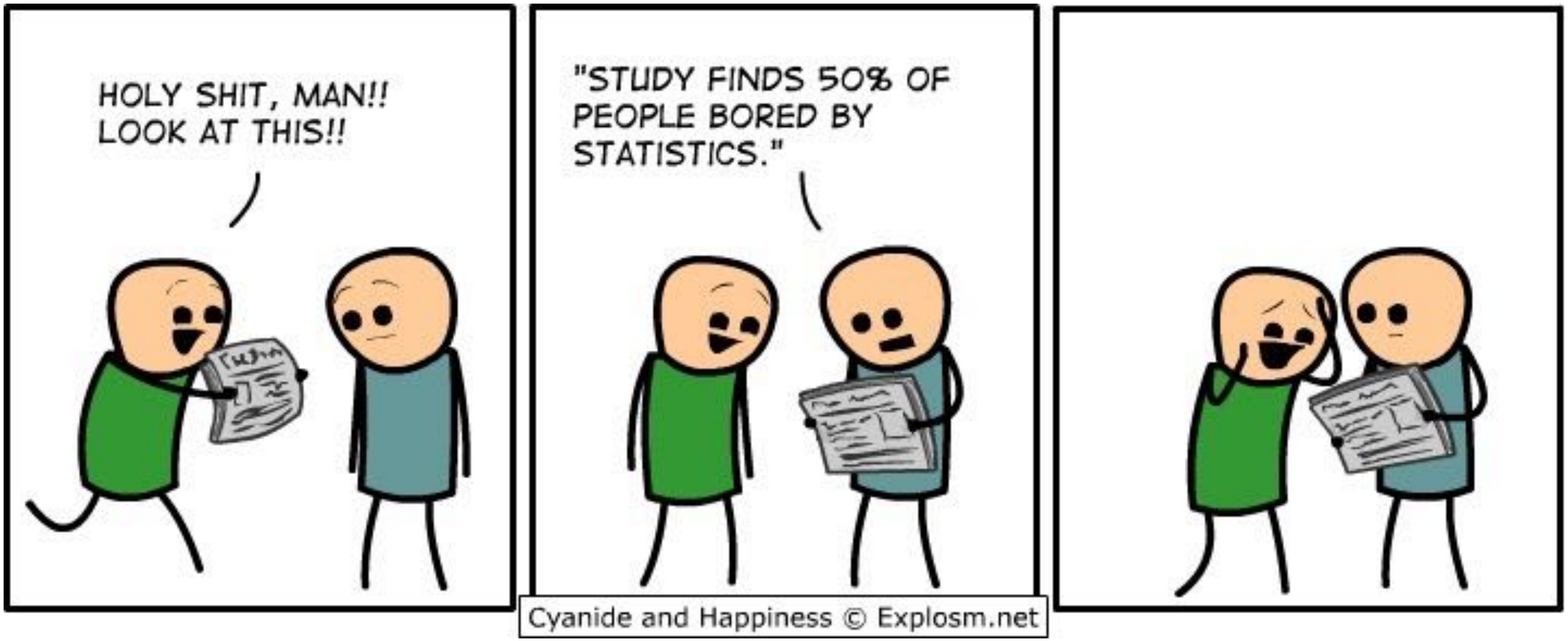
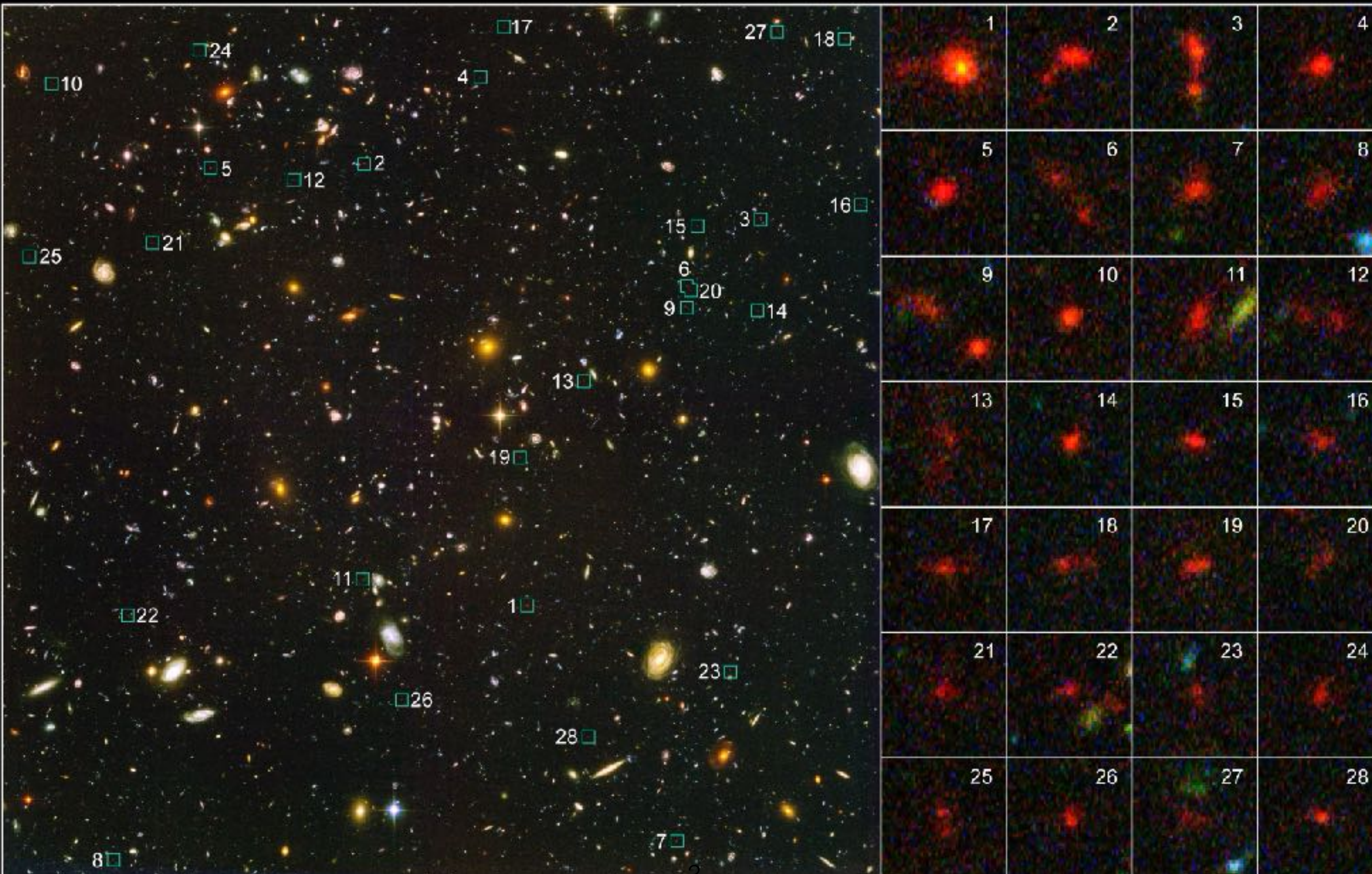
Astronomy 503
Observational Astronomy

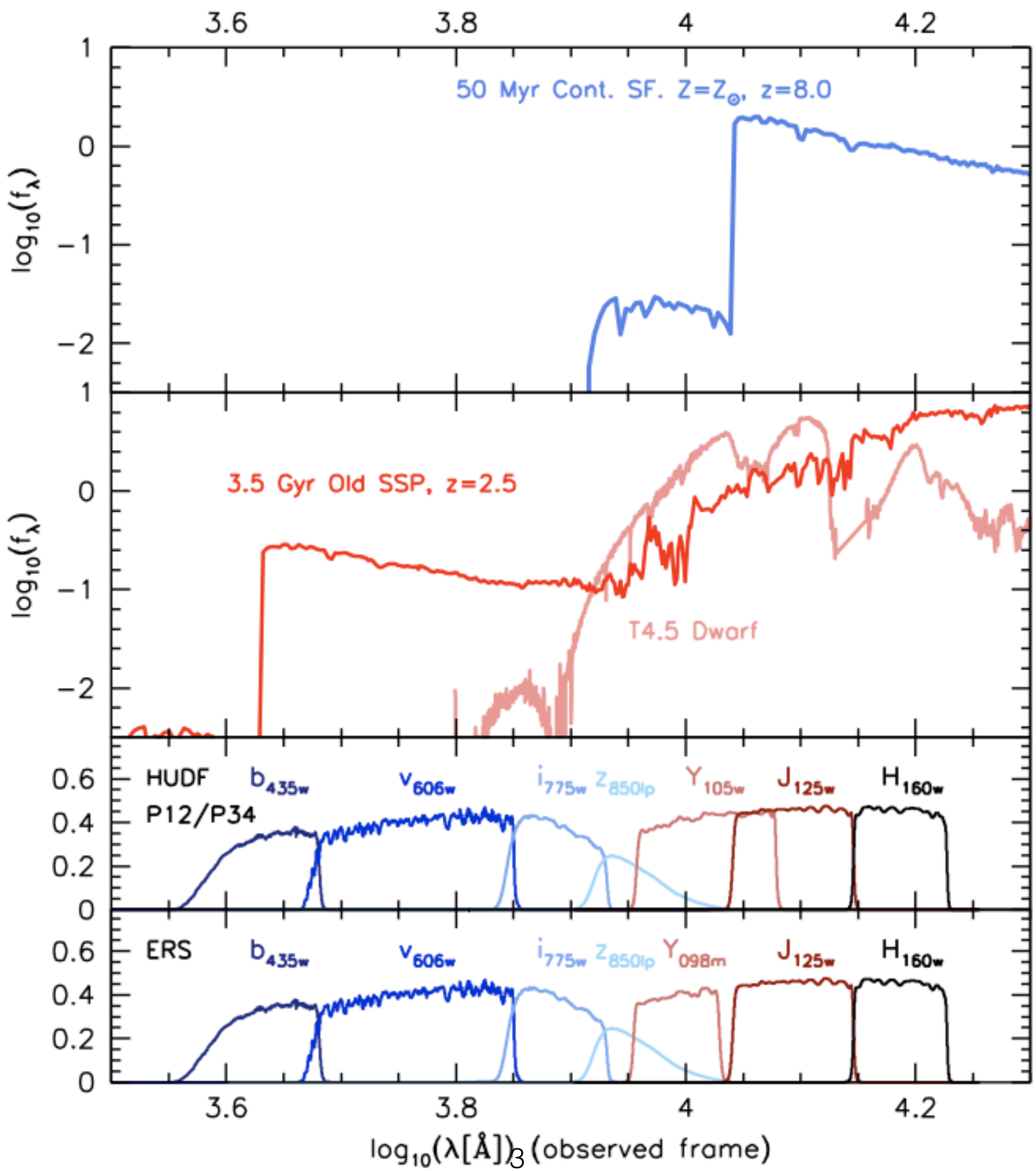


Prof. Gautham Narayan

Lecture 04: Detectors, Statistics & Uncertainty

z~7 galaxies in the Hubble Ultra Deep Field



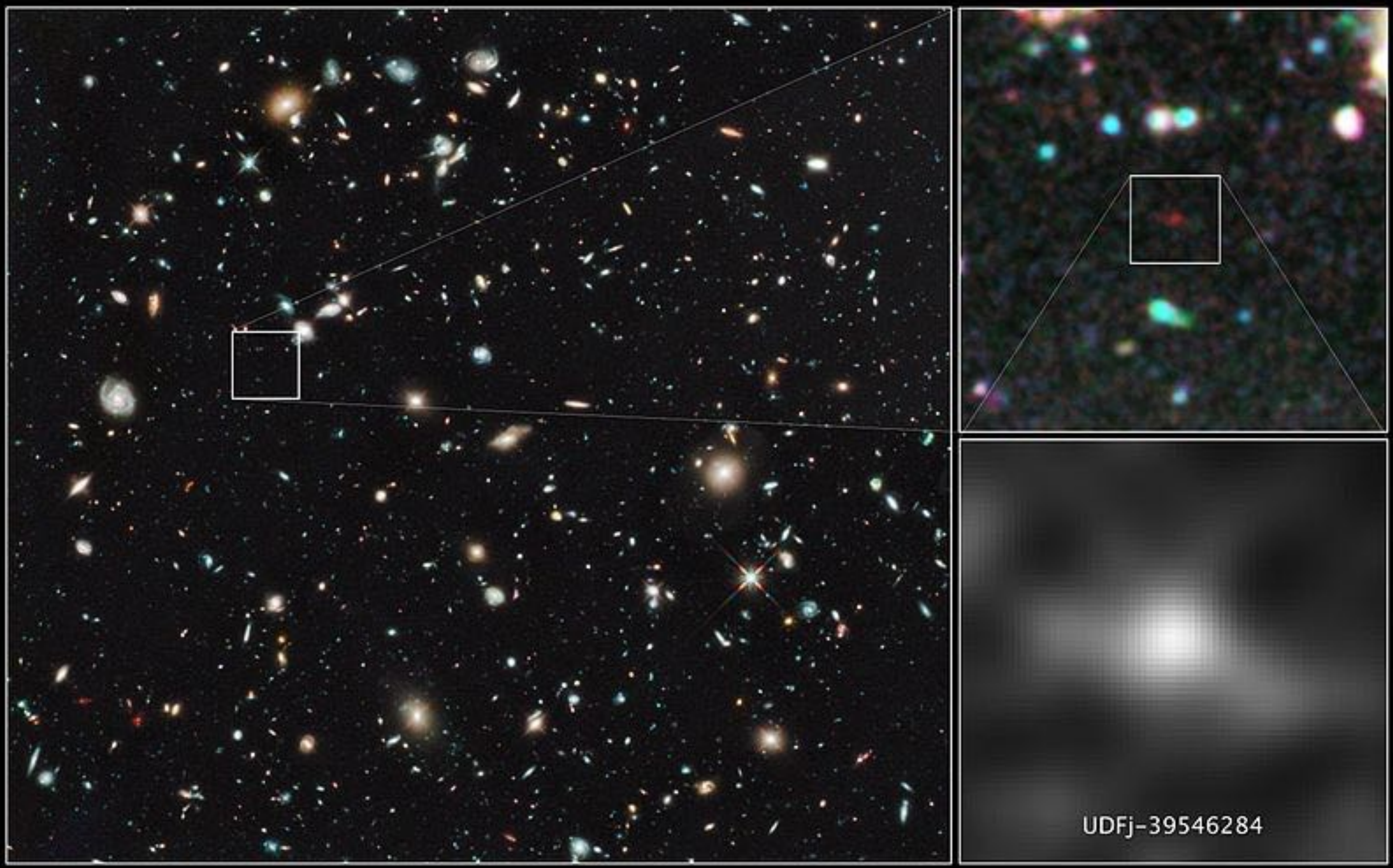


50 Myr Cont. SF. Z=Z$_\odot$, z=8.0
3.5 Gyr Old SSP, z=2.5
T4.5 Dwarf
HUDF
P12/P34
ERS
b_{435w}
v_{606w}
i_{775w}
z_{850lp}
Y_{105w}
Y_{098m}
J_{125w}
H_{160w}
$\log_{10}(f_\lambda)$
$\log_{10}(\lambda[\text{Å}])$ (observed frame)

UDFy-38135539

2010 z=8.6

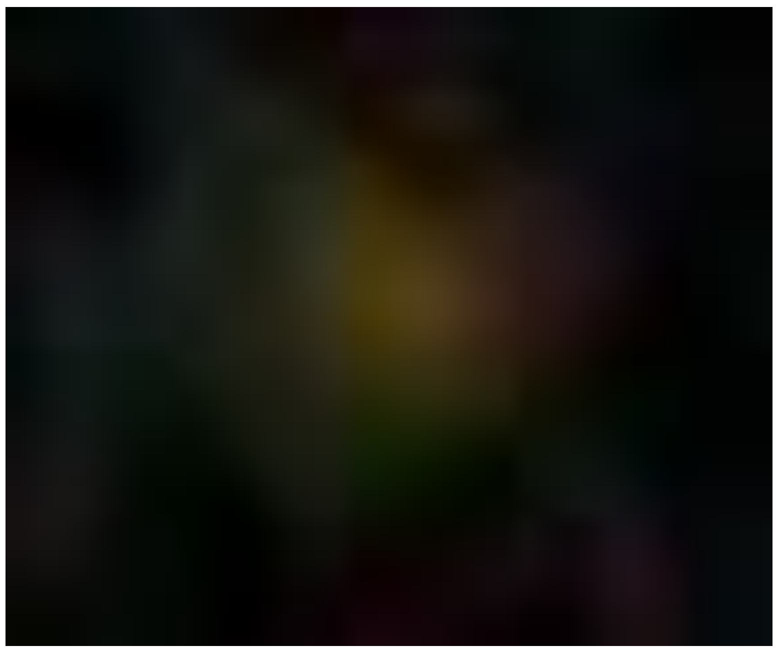
most distant spectroscopically confirmed object

