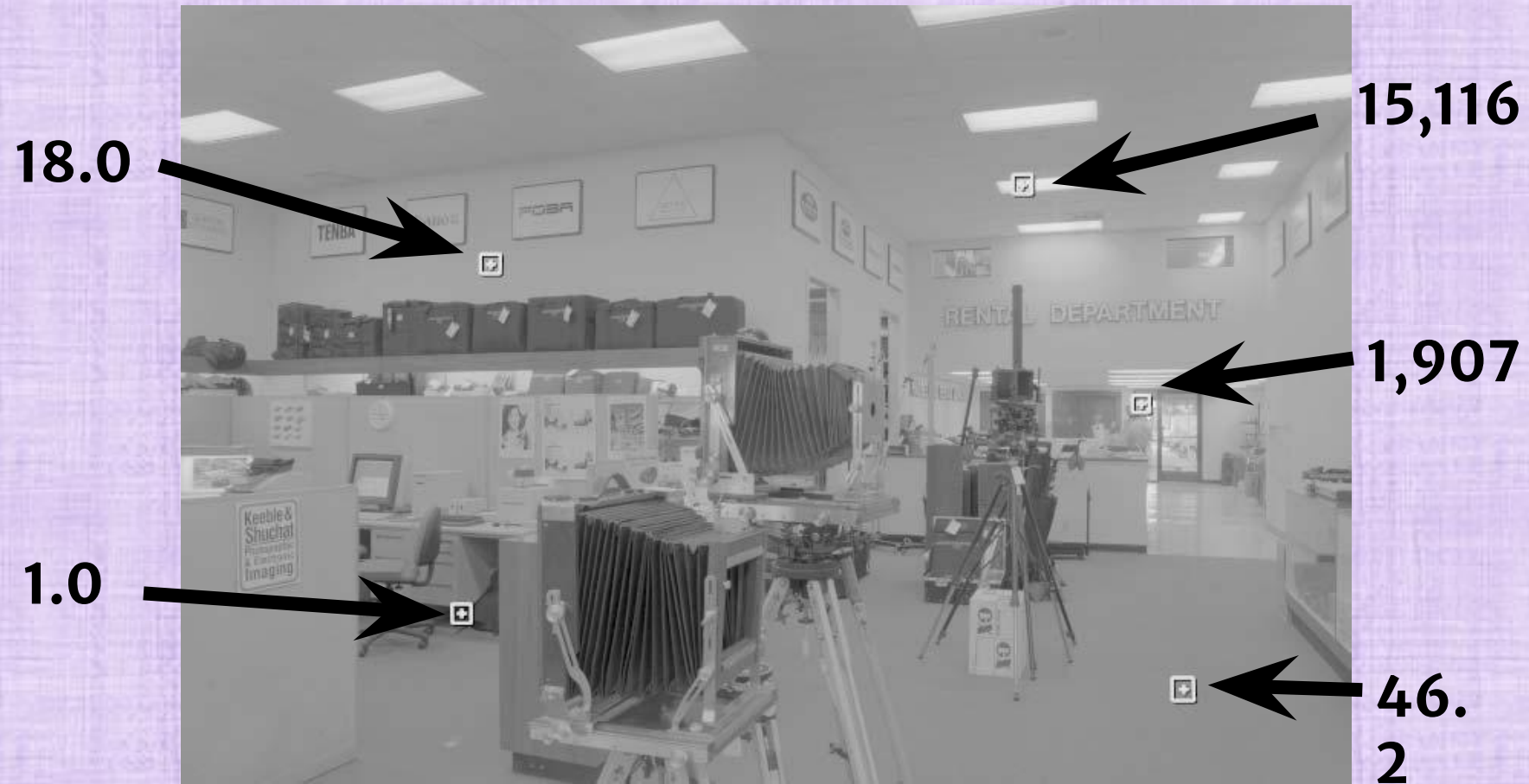
256 levels



Dynamic Range

- **World:**
 - Possible: 100,000,000,000:1 (from the sun to pitch black)
 - Typical real-world scenes: 100,000:1
- **Human Eye:**
 - Static: 100:1
 - Dynamic: 1,000,000:1 (as the eye moves, it adaptively adjusts exposure by changing the pupil size)
- **Media:**
 - Newsprint: 10:1
 - Glossy print: 60:1
 - Samsung F2370H LCD monitor: static 3,000:1, dynamic 150,000:1
 - Static contrast ratio is the luminance ratio between the brightest white and darkest black within a **single** image
 - Dynamic contrast ratio is the luminance ratio between brightest white possible (on any image) and the darkest black possible (on any image) on the same device
 - The contrast ratio in a TV monitor specification is measured in a dark room. In normal office lighting conditions, the effective contrast ratio typically drops from 3,000:1 to less than 200:1