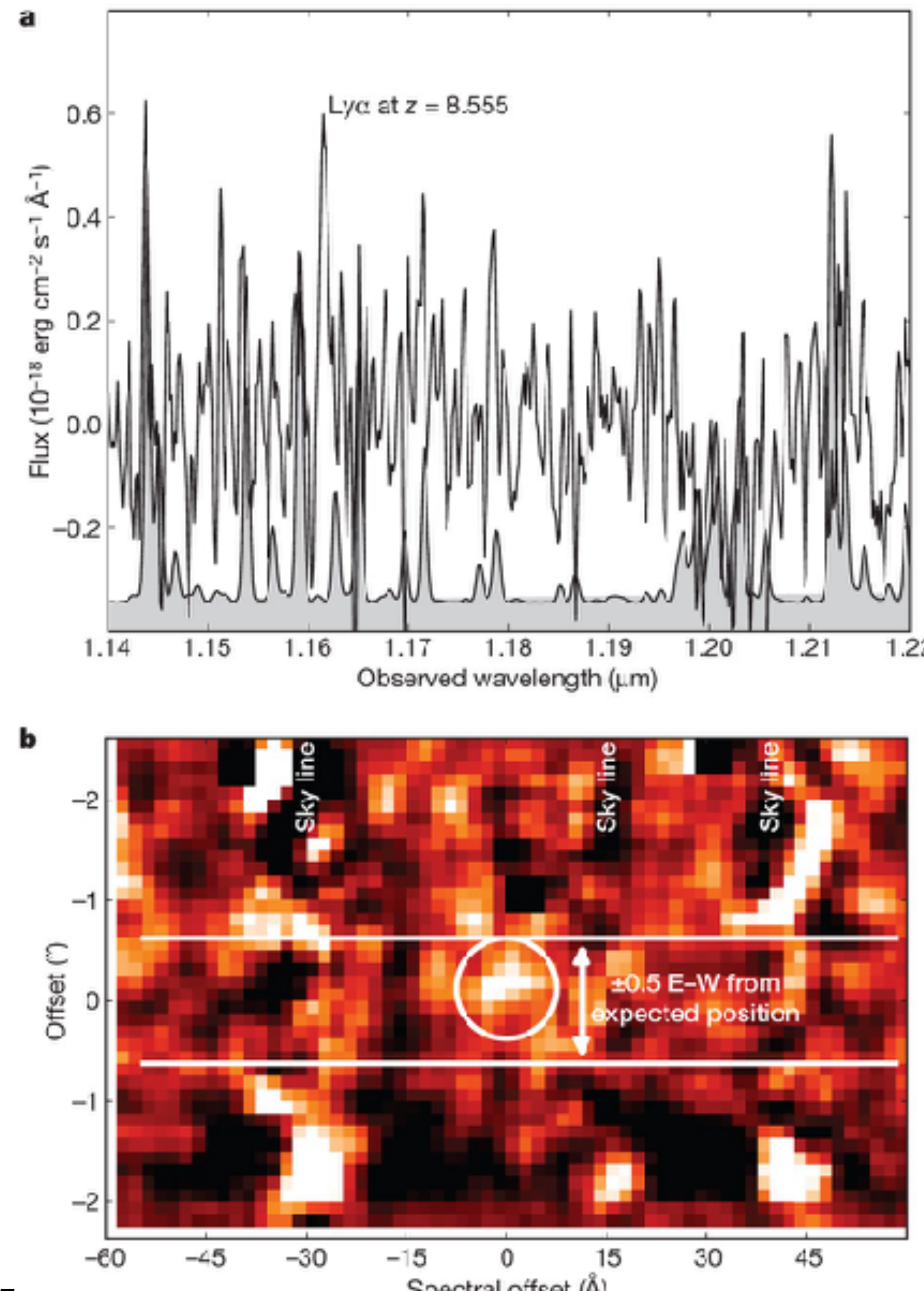


UDFy-38135539

2010 z=8.6

most distant spectroscopically confirmed object



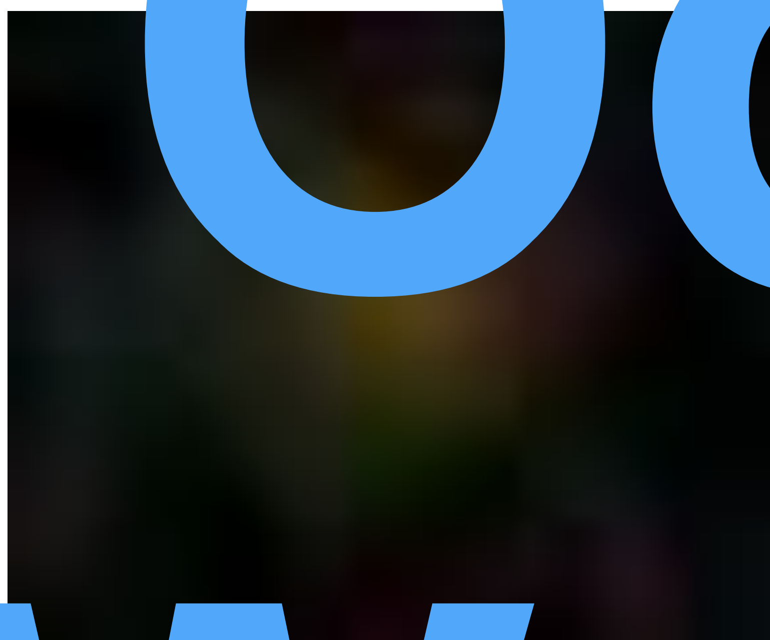


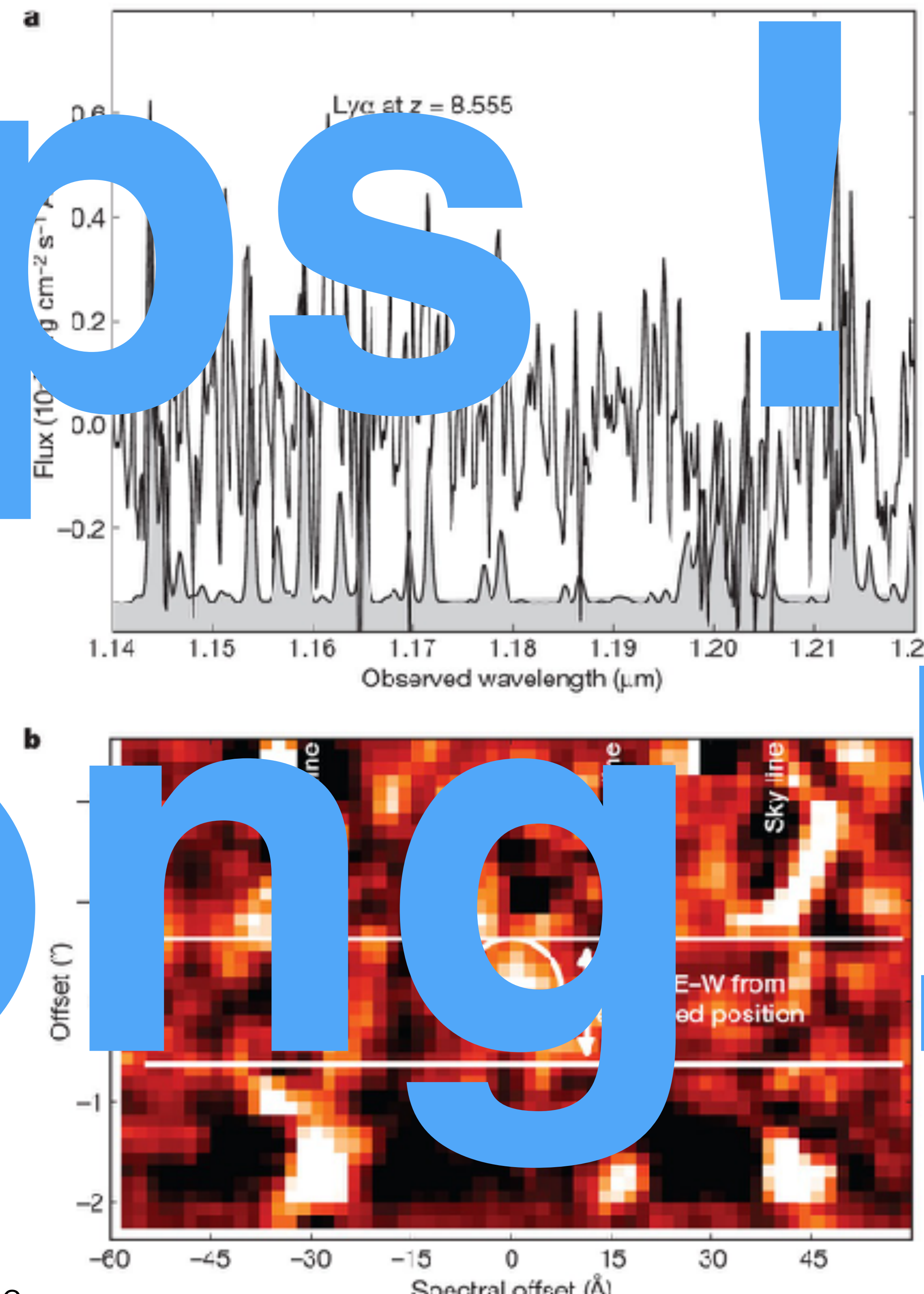
UDFy-38135539

2010 z=8.6

most distant spectroscopically confirmed object

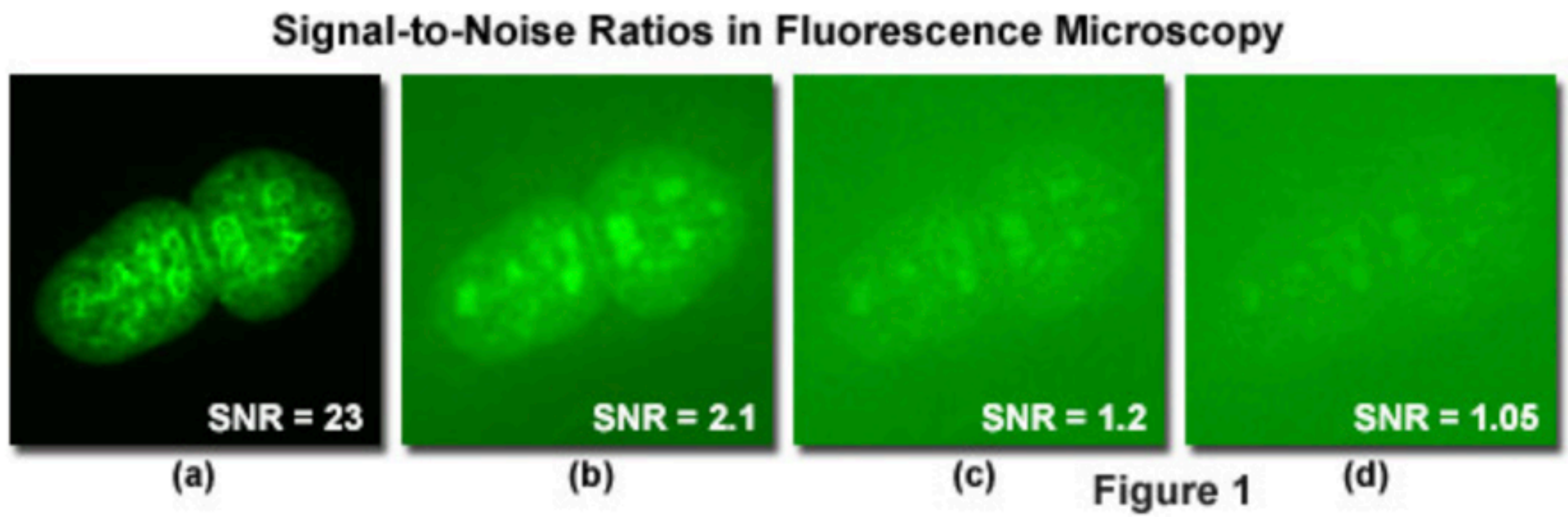
Oops !





Wrong!

How strong a signal is impacts much more than Astro



Fundamentally then, observers need to understand detectors and what their noise sources are

Detectors and Significance

WHAT YOU WANT FROM A DETECTOR

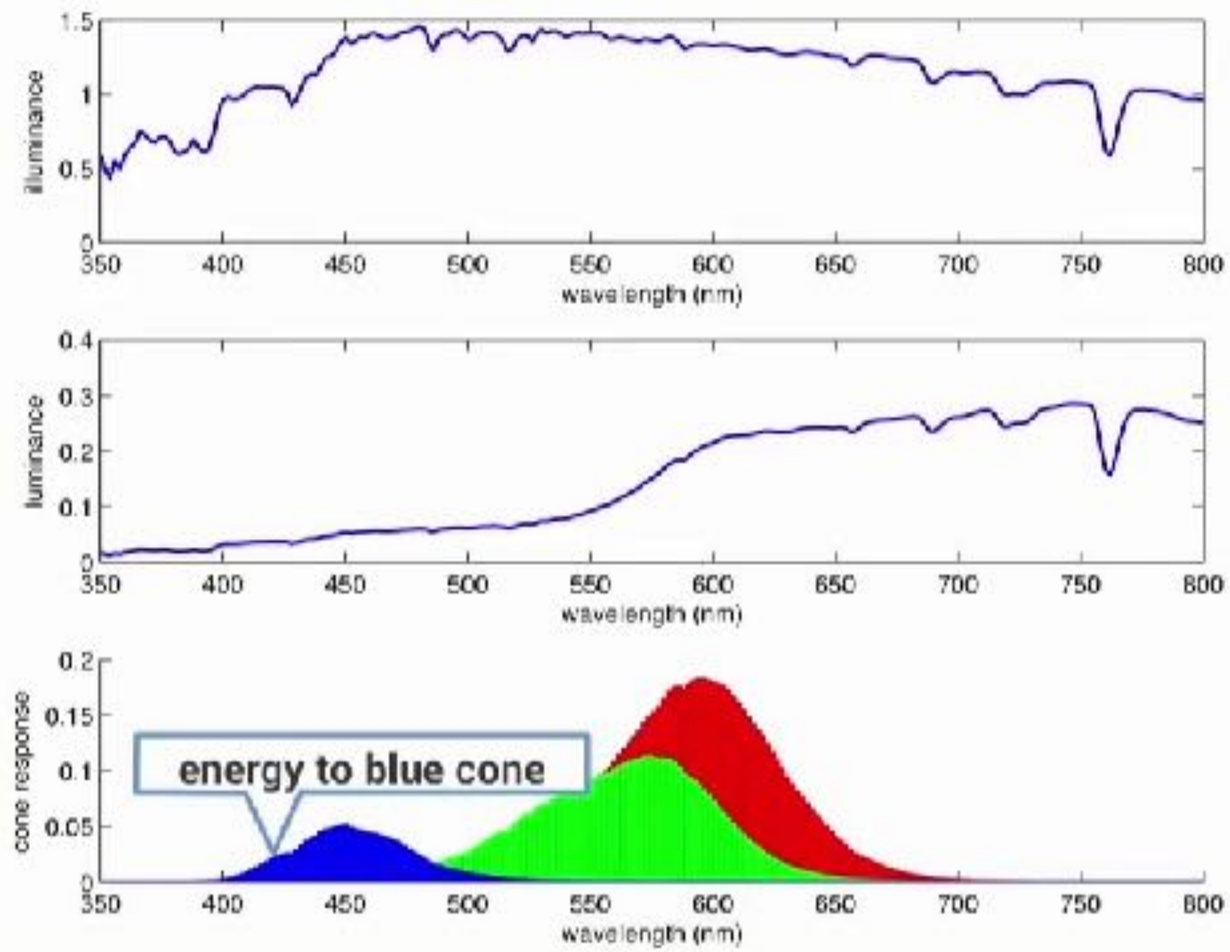
- Detectors often define the science possible. They need to:
  - Detect as many incident photons as possible (their quantum efficiency).
  - Translate the detected photons efficiently to detector output.
  - Be matched to our wavelength or energy region of interest.
  - ideally be configurable and adjustable – so that they can be used in optimized designs.
  - Generate as little noise as possible (i.e. detector response in absence of astronomical source).
  - Be stable, and able to be calibrated.

DETECTOR ALTERNATIVES #1

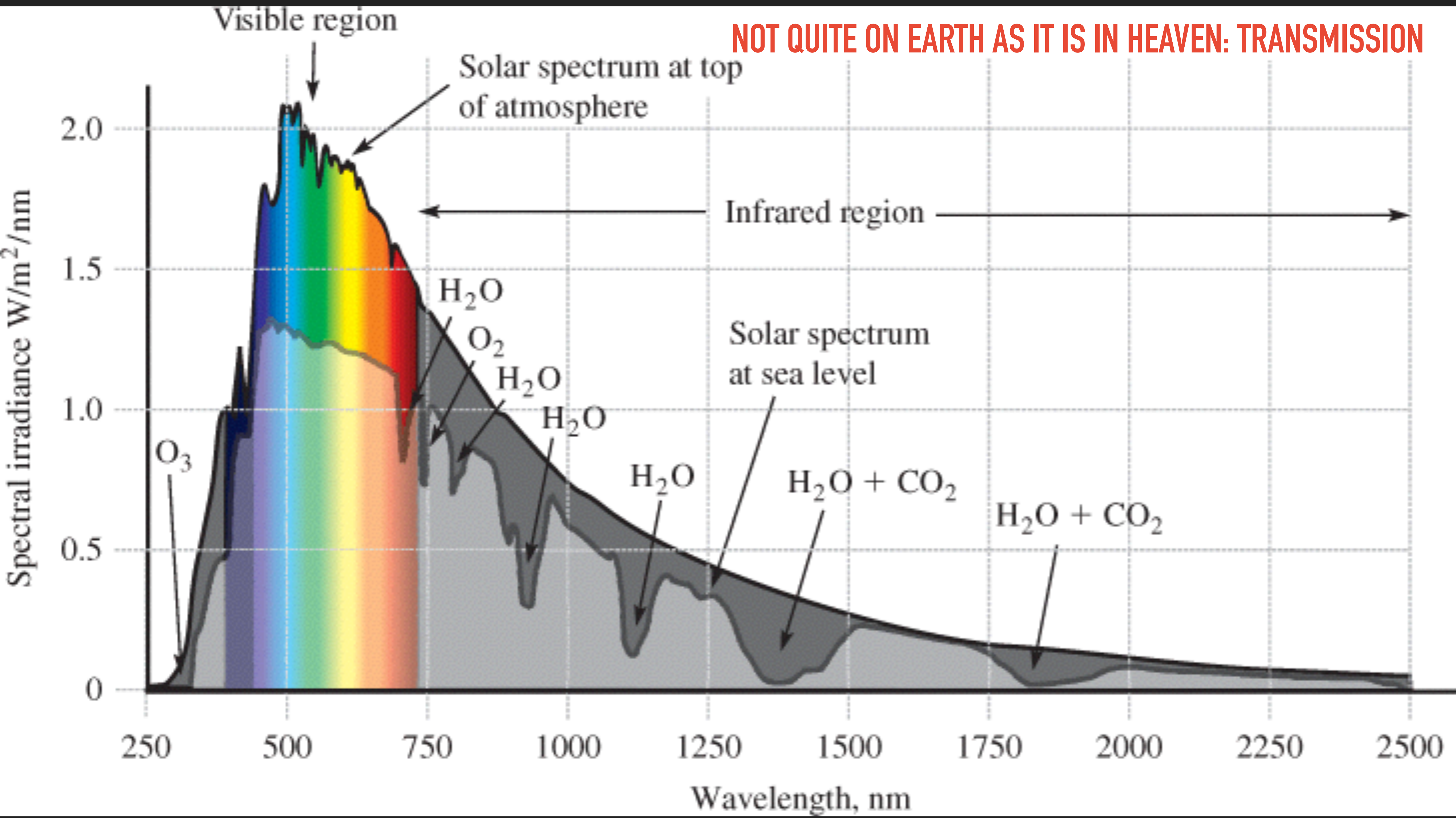
- Once light collected, focused, need to
  - detect photons
  - determine λ/colors
- Historical choice: naked eye – just look!
  - readily available, and cheap!
  - problems:
  - only ~ 1% of photons detected!
  - can't store image
  - only sensitive to small portion of EM spectrum (visible λs)

Color imaging spectra

$$\int_\lambda E(\lambda)R(\lambda)M(\lambda)d\lambda$$



NOT QUITE ON EARTH AS IT IS IN HEAVEN: TRANSMISSION



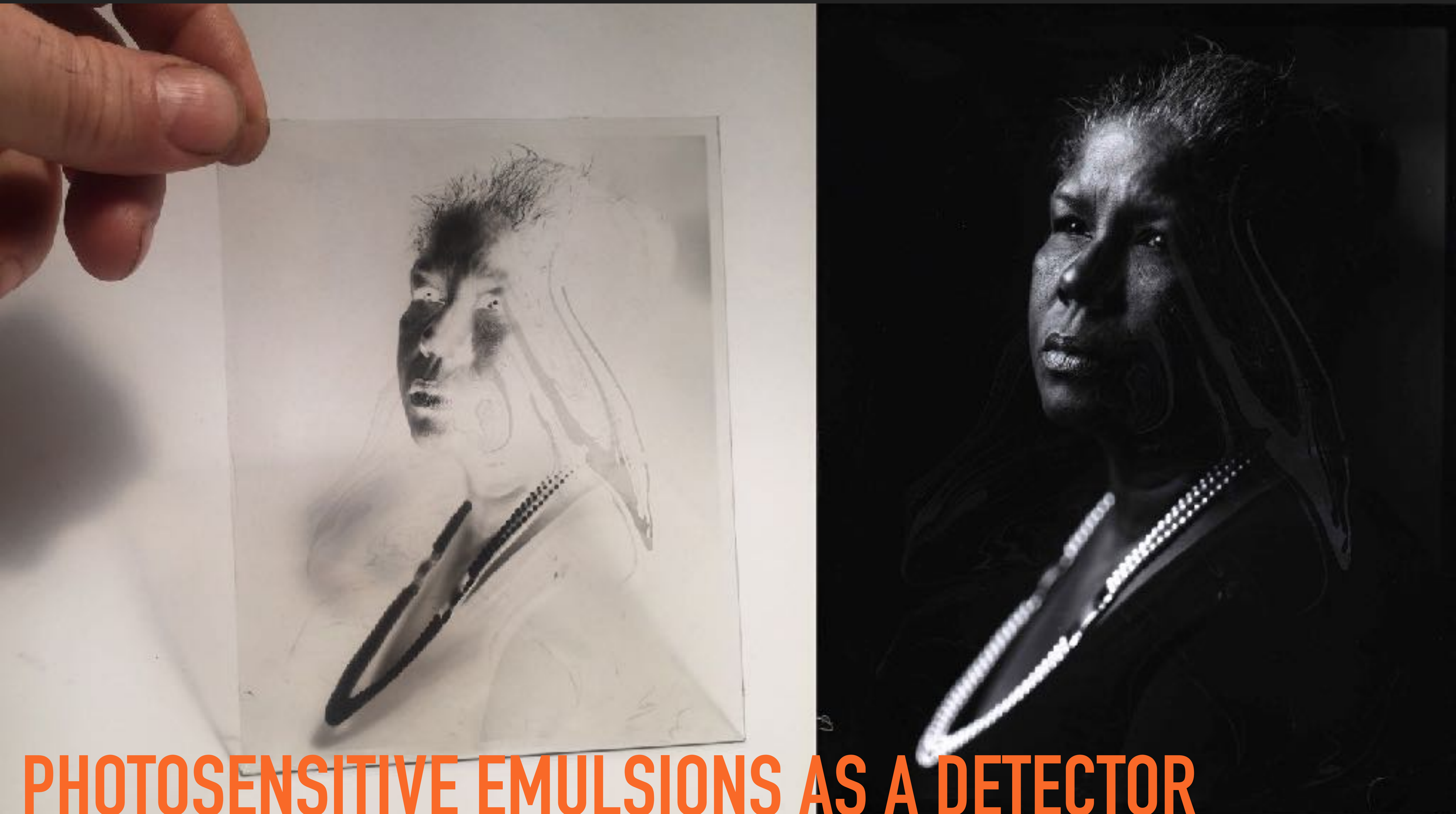
Visible • WFPC2

Infrared • NICMOS

Trapezium Cluster • Orion Nebula

WFPC2 • **Hubble Space Telescope** • NICMOS

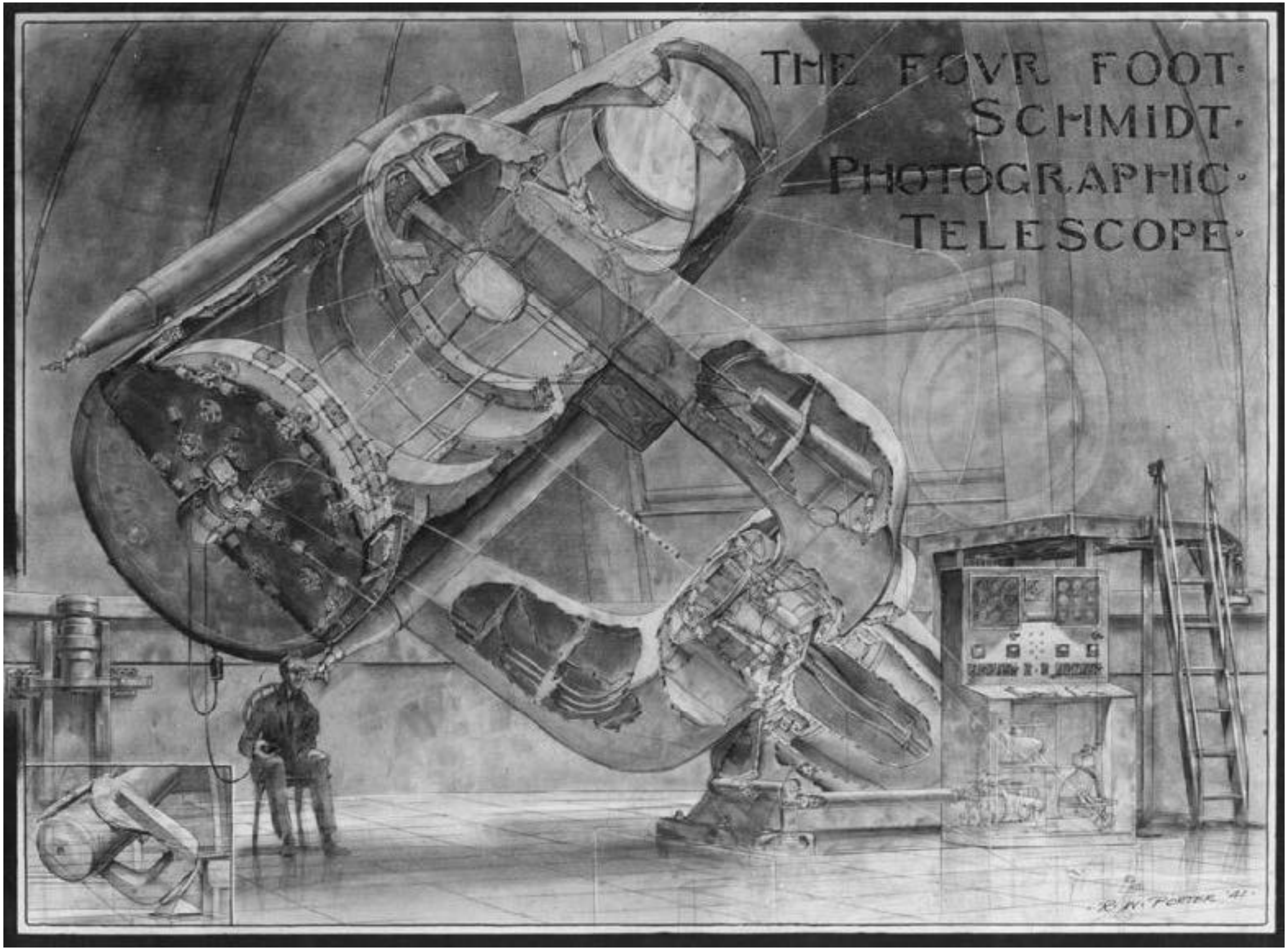
NASA and K. Luhman (Harvard-Smithsonian Center for Astrophysics) • STScI-PRC00-19



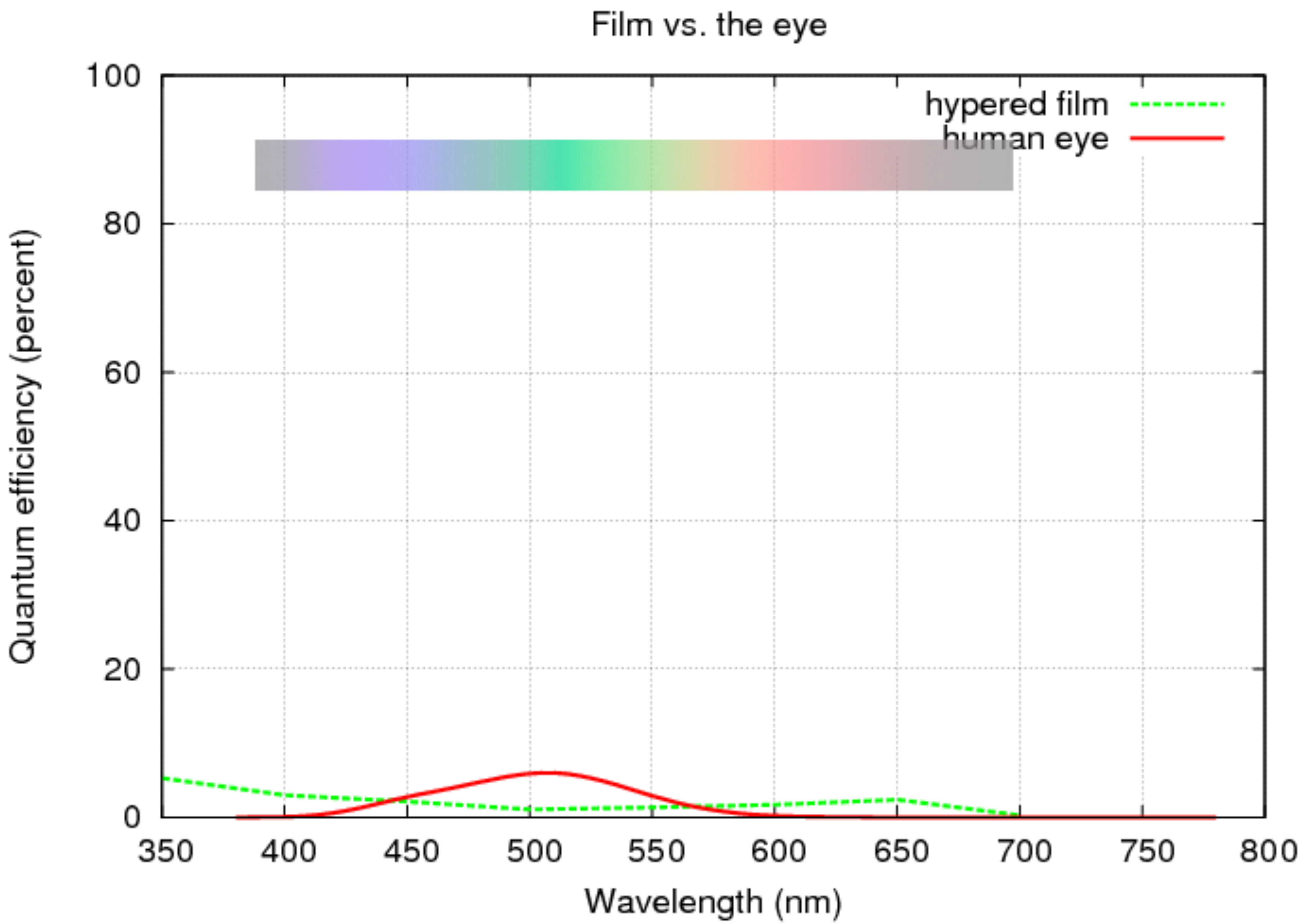
PHOTOSENSITIVE EMULSIONS AS A DETECTOR

DETECTOR ALTERNATIVE #2: PHOTOSENSITIVE FILM

- Daguerrotype (1839) - polished plate of silver coated copper treated with mercury is photosensitive - turns white as exposed to light
    - Very insensitive - Henry Draper's picture of the moon (right) took 20 minutes!
    - Not very uniform, hard to make - not efficient
- Coat a glass plate in a photosensitive emulsion (e.g. silver halide) suspended in gelatin (1850s) - almost transparent when not exposed to light
    - On exposure to light, halide (e.g. bromine) loses an electron (ionization!) and silver ion is converted to metallic silver - blocks light (i.e. appears dark)
- Refind by Kodak/Ilford etc from 1890s-1920s to use faster chemical reactions and a flexible base - **photographic film**

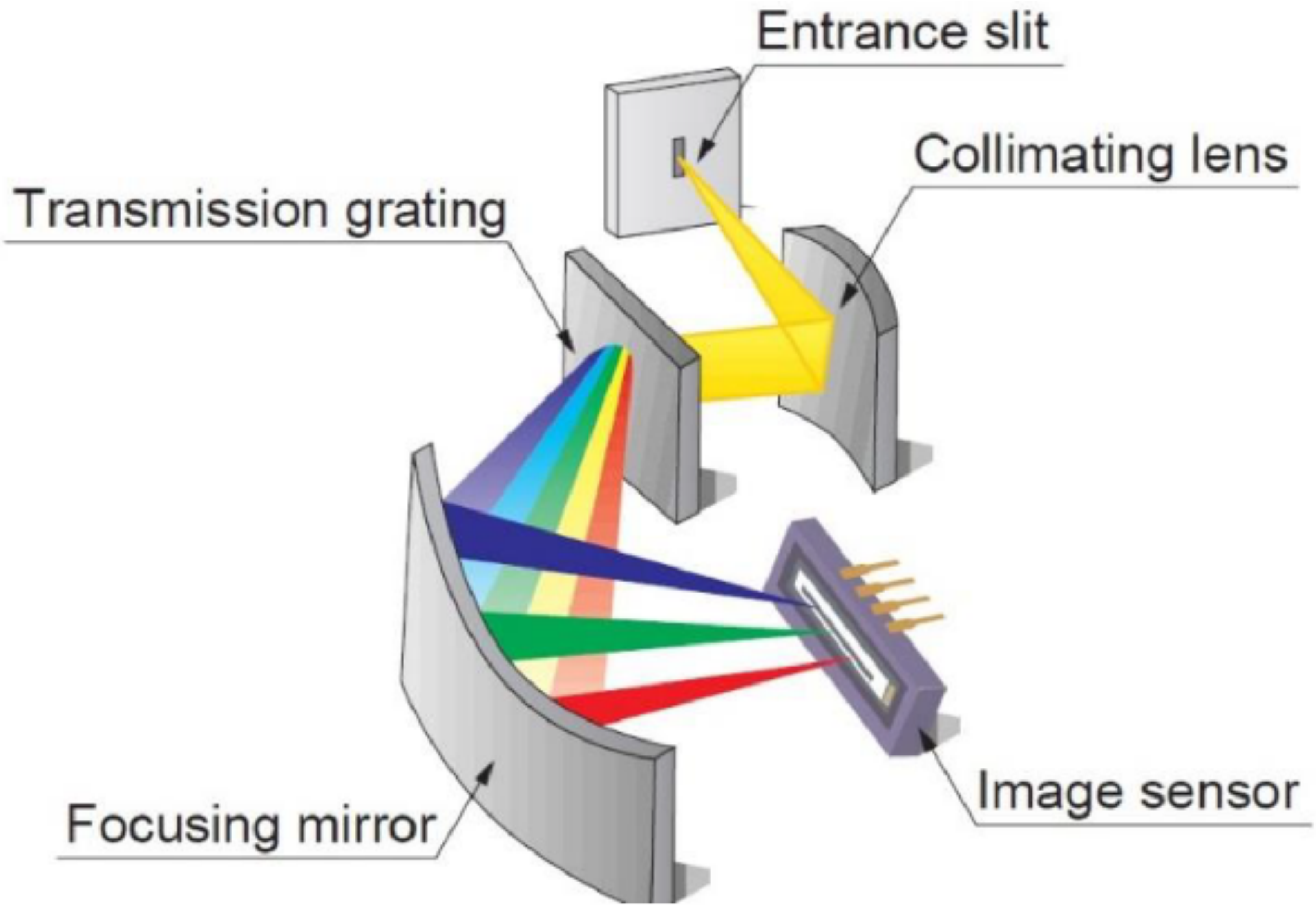

THE FOVR FOOT
SCHMIDT
PHOTOGRAPHIC
TELESCOPE
R. W. PORTER '41


VAR!
N
6-Oct
1923
N



Film vs. the eye
hypered film
human eye
Quantum efficiency (percent)
Wavelength (nm)
0
20
40
60
80
100
350
400
450
500
550
600
650
700
750
800

HOW DO YOU GET COLOR OUT OF BLACK AND WHITE?



(McLean 2008)

- ***Disperser:***
  - Angular dispersion of light into constituent wavelengths.
  - Dispersive elements: ***prisms***, ***gratings***, or ***grisms***.
  - Some may operate in either *reflection* or *transmission* mode.

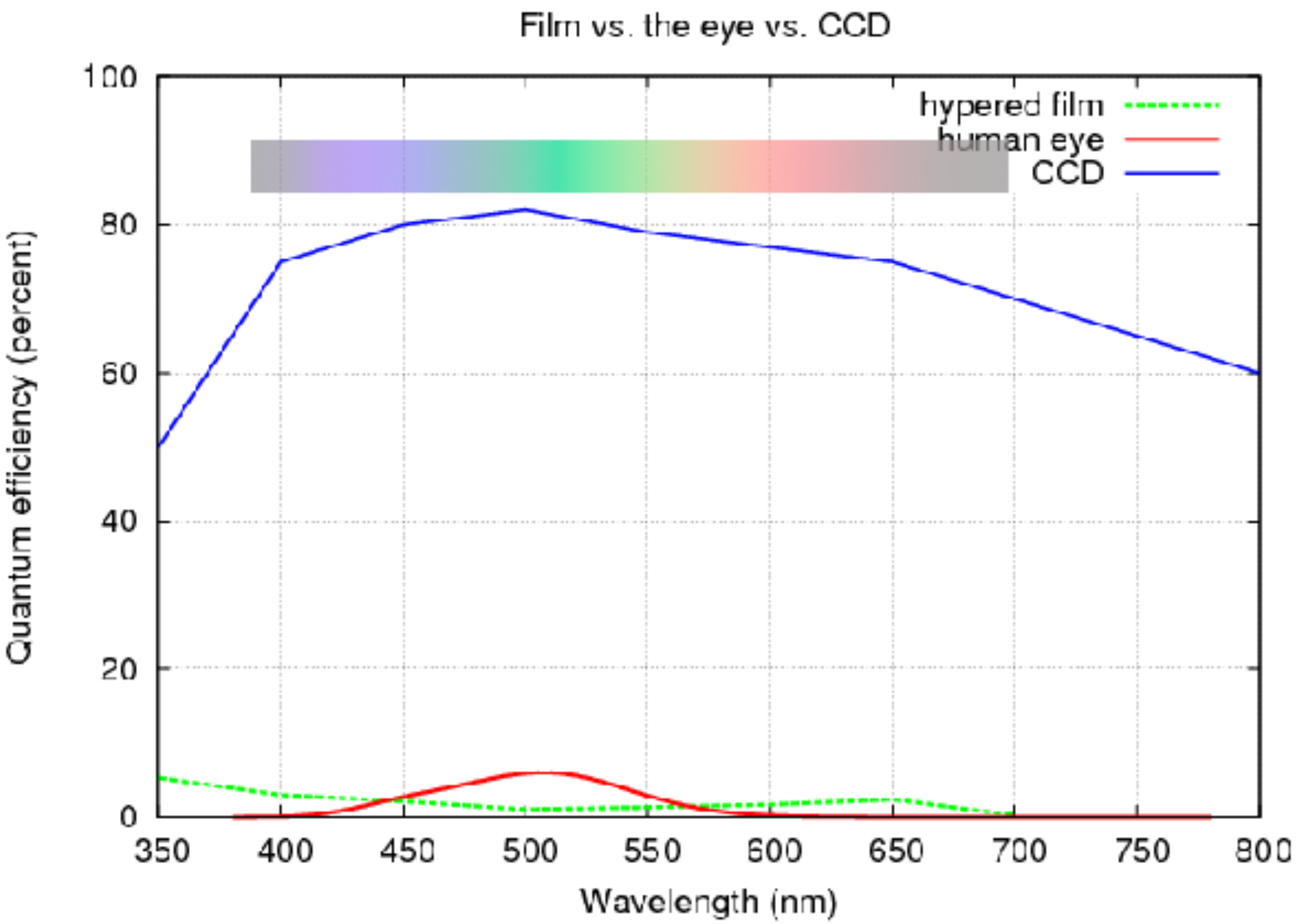
DETECTOR ALTERNATIVE #3: CHARGED-COUPLED DEVICES (CCD)

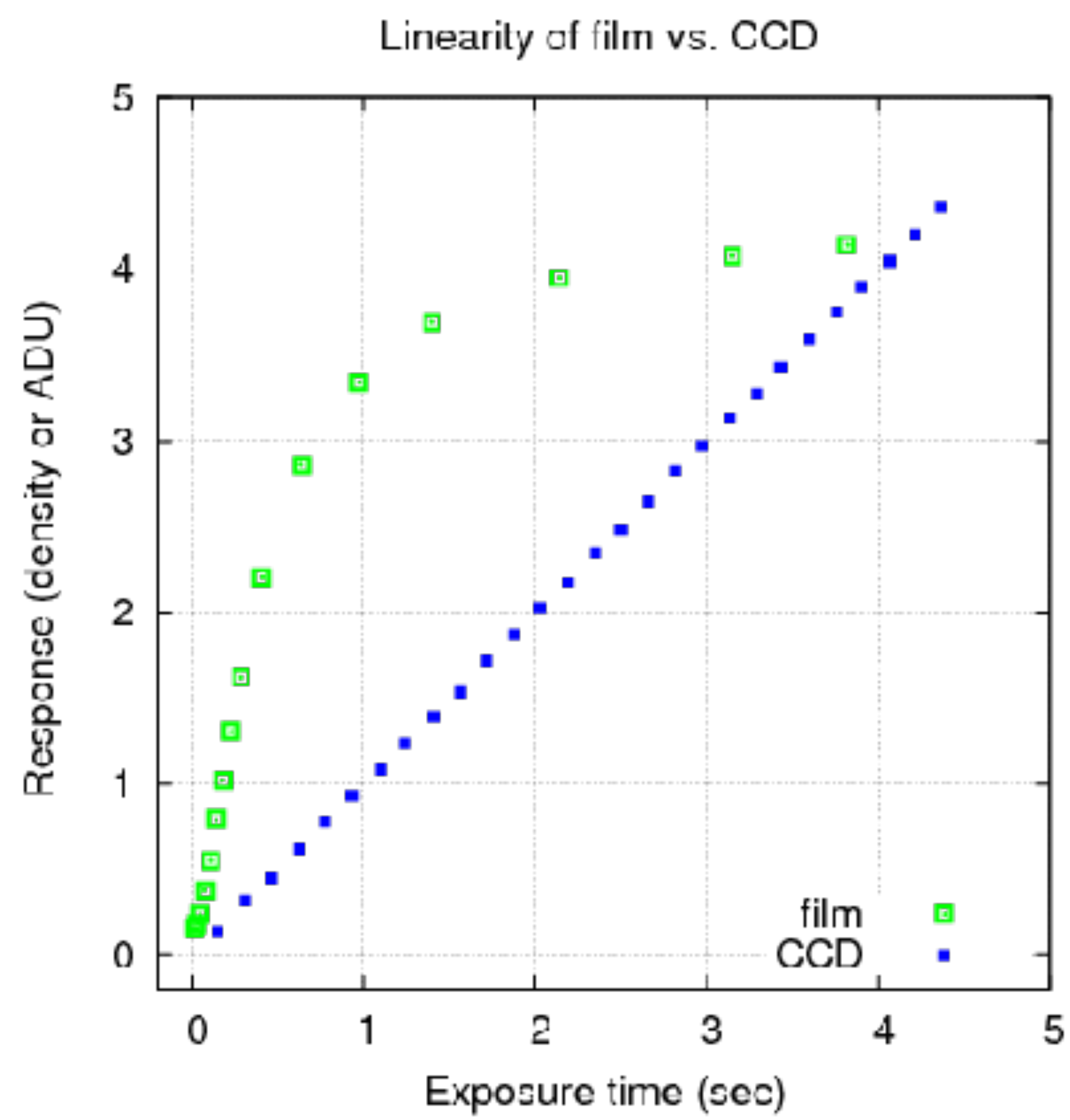
- silicon wafer = semiconductor with a grid/array of capacitors
- serve as picture elements i.e. **pixels**
  - photons → silicon → e knocked out **(photoelectric effect)**
  - charge accumulates in capacitors (e– potential well = “bucket”) for each pixel
  - **total charge is proportional to number of incident photons**
  - after exposure (“integration”) time, read out array
    - charge pattern → image! and flux measurements!

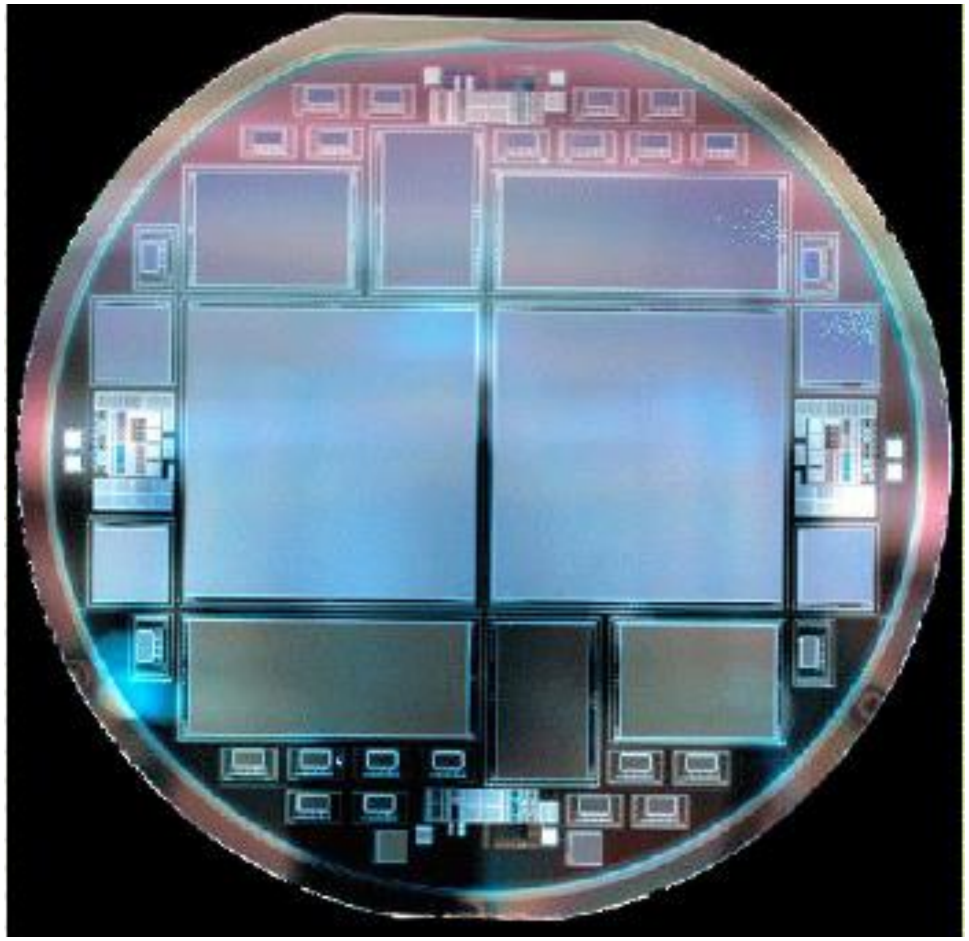
https://astro.unl.edu/classaction/animations/telescopes/buckets.html

- efficiency: > 80% of incident photons detected and linear: charge signal ∝ photon counts ∝ flux

- digital data is easy to work with + can make large detectors, but EXPENSIVE!