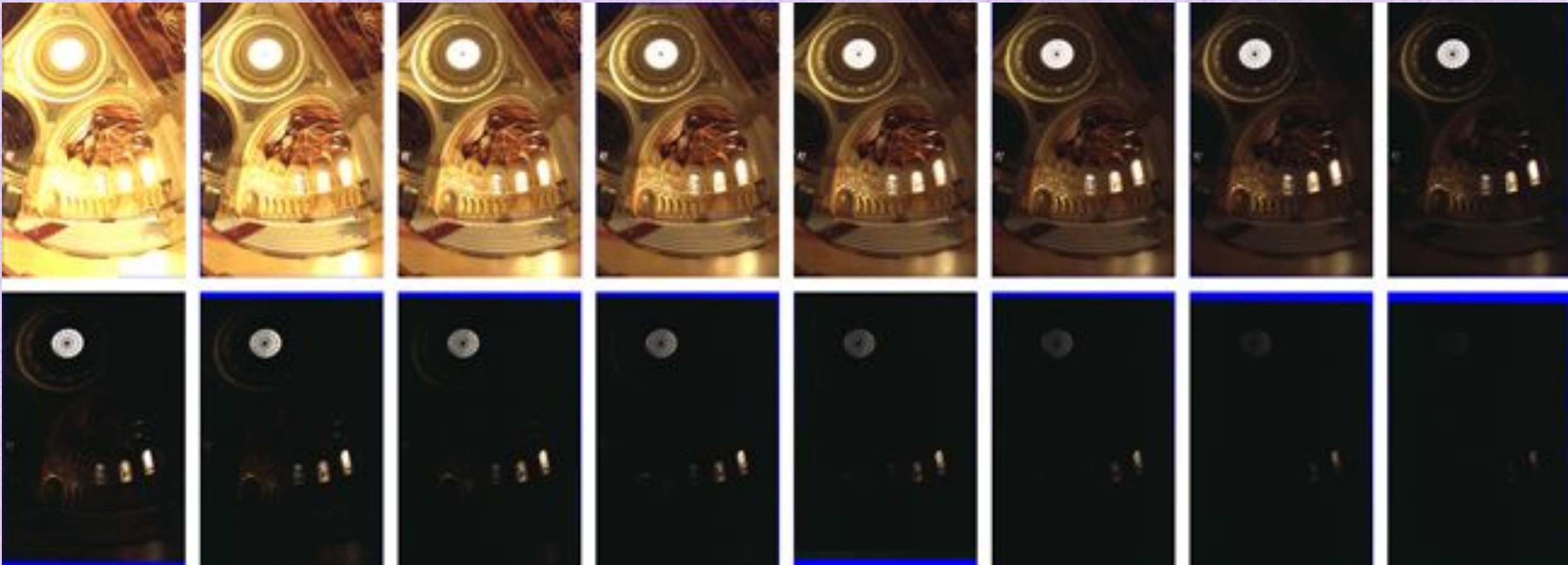
The World has High Dynamic Range



The relative irradiance values of the marked pixels

The World has High Dynamic Range

- 16 photographs of the Stanford Memorial Church taken at 1-stop increments from 30s to 1/1000s
- No single image captures everything desirable in both the darkest and the brightest regions (some pixels are over-saturated and others have no signal at all)



Tone Mapping

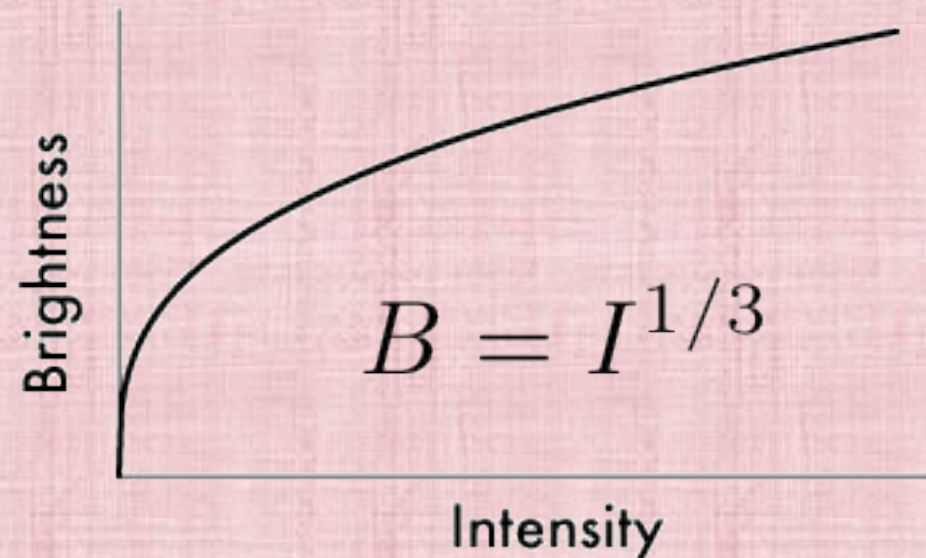
- “Compositing” all the information from all the images gives a result with a High Dynamic Range (i.e., 0-X with $X \gg 255$)
- But that range is too large for the standard image format (i.e., since $X > 255$)
- Solution #1: **Linearly rescale/compress** the values so that $X=255$
 - Small intensity differences are quantized (a range of values map to the same integer), and relative differences (and details) are lost
- Solution #2: **Logarithmic** map to rescale/compress
 - Information is still quantized, but in a more forgiving way exploiting human “perceptual space”
- Solution #3: Other approaches...
 - E.g., Local operators - map each pixel value based on surrounding pixel values (human vision is sensitive to *local* contrast)

Human Perception of Intensities

- We perceive brightness intensity differences better at lower (as opposed to higher) light intensities
- Logarithmic compression uses more resolution for the more-important lower intensities in the image (and thus less resolution for the less-important higher intensities)
- This gives less quantization in the lower intensities of the image (than in the higher intensities), and is thus more optimal for human consumption

$$S = I^p$$

Sense	Exponent
Brightness	0.33
Loudness	0.60
Length	1.00
Heaviness	1.45



Linear vs. Logarithmic Compression



Linear



Logarithmic

Gamma Encoding and Correction

- Maximize the use of the information relative to human perception
- More bits are allocated to the lower intensity (darker) regions of the image than to the higher intensity (lighter) regions
- Gamma correction is applied to the gamma encoded images to convert them back to the original brightness/luminance