A huge wafer of silicon about 6 inches in diameter with multiple CCDs and other devices laid out on it. The large chips are 4k x 4k pixels

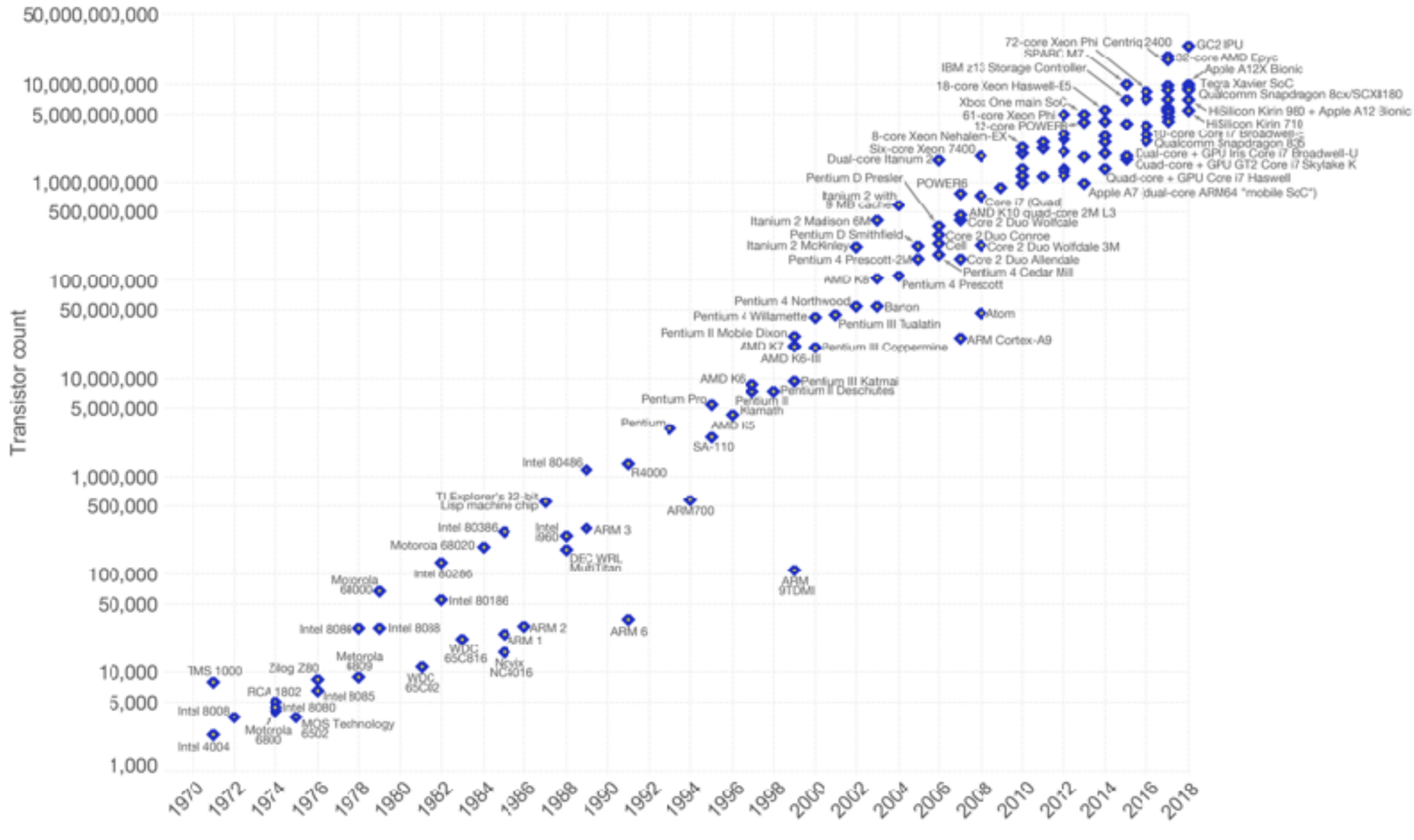
A ***mosaic*** of 40, 2048 x 4612 pixel CCDs yielding a camera with 378 megapixels on the CFHT this camera has a 1 degree field with 0.18” resolution

(McLean 2008)

Moore's Law – The number of transistors on integrated circuit chips (1971-2018)

Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years. This advancement is important as other aspects of technological progress – such as processing speed or the price of electronic products – are linked to Moore's law.

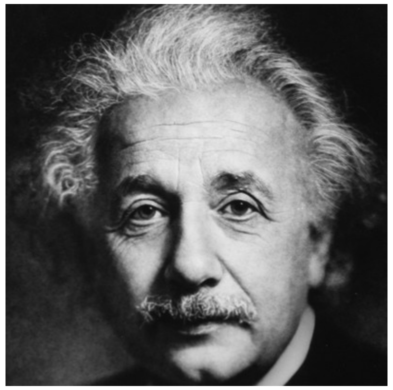
Our World in Data



Data source: Wikipedia (https://en.wikipedia.org/wiki/Transistor_count)
The data visualization is available at OurWorldinData.org. There you find more visualizations and research on this topic.

Photons

Einstein's explanation: light comes in discrete units called **photons**



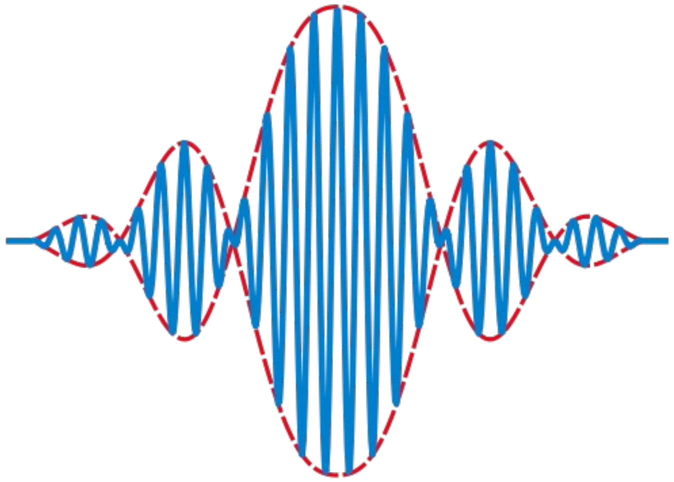
Albert Einstein
(1879 – 1955)

Energy of one photon:

$$E = h\nu$$

where **Planck's constant**

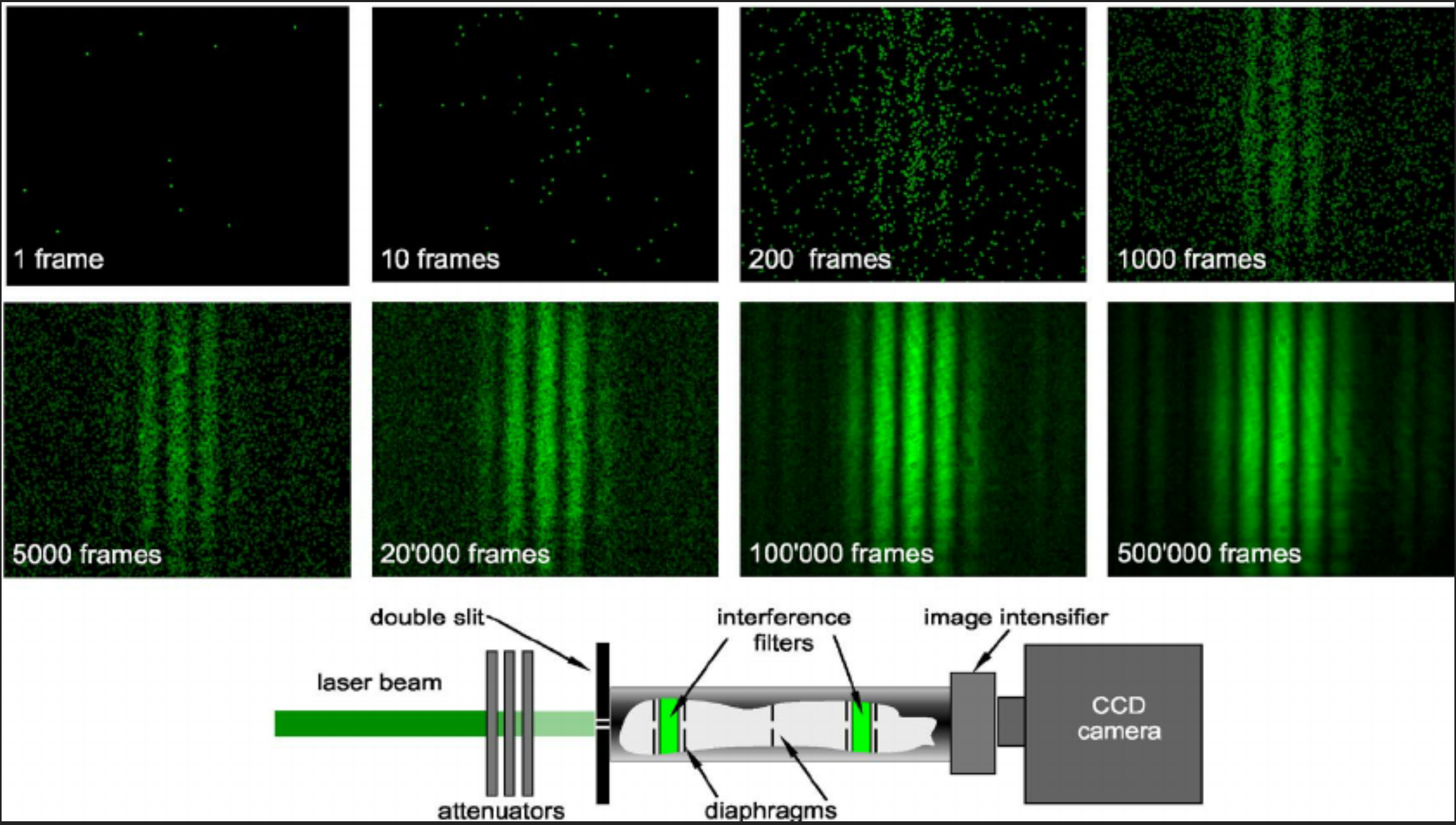
$$h \approx 6.626 \times 10^{-27} \text{ erg s} = 6.626 \times 10^{-34} \text{ J s}$$



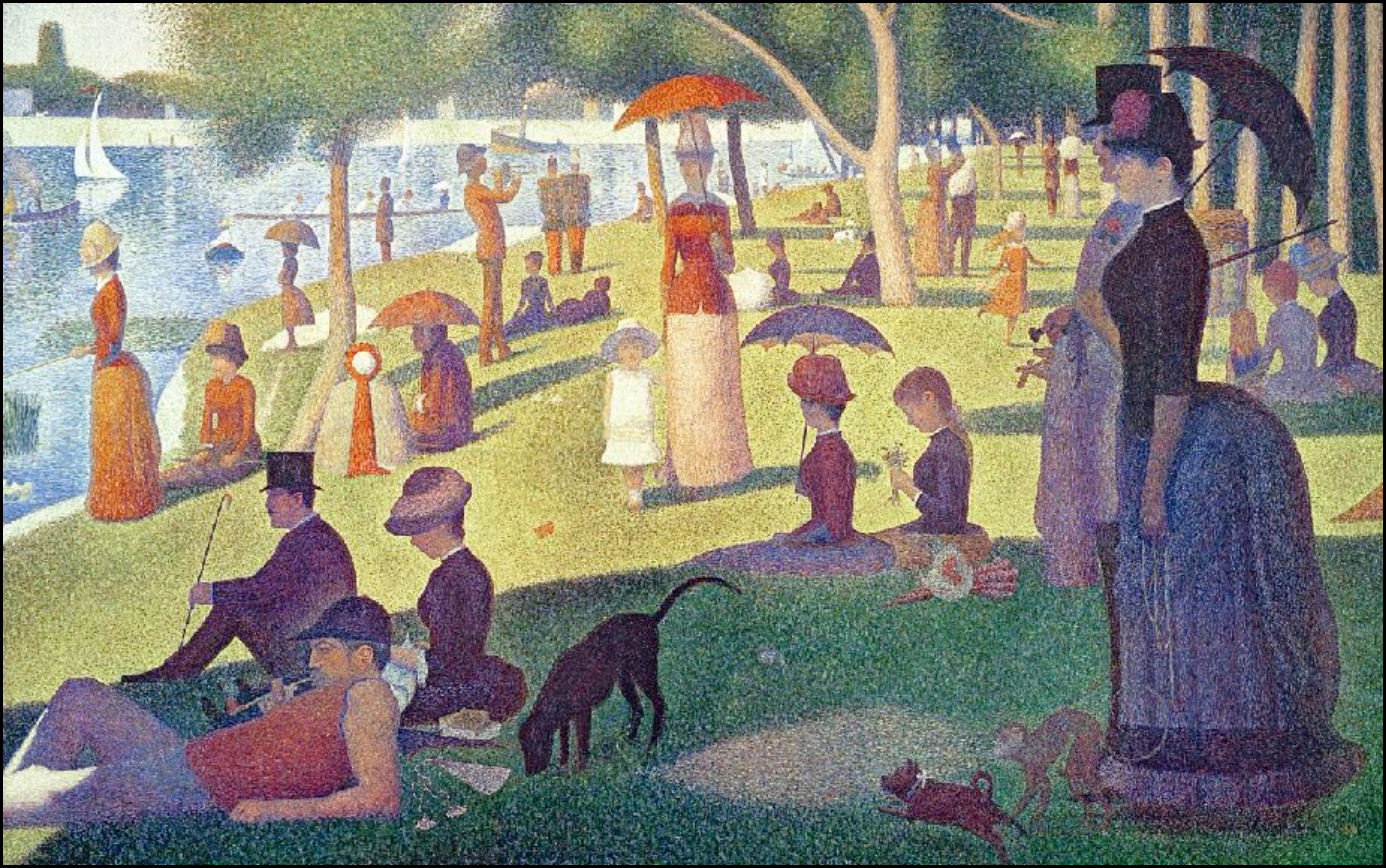
Wikimedia Commons

A photon is a localized electromagnetic wave

Behaves like a particle... though not like a “classical” particle

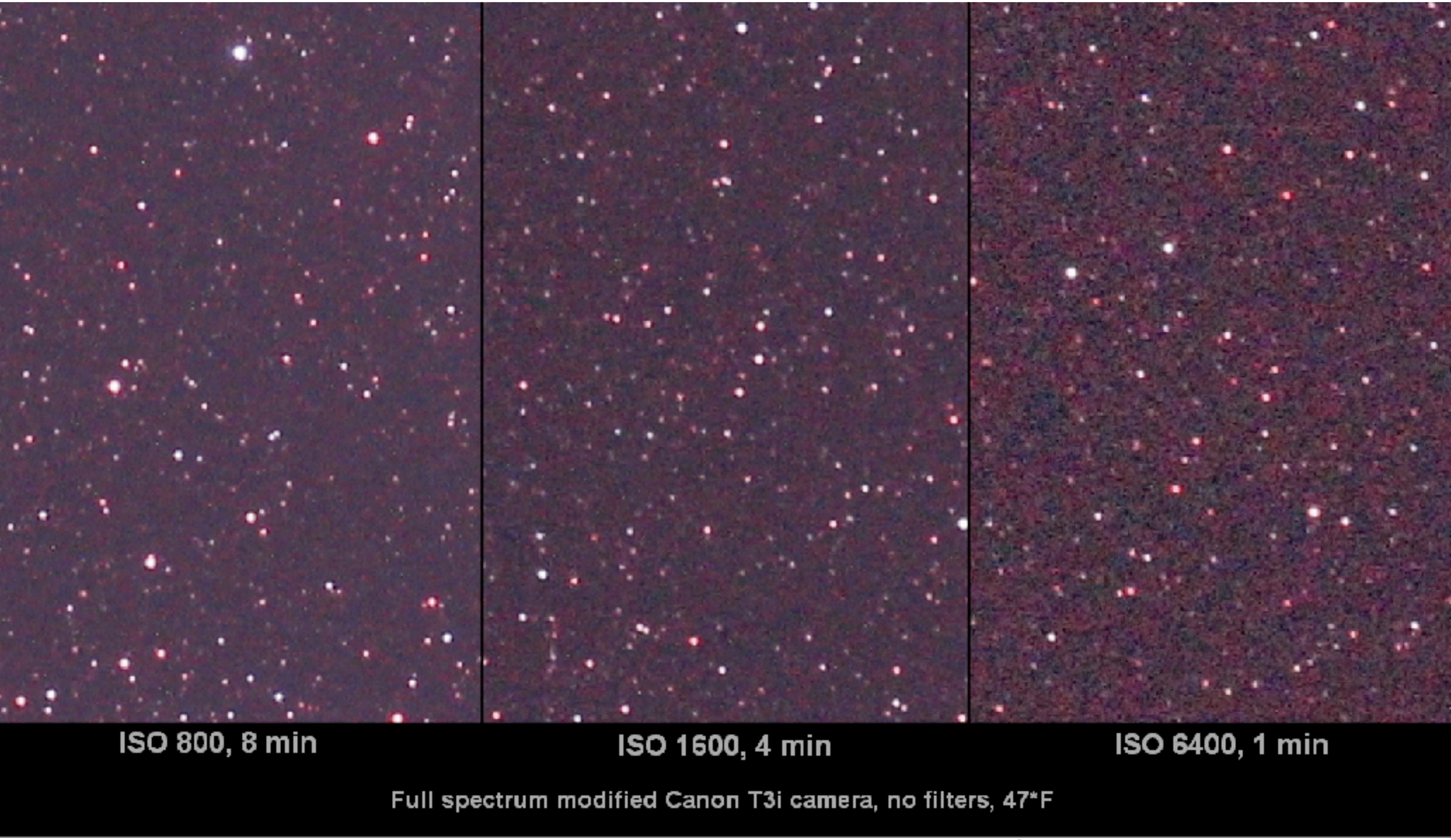

1 frame
10 frames
200 frames
1000 frames
5000 frames
20'000 frames
100'000 frames
500'000 frames
double slit
interference
filters
image intensifier
laser beam
CCD
camera
attenuators
diaphragms

IN FACT, IF YOUR EYES COULD BE COMMANDED TO INTEGRATE



“Signal” i.e. electrons are generated from more than just astrophysical sources

A single bias frame captured with a DSLR camera (ISO 800, 1/8000").



Fundamentally bias is the signal you get from electronics and is therefore impacted by voltages

Figure 4.3: Bias Gradients Seen Within Each Image Quadrant in the WFC Superbias (In Units of Counts) after SM4

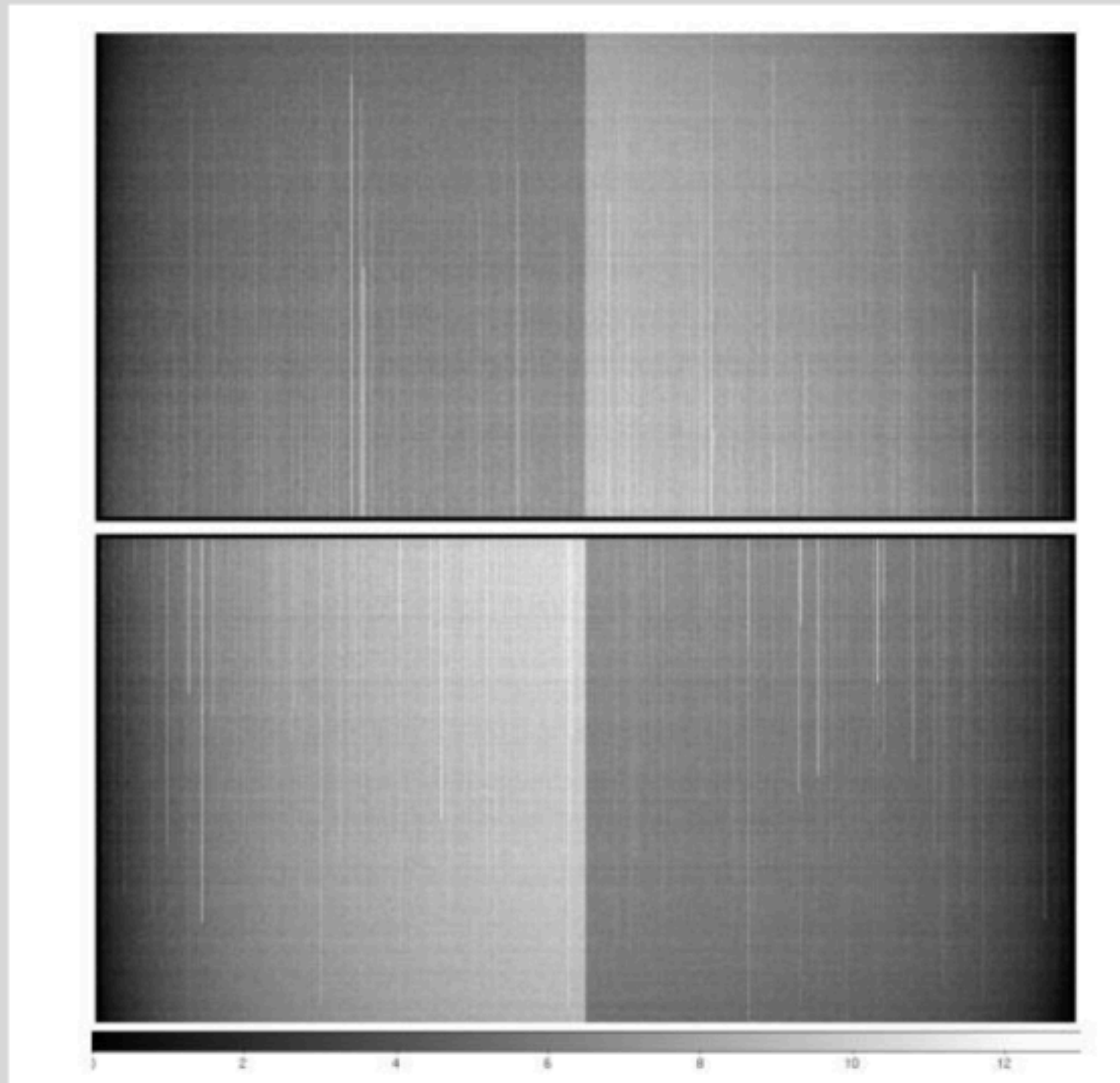
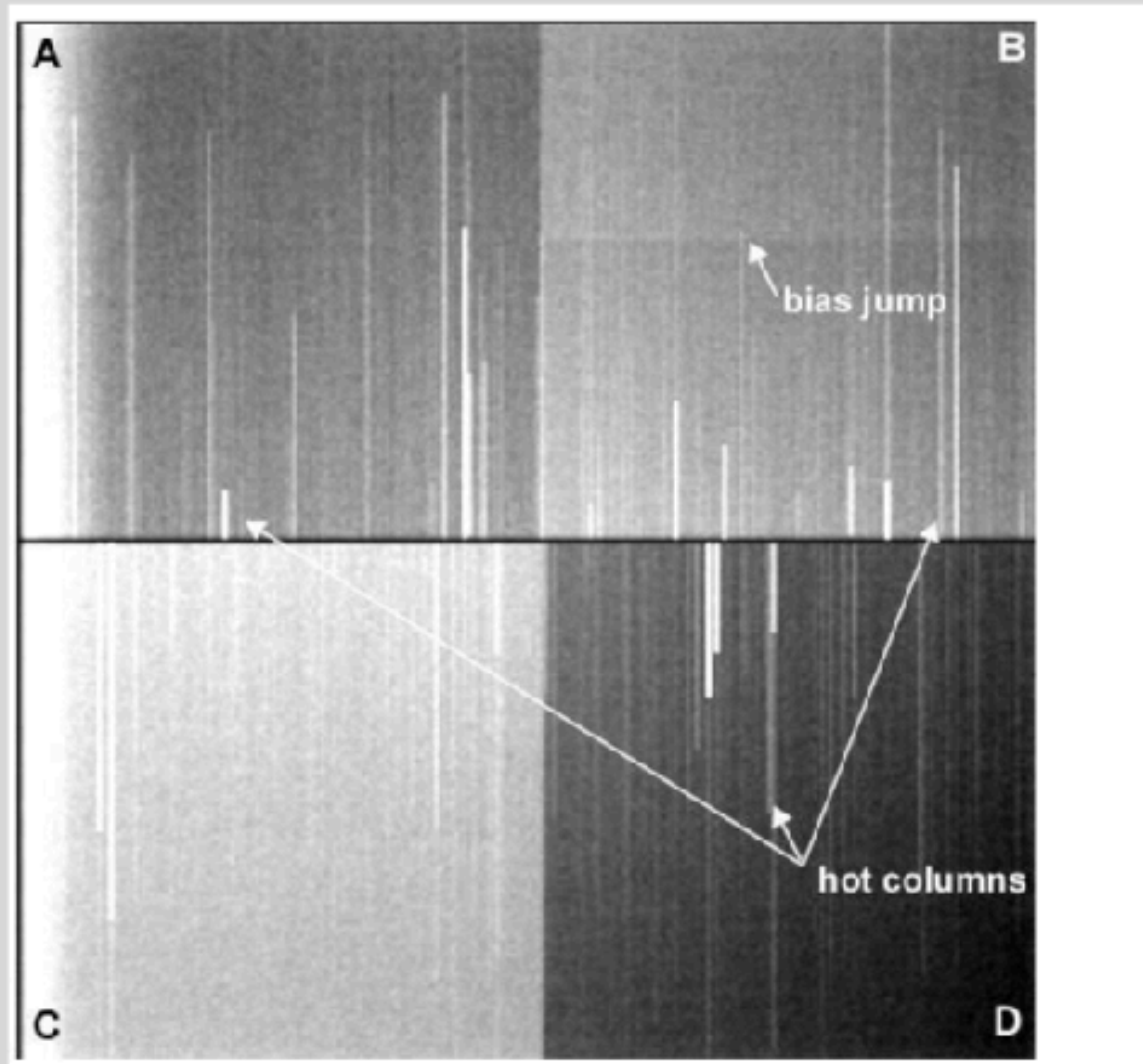
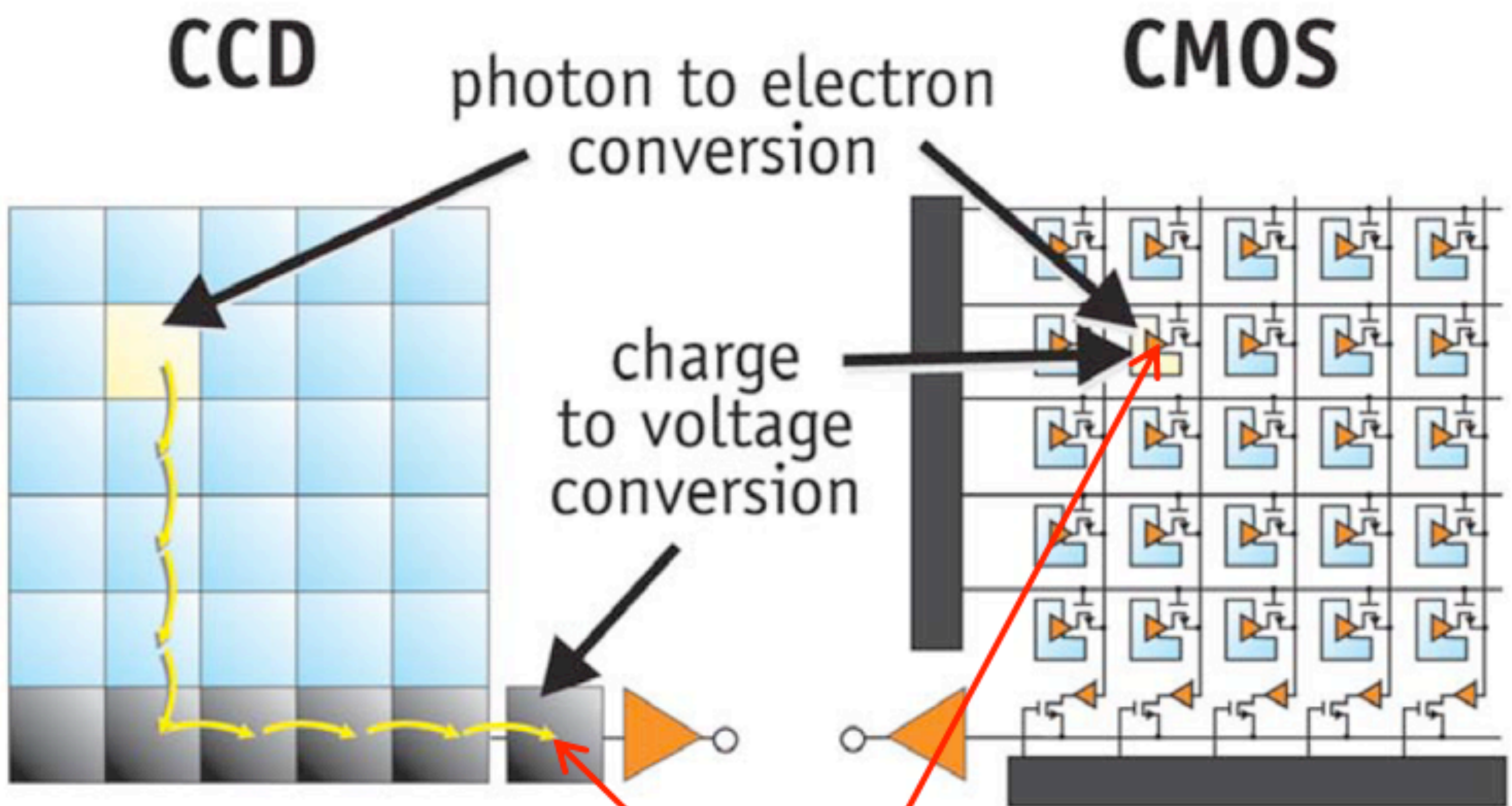


Figure 4.6: Bias Frame – Bias Jump is seen in the WFC1 Quadrant B Only. The vertical stripes in the data are hot columns and are unrelated to bias.

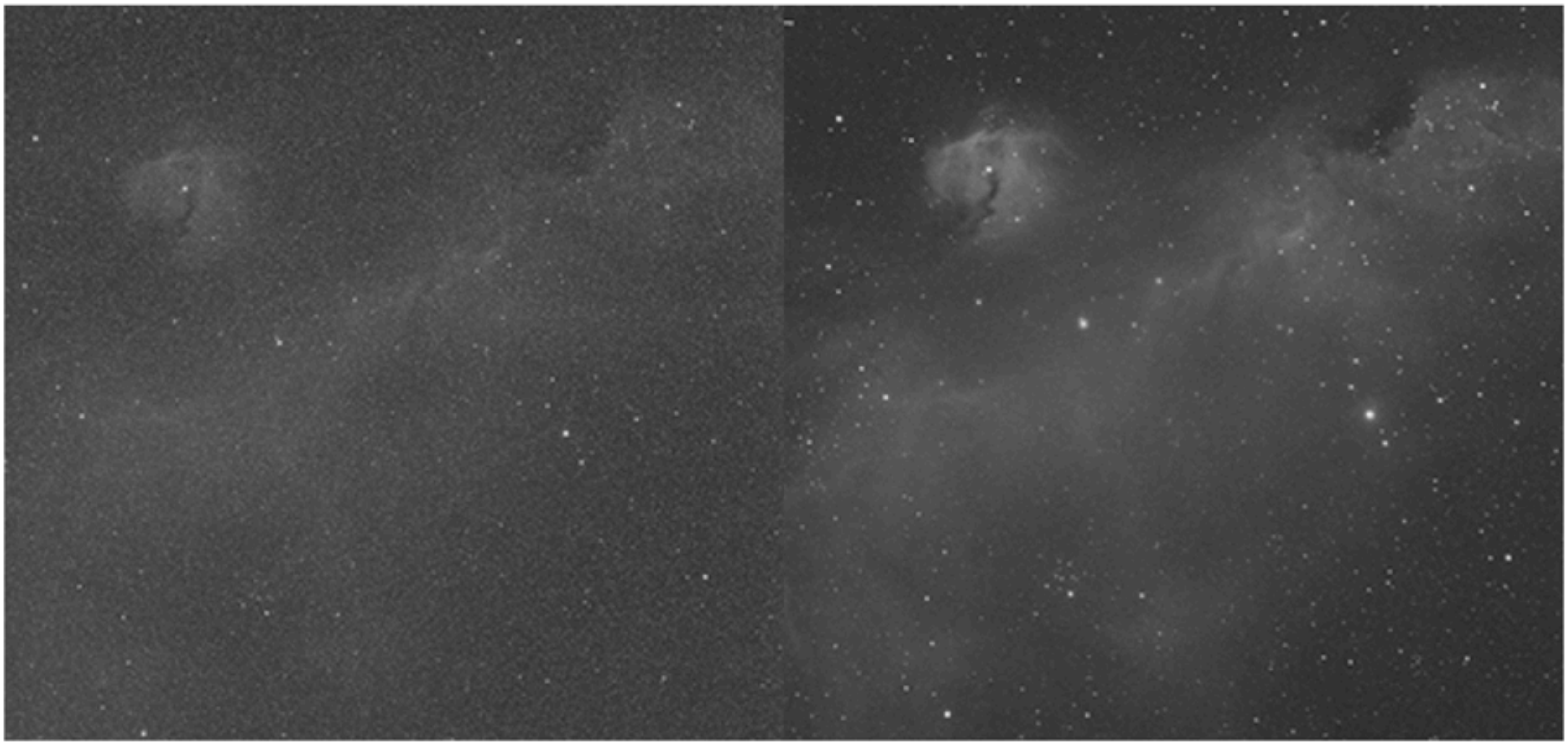




CCDs move photogenerated charge from pixel to pixel and convert it to voltage at an output node. CMOS imagers convert charge to voltage inside each pixel.

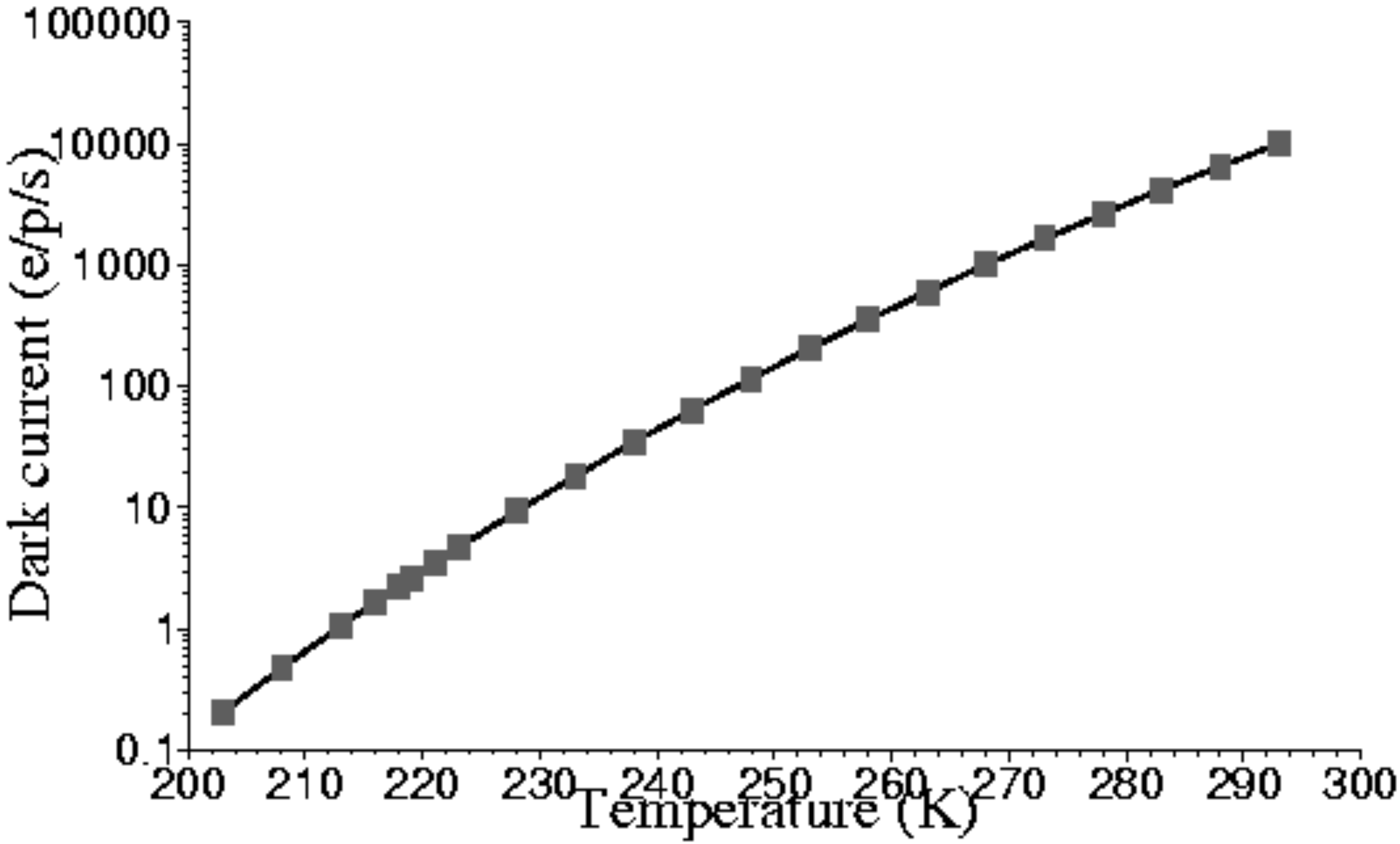
Even without bias (on the left), you'll still have other sources of noise

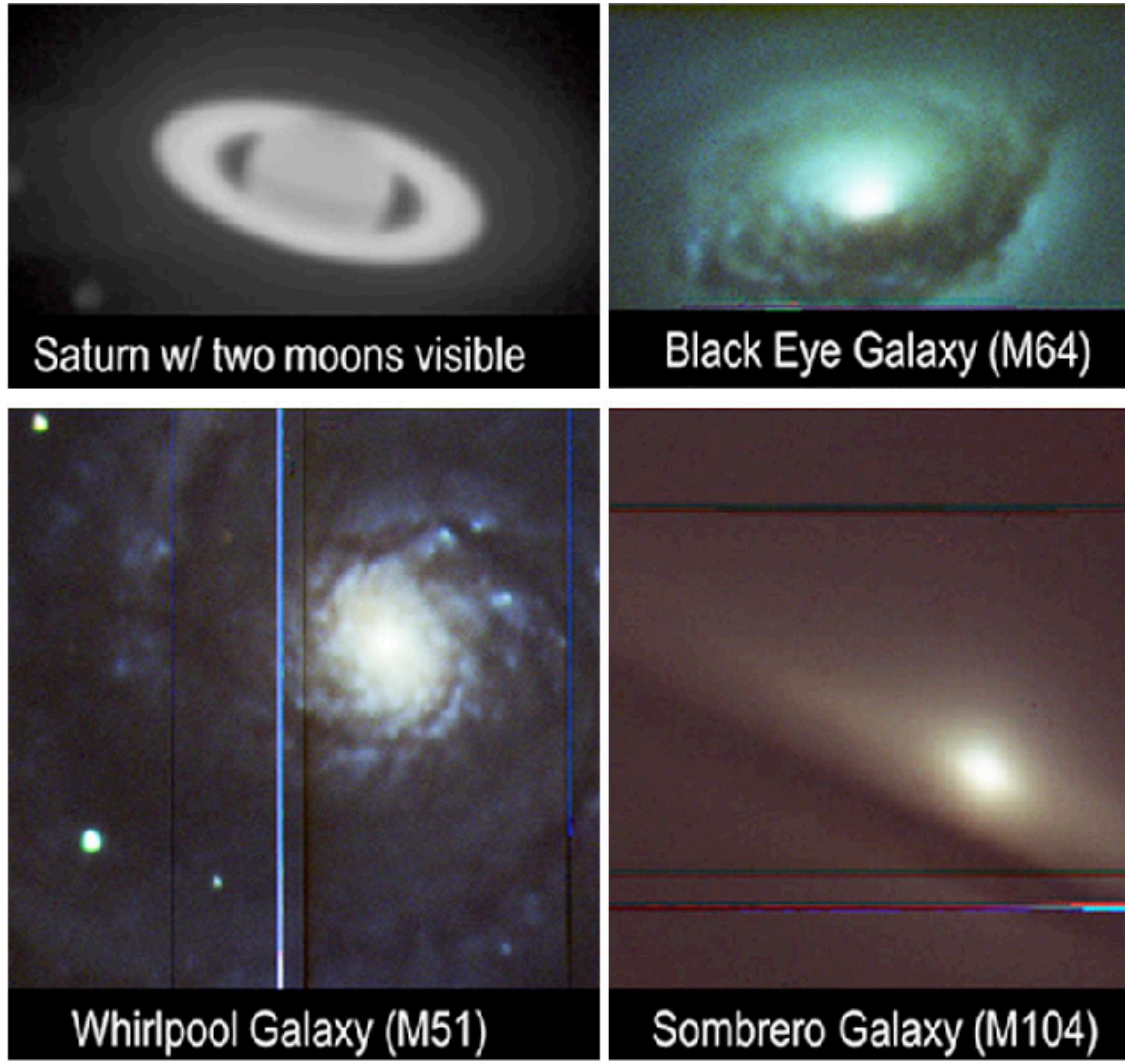
Same source, same exposure time, site and atmosphere, same detector + telescope, both have bias removed



One is clearly lower S/N than the other.
So what might be different? Hint: Planck Distribution

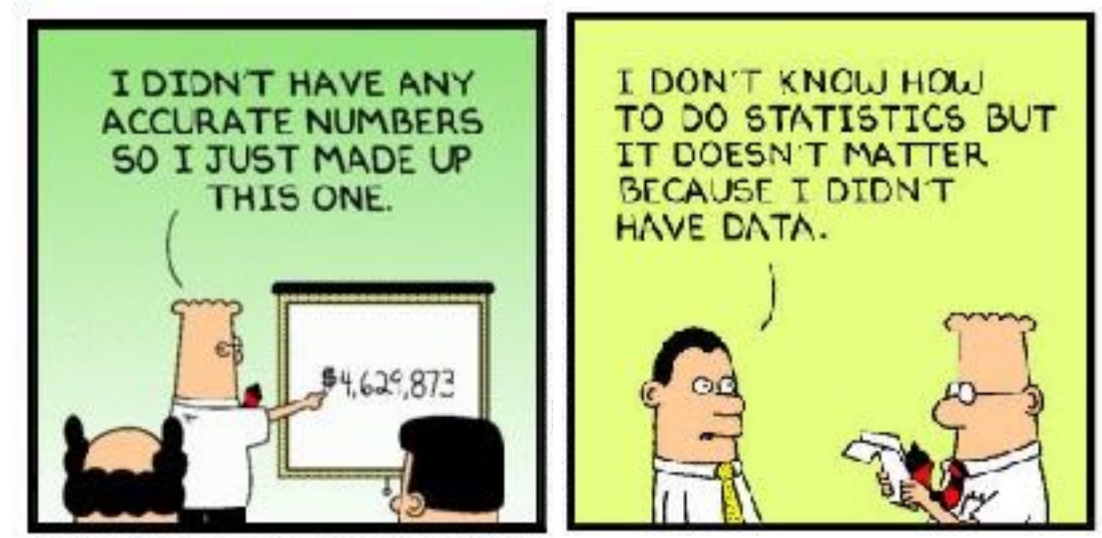
Dark current is unavoidable at some level, but can be reduced by cooling





Most detectors are stuck in a dewar and cooled electrically or with LN2/coolant to reduce dark noise https://www.astropy.org/ccd-reduction-and-photometry-guide/v/dev/notebooks/01-03-Construction-of-an-artificial-but-realistic-image.html

What is a Statistic?



A statistic summarizes data.

Examples:
- mean brightness of an object
- center of a PSF of a source
- a half-light radius of a galaxy
- number of galaxies per square degree

A statistic can be related to the assumed probability distribution parameters

"A number without an error is useless."

What use are statistical methods?

- Help you make a decision!

  Is a signal in a set of observations meaningful?

  Do the data fit our model of the phenomenon under study?

- Simulate Observations

  Plan size of sample, etc.

  What would happen if we repeated the observations?

- Compare different observations

  Are two sets of data consistent with each other?

  Are the observations truly independent?

5

Possible Statistics

"Most Likely" Statistics

Average $$\mu = \sum_j \frac{x_j}{n}$$

Median: Order data. $$med = x_j \, where \, j = N/2 + 1/2 \, (odd)$$
$$= \frac{x_j + x_{j+1}}{2} \, where \, j = N/2 \, (even)$$

Mode: Most frequently occurring value(s).

"Spread" Statistics

Variance $$s^2 = \sum_j \frac{(x_j - \mu)^2}{n-1}$$

Root mean square $$s = \sqrt{\sum_j \frac{(x_j - \mu)^2}{n-1}}$$

Mean deviation $$\overline{\Delta x} = \sum_j \frac{|x_j - \mu|}{n-1}$$

Sample Maximum

9

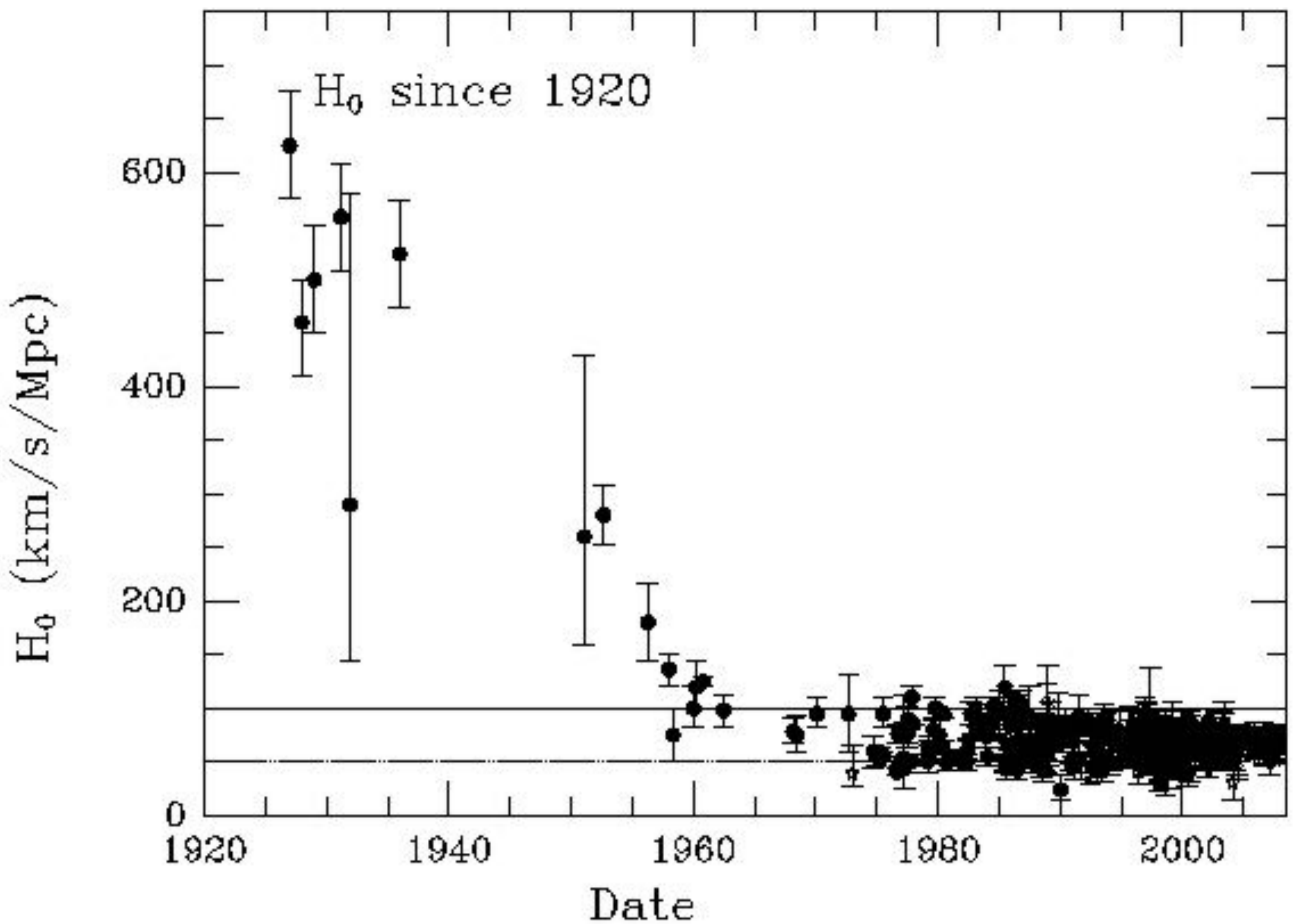
Basic Quantities

- Mean (average): $$\mu = \frac{1}{N}\sum_i x_i$$

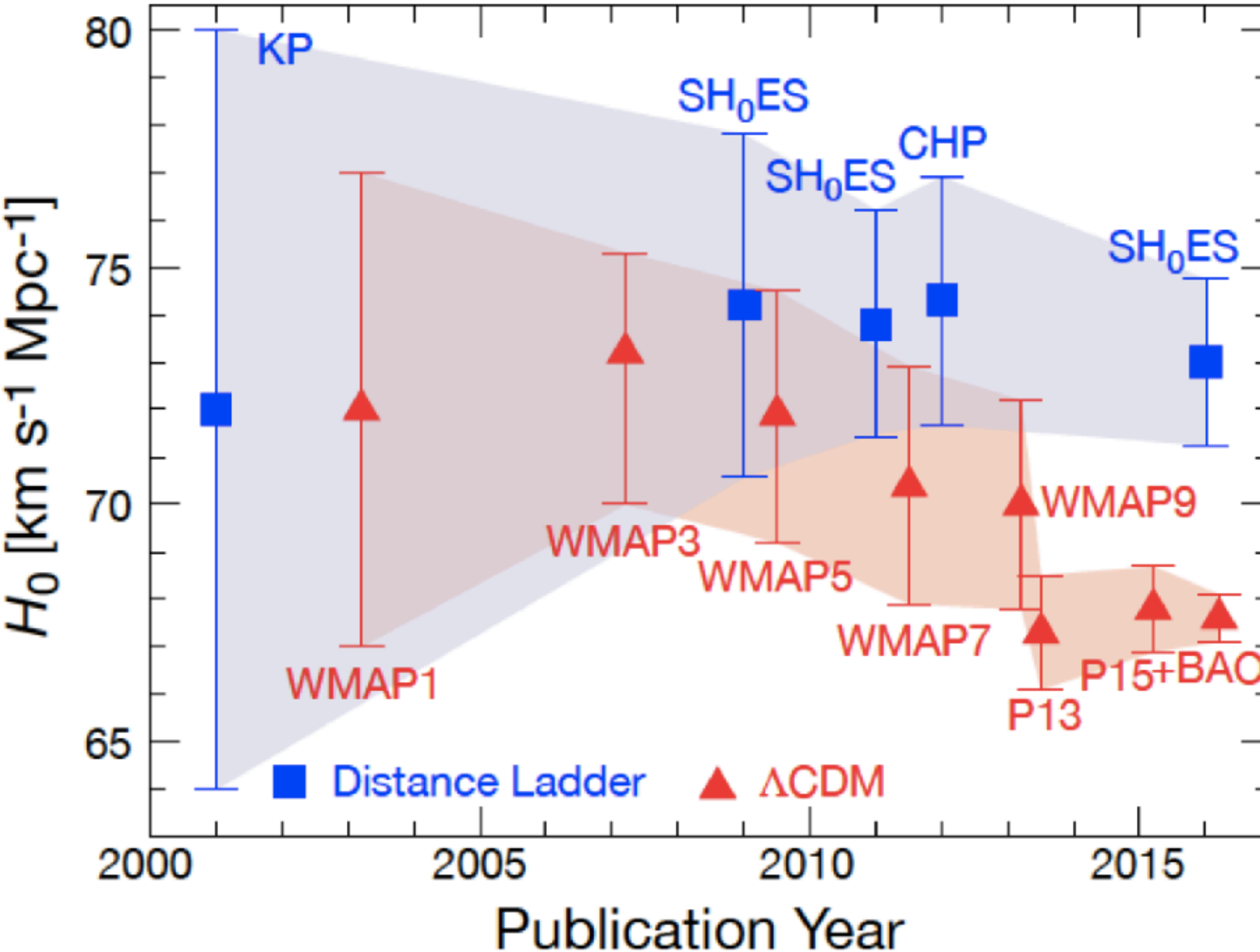
- Weighted mean: factors in uncertainty in individual measurements.

$$\mu = \frac{\sum_i x_i/\sigma_i^2}{\sum_i 1/\sigma_i^2}$$

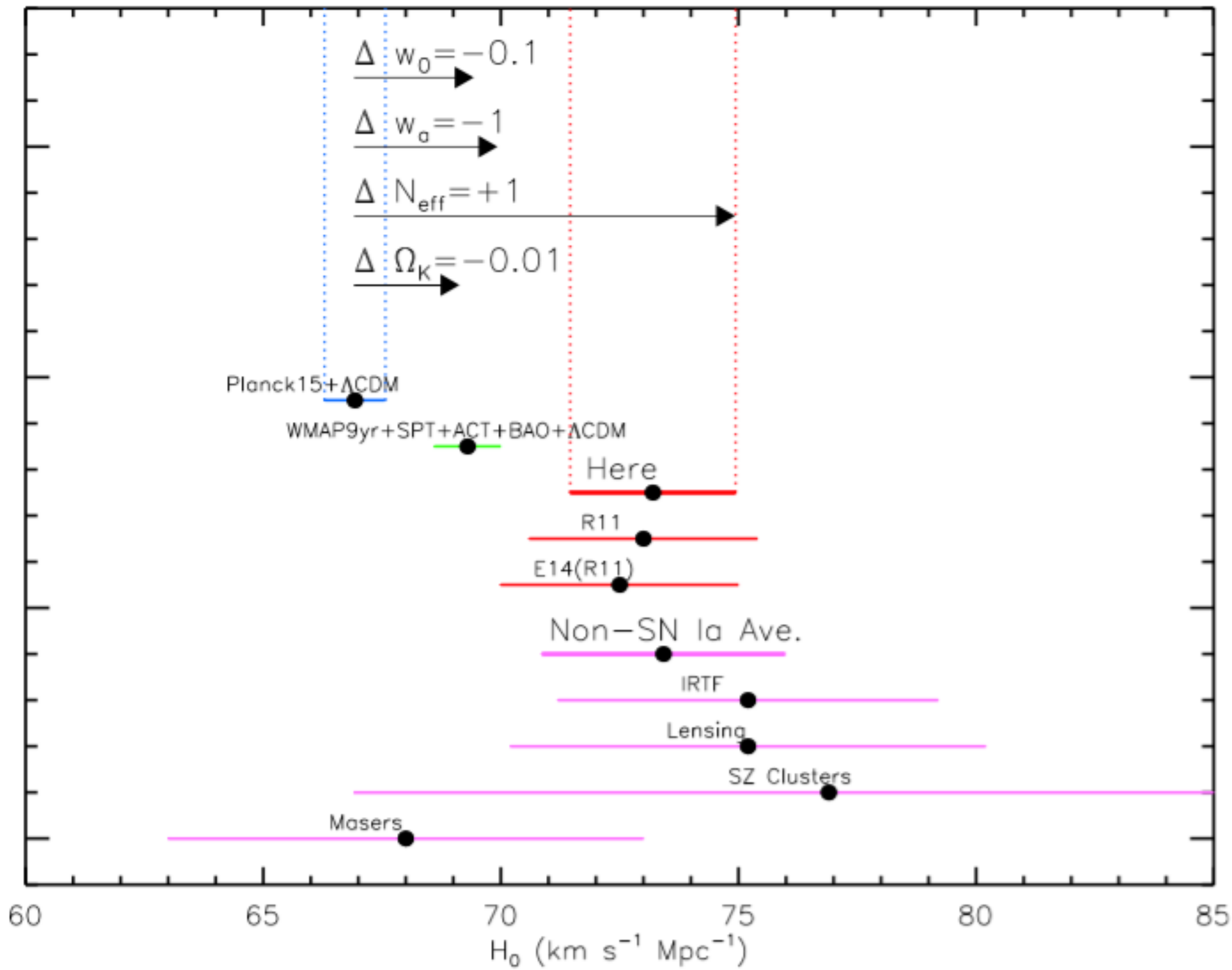
when would you use each of these?



H0 since 1920
600
400
200
0
1920
1940
1960
1980
2000
Date
H0 (km/s/Mpc)



KP
SH0ES
SH0ES
CHP
SH0ES
WMAP1
WMAP3
WMAP5
WMAP7
WMAP9
P13
P15+BAO
Distance Ladder
ΛCDM
80
75
70
65
2000
2005
2010
2015
H_0 [km s^{-1} Mpc^{-1}]
Publication Year


Δ w_0=−0.1
Δ w_a=−1
Δ N_{eff}=+1
Δ Ω_K=−0.01
Planck15+ΛCDM
WMAP9yr+SPT+ACT+BAO+ΛCDM
Here
R11
E14(R11)
Non−SN Ia Ave.
IRTF
Lensing
SZ Clusters
Masers
60
65
70
75
80
85
H_0 (km s^{-1} Mpc^{-1})

Know your statistic

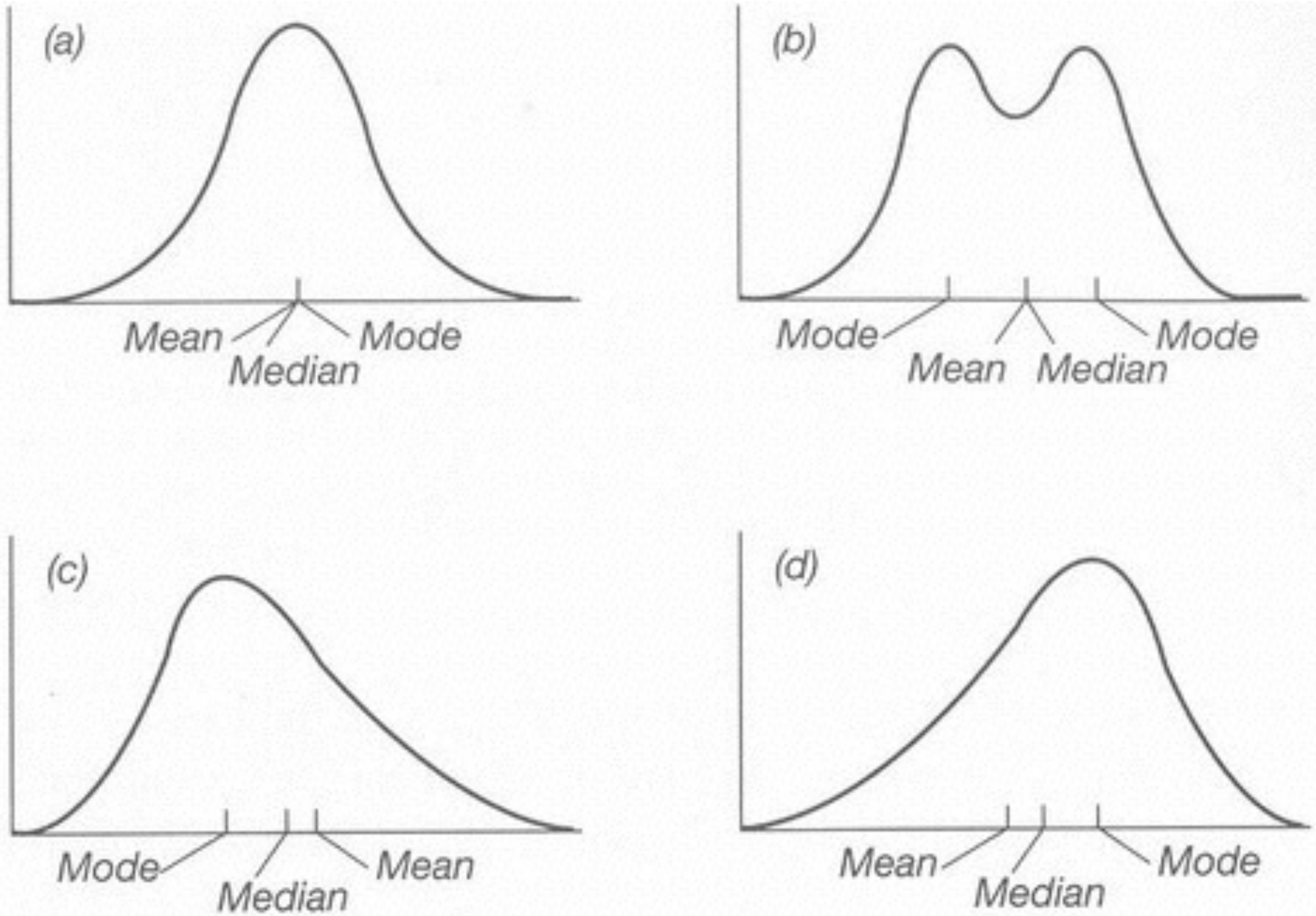
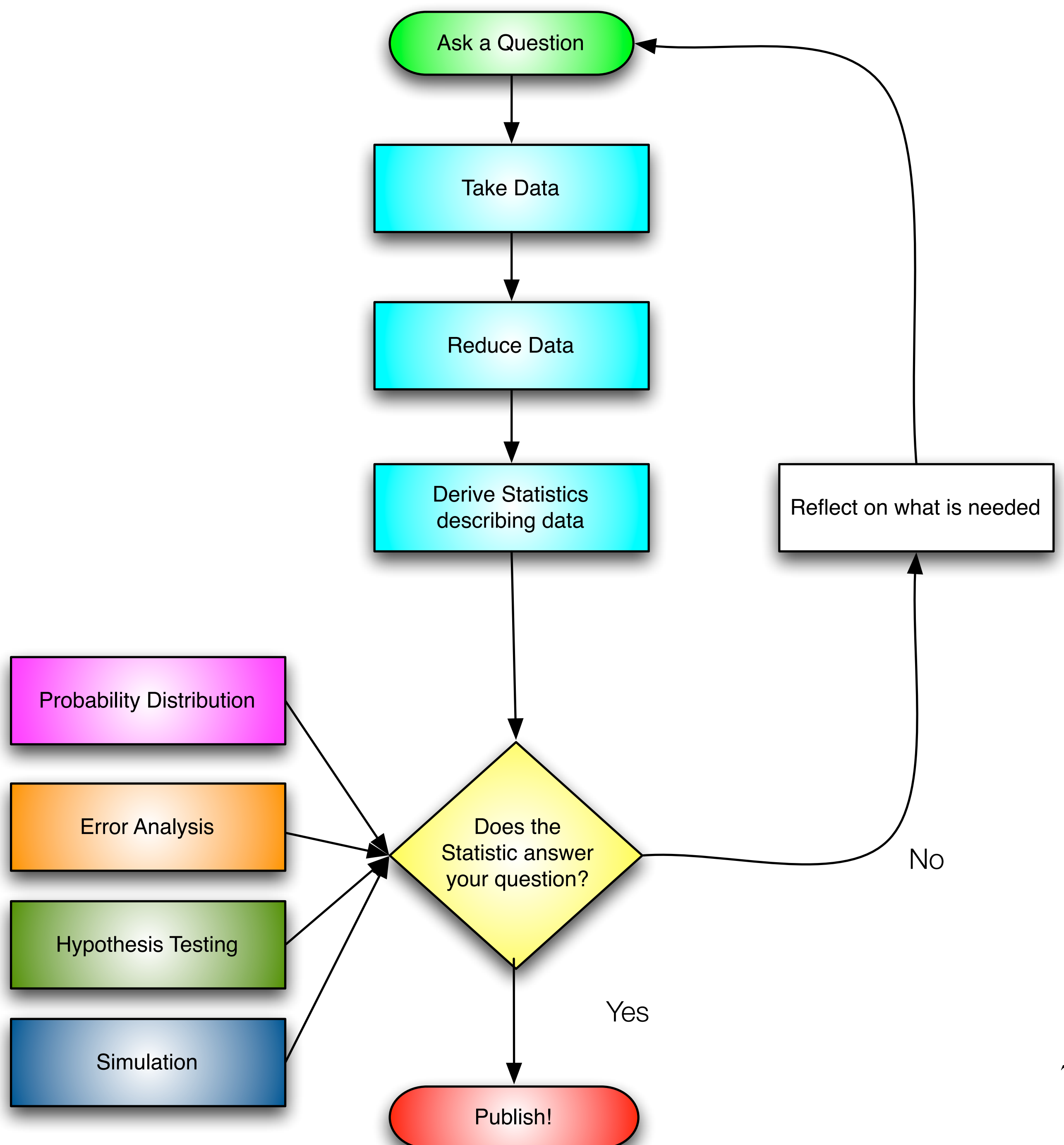


Figure 3.2 Frequency distributions showing measures of central tendency. Values of the variable are along the abscissa (horizontal axis), and the frequencies are along the ordinate (vertical axis). Distributions (a) and (b) are symmetrical, (c) is positively skewed, and (d) is negatively skewed. Distributions (a), (c), and (d) are unimodal, and distribution (b) is bimodal. In a unimodal asymmetric distribution, the median lies about one-third the distance between the mean and the mode.*

from Biostatistical Analysis, fourth edition, Simon & Schuster 1999

Process of Decision Making



7

When to use:

What statistic would you use to estimate the true bias level for a CCD that is viewing a blank sky, but has cosmic rays (resulting in anomolously high counts for a few pixels)?

What statistic would be most useful for understanding the typical color of a guide star one would use for adaptive optics?

11

Good Statistics

Good statistics should be:

Unbiased - should converge to the right value with more data points.

Robust - should not be affected by a few bad data points.

Consistent - should not be affected, systematically by the size of your sample.

Close - should converge as quickly as possible with increasing data.

12

Relation to Probability distributions

Statistics are based on data only!

However, they are often most useful as estimators of parameters of probability distribution.

This is a "frequentist" approach where the distribution is used to determine how often we might obtain the resulting statistic, so that we can decide whether this is the correct model.

13

Systematic Errors

- Errors must follow the “principle of indifference” to really behave the way described on the previous slide.
  - There may be some hidden correlation which causes this to be violated.
- Errors must also be independent.
  - An error which is the exact same in each measurement sequence will not allow more precise measurement.
- If either of these are violated, we typically refer to these effects as “systematic” errors (or biases). They may provide a practical limit to the precision of a measurement.

16

Some definitions

Statistic – a number or set of numbers that describe a set of data.

Probability distribution – the relative chances for different outcomes for your set of data.

Sample distribution – Set of data that allow us to estimate useful values of the object under study.

Parent distribution – Presumed probability distribution of the data that one would measure if an infinite data set were acquired.

Mean - the 1st moment of a distribution, which gives information about the most likely value one will observe.

Variance – the 2nd moment of a distribution, which gives information about the range of values one will observe.

8

Probability distribution functions

Discrete

- Random variable ***X*** with discrete outcomes x_i, each with probability p_i.
- Probability $p_r(X = x_i) = p_i$
- Cumulative probability:

$$P(x) = \sum_{i}^{x_i < x} p_i.$$

- Probability density:

$$p(x) = \sum_i p_i \, \delta(x - x_i).$$

- Normalization:

$$\sum_i p_i = 1.$$

Continuous

- A probability density function (PDF) ***p(x)*** that defines the probability of drawing a sample with a value between ***a*** and ***b*** of:

$$p_r(a < x < b) = \int_a^b p(x)\,dx$$

- Cumulative distribution function (CDF):

$$P(x) = \int_{-\infty}^{x} p(u)\,du$$

- Normalization:

$$\int_{-\infty}^{+\infty} p(x)\,dx = 1$$

Example of discrete PDF and CDF

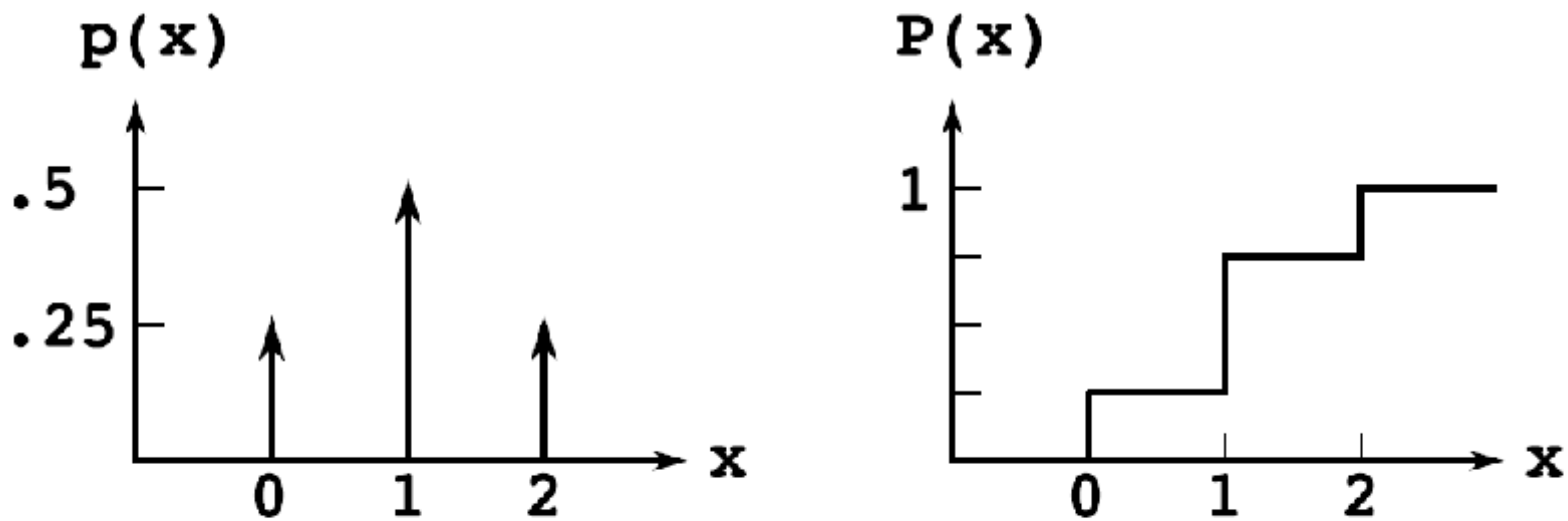


Fig. 6.1.— Probability density p(x) and cumulative probability P(x) for the numbers of "heads" achieved after two flips of a coin.

$$P(x) = \int_{-\infty}^{x} p(u)\,du$$

$$p(x) = \frac{dP(x)}{dx}$$

Moments of a PDF

- Expectation value ***<x>*** or ***E[x]*** of a random variable ***x*** – the mean value obtained in an infinite number of trials – an ***ensemble average*** over realizations.
- In terms of a PDF ***p(x)***:

$$< \mathrm{x} > = \int_{-\infty}^{\infty} \mathrm{x}\ \mathrm{p}(\mathrm{x})\ \mathrm{dx}.$$

- The *n-th central moment* is:

$$< (x - < x >)^n >$$

Mean and variance: 1st and 2nd moments

- The first and second moments of the PDF are the mean **μ** and variance **σ²** respectively:

$$\mu = < \mathrm{x} > = \int_{-\infty}^{\infty} \mathrm{x}\ \mathrm{p}(\mathrm{x})\ \mathrm{dx}.$$

$$\sigma^2 = \int_{-\infty}^{\infty} (\mathrm{x} - \mu)^2\ \mathrm{p}(\mathrm{x})\ \mathrm{dx} = < \mathrm{x}^2 > - < \mathrm{x} >^2,$$

Distribution Examples

- Binomial:
  - Chance of seeing at least one AGN with p=0.03 in a sample of 100 galaxies.
  - Chance of being weathered out in a three night run, if nights are usable 90% of the time.
- Poisson:
  - Photometry precision in a photon counting detector with N photons detected.
  - Uncertainty in star counts in two spectral bins.
- Gaussian:
  - Angular separation measurement variations of a binary made in many individual images.
  - FWHM measurement of a marginally resolved source.

4

Binomial PDF

0 or 1

- The probability of ***x*** events in ***n*** tries if the probability of each event is ***κ:***

$$\text{Probability density function: } p(x;n,\kappa) = \binom{n}{x}\kappa^{x}(1-\kappa)^{n-x}$$

$$\text{Binomial coefficient: } \binom{n}{x} = \frac{n!}{(n-x)!\,x!}$$

- Derivation:
  - Probability of zero events: $p(0) = (1-\kappa)^{n}$
  - Probability of 1 event: $p(1) = n\kappa(1-\kappa)^{n-1}$
  - …

The Binomial distribution

- You are observing something that has a probability, p, of occurring in a single observation.
- You observe it M times.
- Want chance of obtaining n successes. For one, particular sequence of observations the probability is:

$$P_1(n) = p^n (1-p)^{M-n}$$

- There are many sequences which yield n successes:

$$P(n) = \frac{M!}{n!(M-n)!} p^n (1-p)^{M-n} = \begin{pmatrix} M \\ n \end{pmatrix} p^n (1-p)^{M-n}$$

Mean	Variance
Mp	Mp(1-p)

Often said
“M choose n”
14

multiple discrete outcomes:

The Poisson Distribution

- Consider the binomial case where p → 0, but Mp → μ. The binomial distribution, then becomes:

$$P(n) = \mu^n \frac{e^{-\mu}}{n!}$$

Mean	Variance
Mp=μ	Mp(1-p)~Mp=μ

15

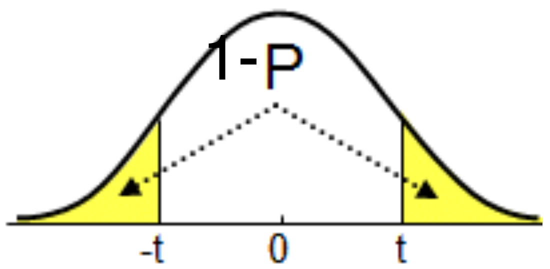
Gaussian Distribution

The Gaussian distribution is often used (sometimes incorrectly) to express confidence.

$$P(|x < \mu + \sigma|) > 0.68$$

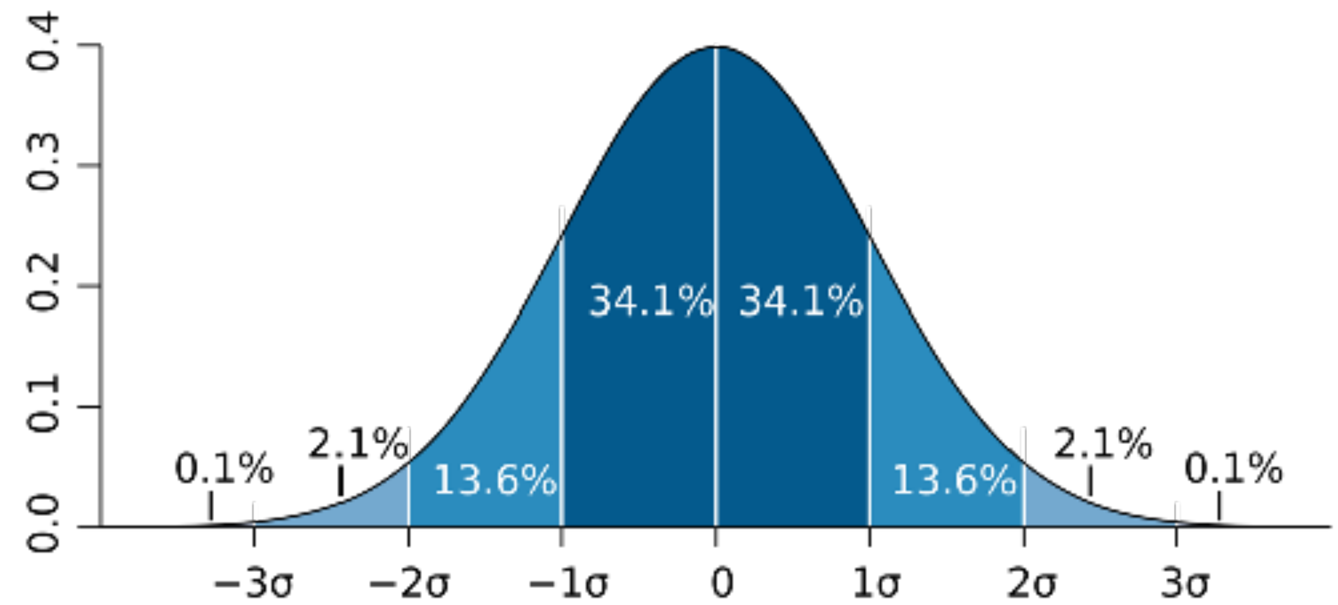
$$P(|x < \mu + 2\sigma|) > 0.95$$

$$P(|x < \mu + 3\sigma|) > 0.997$$



17

Standard Deviation



- Variance:

$$\sigma^2 \equiv \langle (x-\mu)^2 \rangle,$$

$$\sigma = \sqrt{\langle x^2 \rangle - \langle x \rangle^2}$$

- Standard Deviation:

$$\sigma_x = \sqrt{\frac{1}{N-1}\sum_i (x_i - \bar{x})^2}$$

Standard deviation is important for determining uncertainties. For a Gaussian distribution,

⇒ 68.3% probability that the true values is within 1σ of the mean

⇒ 95.4% probability that the true values is within 2σ of the mean

⇒ 99.7% probability that the true values is within 3σ of the mean

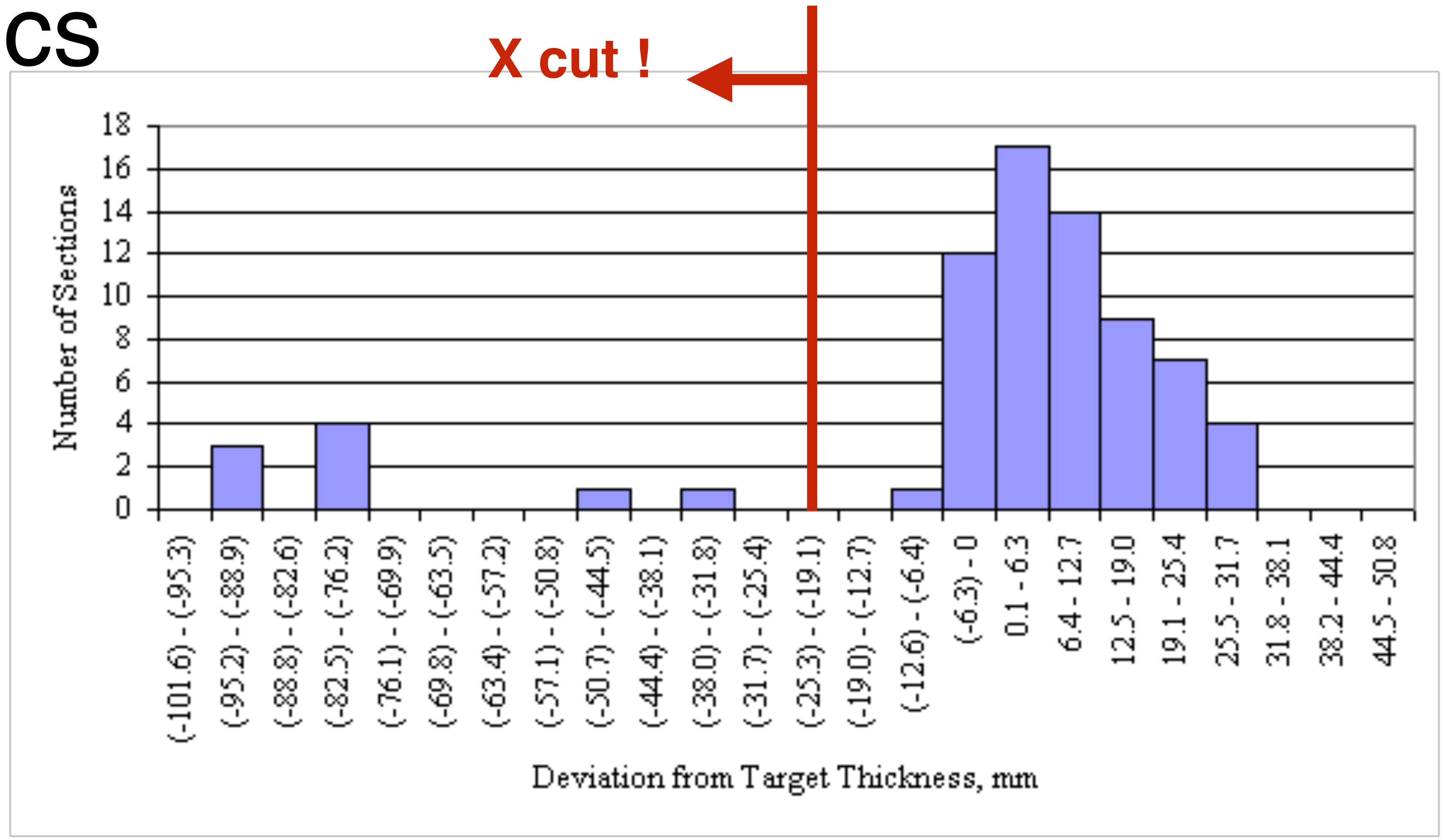
More generally, the probability that a point lies within N σ of the mean is

$$P = erf(N/\sqrt{2})$$

Note that for uniform observations, the standard deviation is proportional to $N^{-1/2}$.

Often a good approximation, but not robust to outliers.

Robust Statistics



- Iterative ***"Sigma Clipping"*** is a common technique in astronomy for dealing with outliers, particularly when combining images to eliminate artifacts such as cosmic rays. Another example would be in computing the velocity dispersion of a galaxy cluster, for which the "outliers" are galaxies not associated with the cluster.
- Basic method:
  1. Compute the mean (median) and standard deviation.
  2. Reject all points >N σ away from the mean (median) as outliers. A typical value for N might be 5 or 10.
  3. Recompute the mean (median) and standard deviation, and again reject outliers.
  4. Repeat until you are no longer rejecting any points.

Poisson Distribution

The Poisson distribution is of fundamental importance in astronomy. Put simply, the Poisson distribution
asks how many times an event is likely to happen in a given amount of time. Examples of applications in
the real would include:

How many bagels will be sold at Einstein's Bagels today?
How many cell phones will be sold in the US this week?

In astronomy, the question is: How many photons will hit my detector while I'm observing an object?

In all cases the answer can be any ***non-negative integer***.
The probability (P) of getting a certain number (x) of occurrences is:

$$P_p(x, \mu) = \frac{\mu^x}{x!} e^{-\mu}$$

where μ is the expected number of events when averaged over a very long period.

Poisson Distribution

Example: Assume that the average number of courses taken per semester by students at UF is 3 (and for the moment assume that you're allowed to take anywhere from 0 to infinitely many). If you ask someone how many they are taking, and if the distribution is Poisson, then:

P(0) = 0.02
P(1) = 0.07
P(2) = 0.15
P(3) = 0.20
P(4) = 0.20
P(5) = 0.16
P(6) = 0.10
P(7) = 0.06

Do you expect this distribution to be Poisson?

$$P(N = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

$$P_p(x, \mu) = \frac{\mu^x}{x!}\mathrm{e}^{-\mu}$$

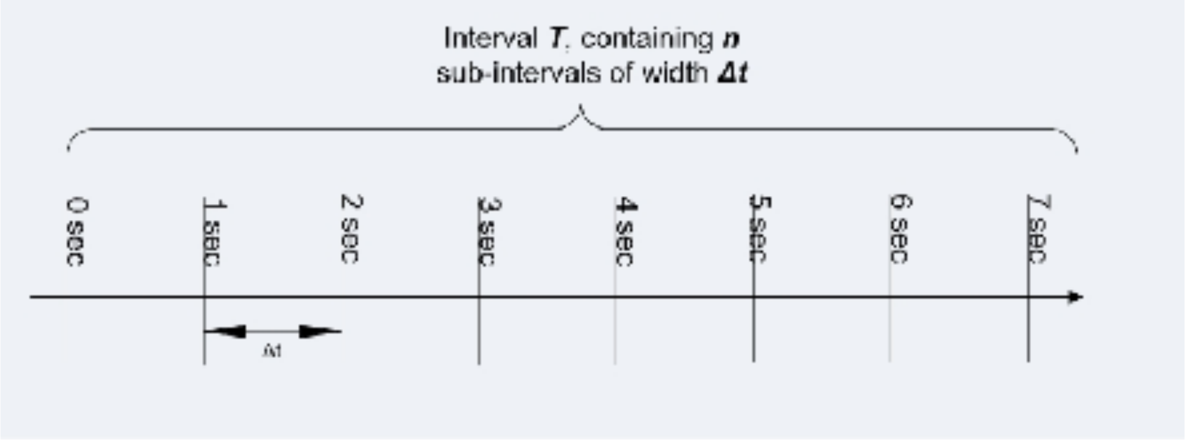
In the limit of λ becoming very large, the mean and standard deviation for a Poisson distribution approach:

$$\sigma = \sqrt{\lambda} \approx \sqrt{N}$$

For optical astronomy, where you have a lot of photons, these are very good approximations.

Poisson PDF

- Derived from the binomial distribution and vital for photon counting statistics. Consider photons arriving randomly over interval ***T*** with mean photon arrival rate ***F*** photons/s.



- Expected photon count over ***T*** is ***FT***. For large ***n***, the probability of a photon arrival in ***Δt*** is ***κ=FT/n***

$$\text{Binomial PDF:} \quad p(x;n,\kappa) = \binom{n}{x}\kappa^{x}(1-\kappa)^{n-x}$$

$$\text{Probability of photon count } x \text{ over } T\text{:} \quad p(x;n,FT) = \binom{n}{x}\left(\frac{FT}{n}\right)^{x}(1-\frac{FT}{N})^{n-x}$$

$$\lim_{N\to\infty} p(x) = p(x;FT) = \frac{(FT)^{x}e^{-FT}}{x!} \qquad \text{(the Poisson distribution)}$$

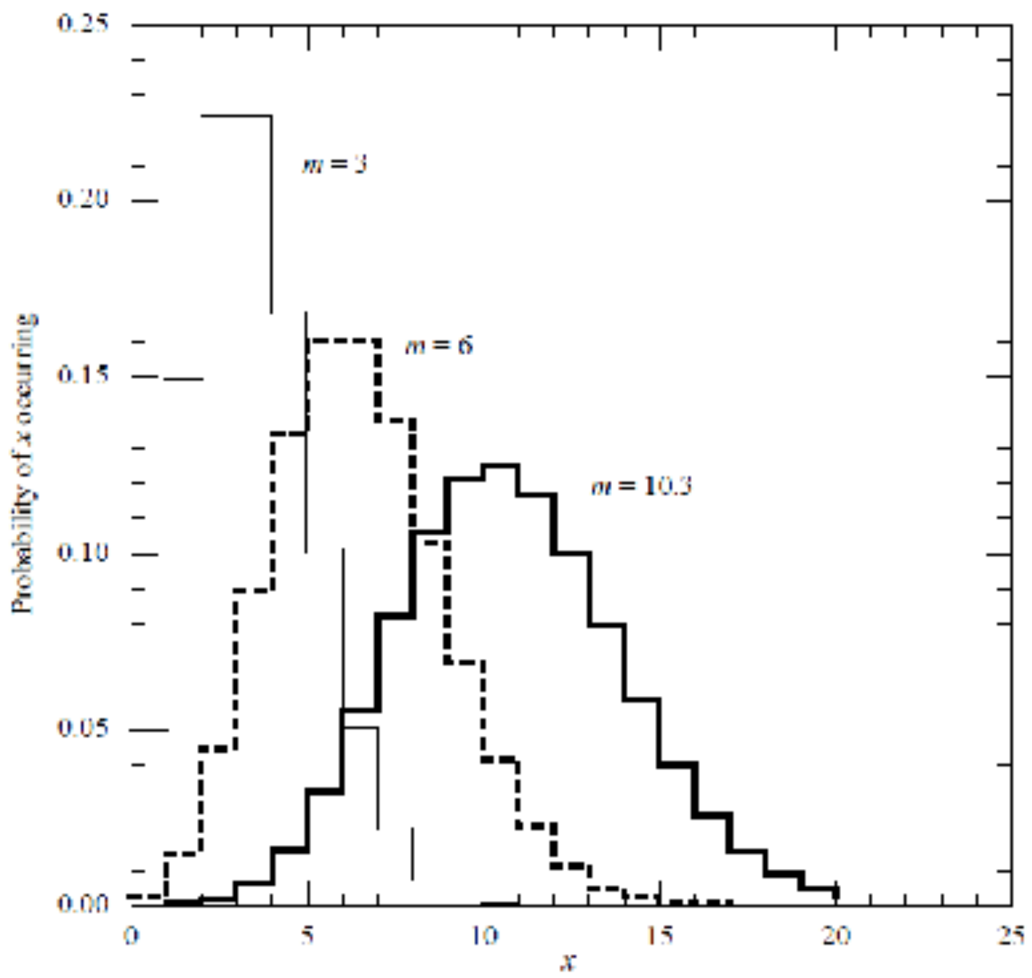
Poisson PDF

Define mean rate: $m = FT$

Probability density function: $p(x;m) = \dfrac{m^x e^{-m}}{x!}$

Expectation value: $\langle x \rangle = m$

Variance: $\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2 = m$



(Bradt 2004)

Expected Value

$$E(x) = \langle x \rangle = \int_X x \cdot p(x) dx$$

$$E(f(x)) = \langle f(x) \rangle = \int_X f(x) \cdot p(x) dx$$

Variance

$$\mathrm{Var}(x) = E([x - \langle x \rangle]^2)$$

nth moment (non-central)

$$\mu_n(x) = E(x^n)$$

nth moment (central)

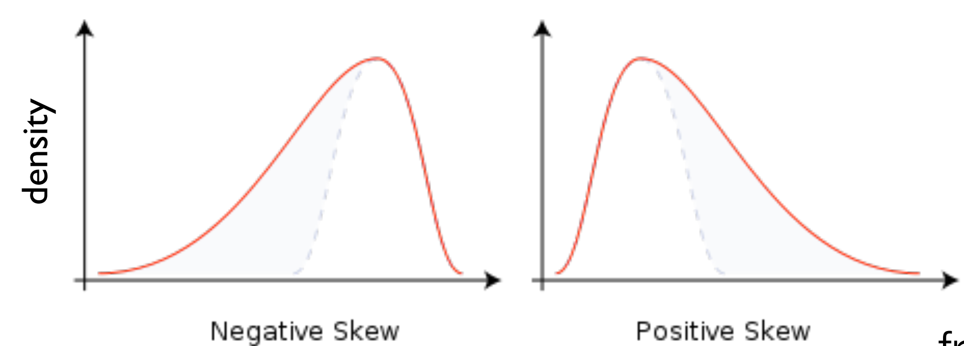
$$\widetilde{\mu}_n(x) = E([x - \langle x \rangle]^n)$$

3d and 4th moments of a distribution

- Skewness, asymmetry

$$\mu_3/\sigma^3 = \int_{-\infty}^{\infty} (x-\mu)^3 f(x) dx/\sigma^3$$

Normal: 0
Poisson: 1/√λ



from wikipedia

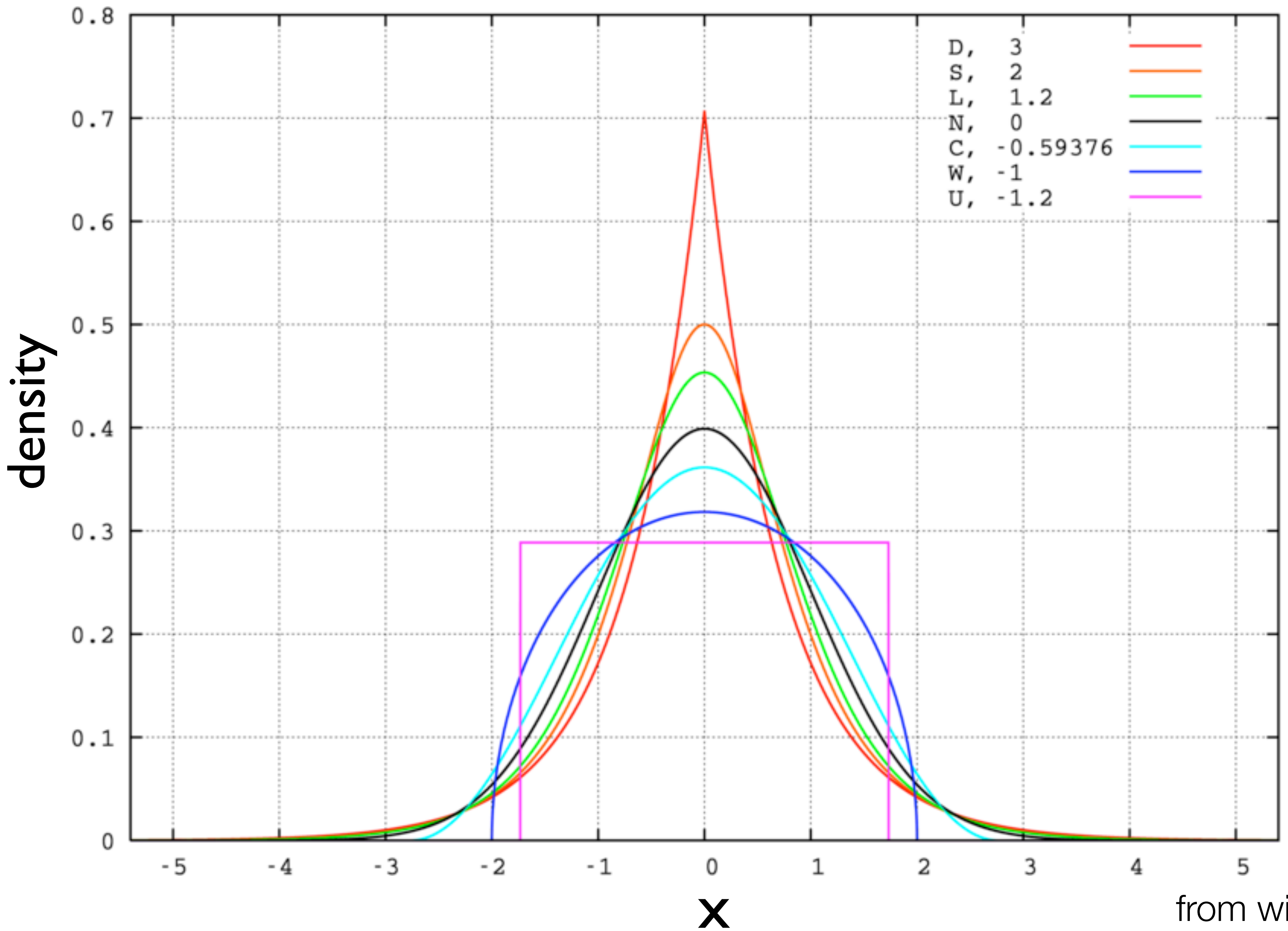
- Kurtosis

$$\mu_4/\sigma^4 = \int_{-\infty}^{\infty} (x-\mu)^4 f(x) dx/\sigma^4$$

$$\mu_4/\sigma^4 - 3$$

Normal: 0
Poisson: 1/λ

Example of different values of kurtosis: "boxiness" -- tail heaviness



from wikipedia

Gaussian or normal PDF

Probability density function: $p(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

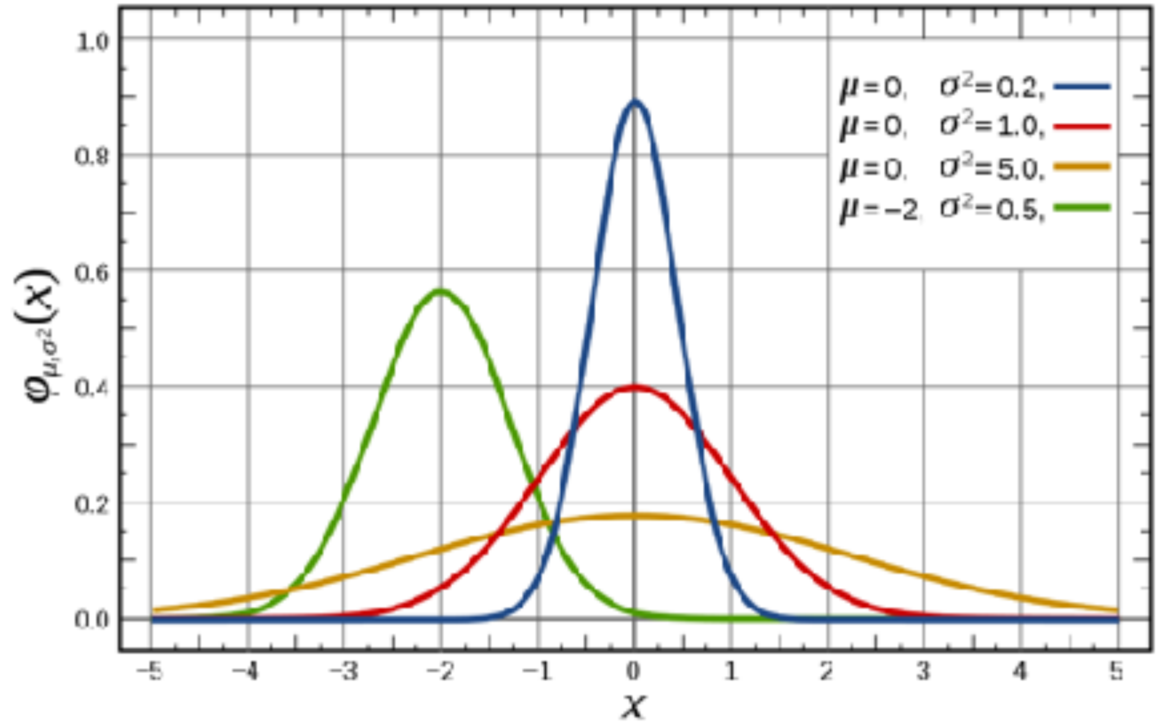
$p_r(-\sigma \le x \le \sigma) = 1 - 0.32$

$p_r(-2\sigma \le x \le 2\sigma) = 1 - 0.046$

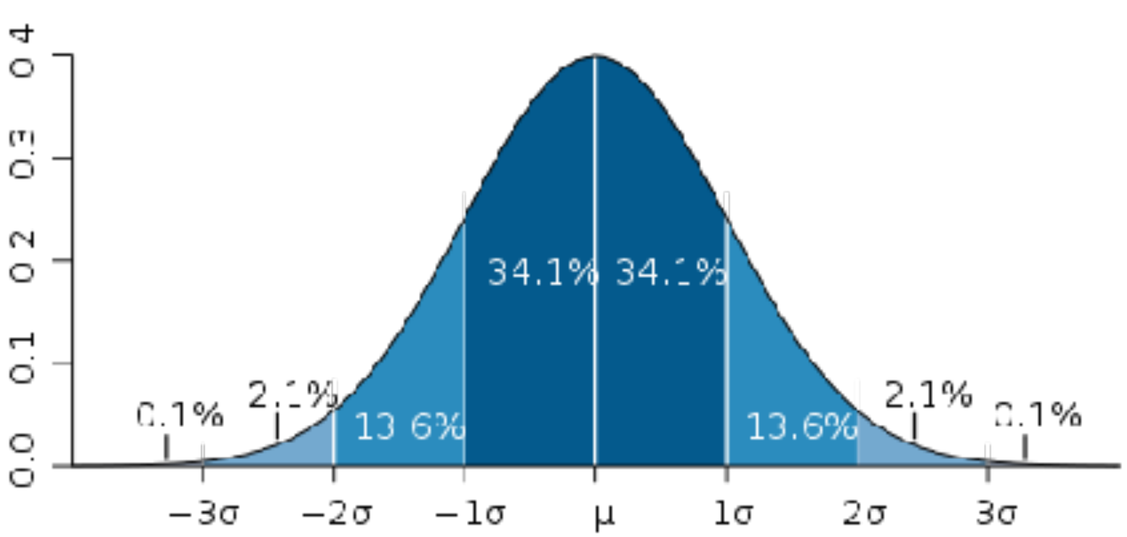
$p_r(-3\sigma \le x \le 3\sigma) = 1 - 0.0027$

Expectation value: $\langle x \rangle = \mu$

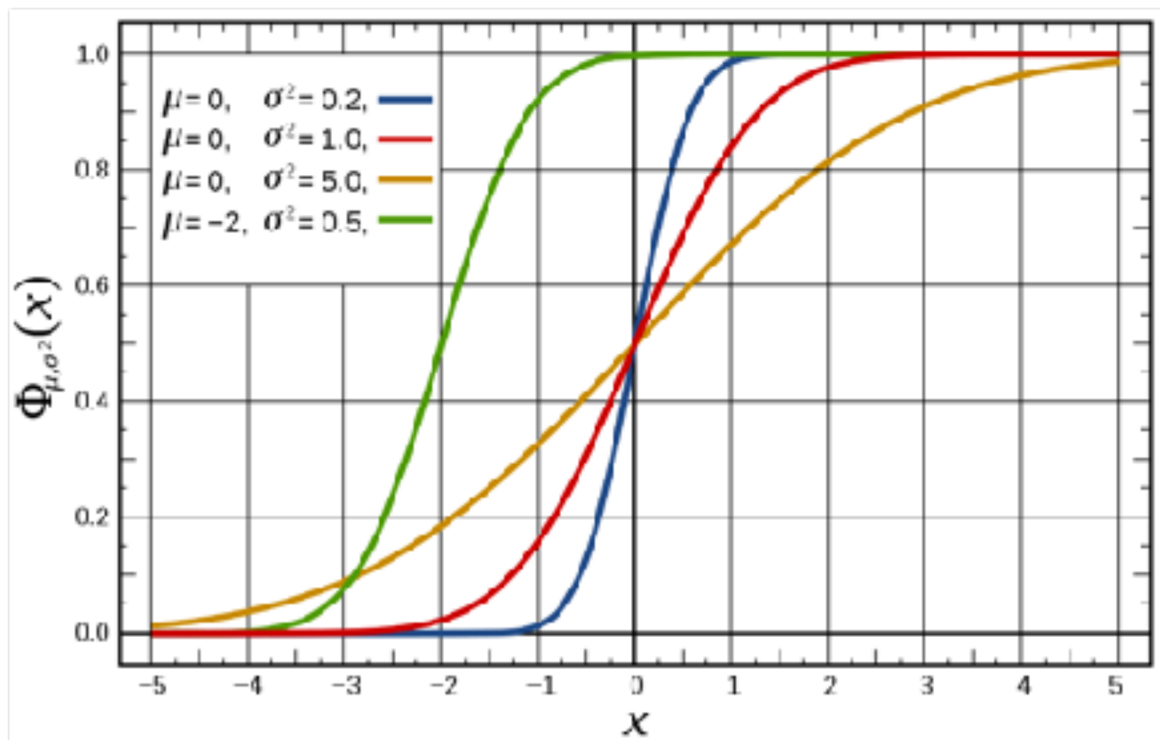
Variance: $\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2 = \sigma^2$



(Gaussian PDF; CC)



(Gaussian PDF; CC)



(Gaussian CDF; CC)

Gaussian CDF

- For a Gaussian PDF:

$$p(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-(x-\mu)^2/2\sigma^2}. \qquad \equiv N(\mu, \sigma^2)$$

- The cumulative distribution function is:

$$P(x) = \int_{-\infty}^{x} p(k)\, dk$$

$$= \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sqrt{2}\,\sigma}\right)\right],$$

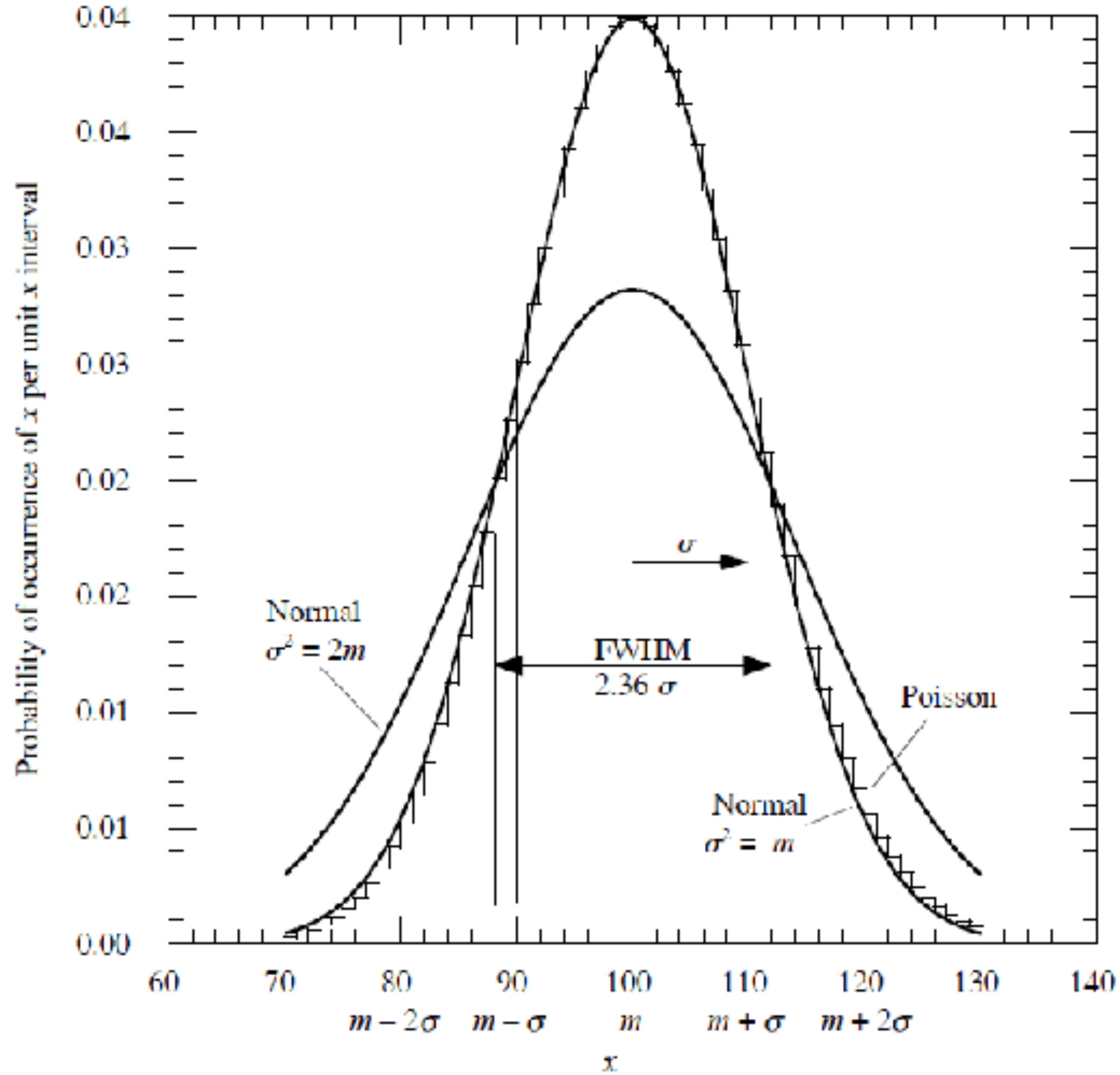
- Where the error function is defined as:

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt.$$

(computed numerically)

Convergence of Poisson and Gaussian distributions

- For modest ***m*** (> 10), the Poisson distribution is relatively well approximated by a Gaussian distribution.



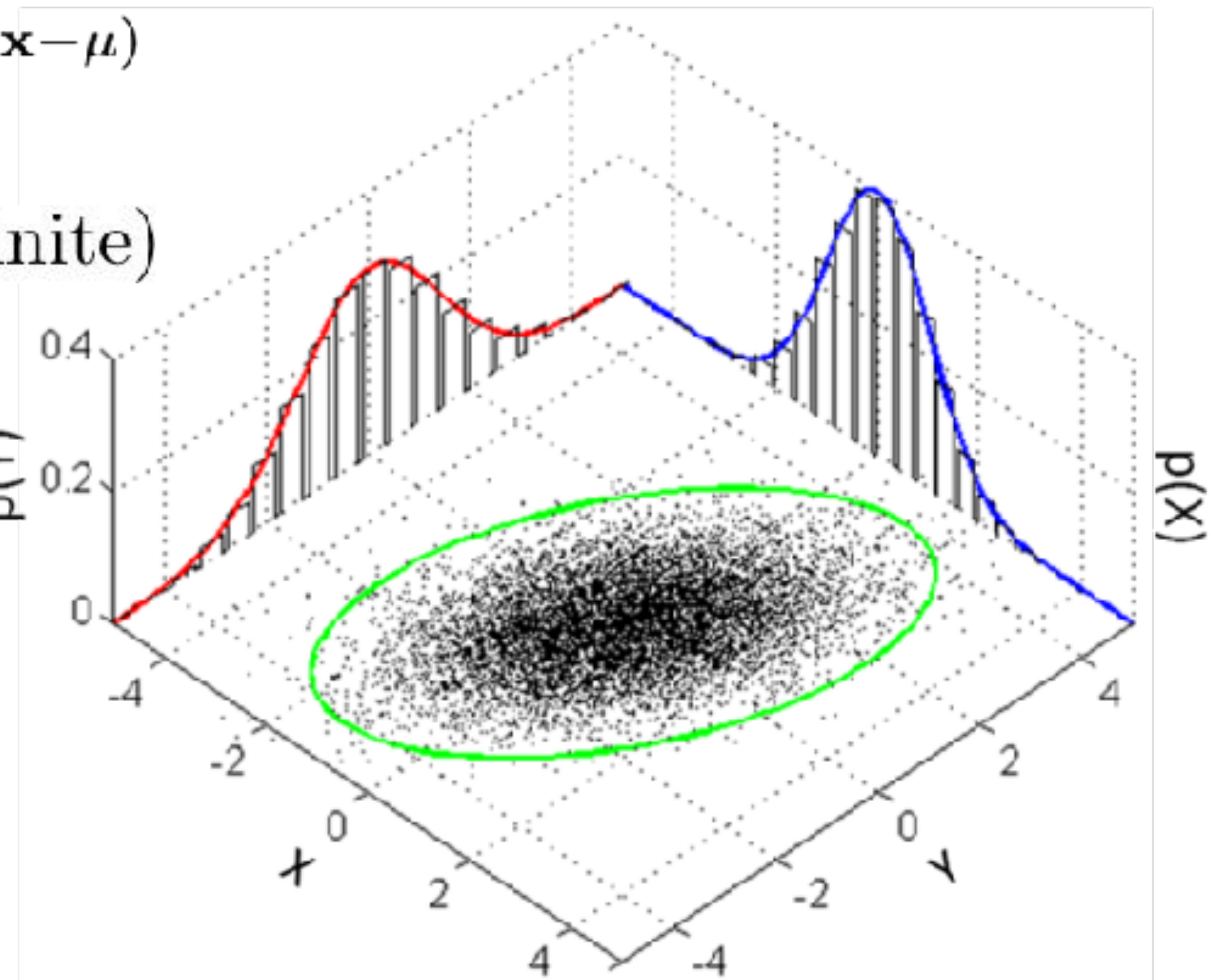
(Normal and Poisson distributions for m=100; Bradt 2004)

THE MULTIVARIATE NORMAL DISTRIBUTION

$$\phi(\mathbf{x}; \mu, \mathbf{\Sigma}) = \frac{1}{\sqrt{2\pi \det(\mathbf{\Sigma})}} e^{\frac{-1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1} (\mathbf{x}-\mu)}$$

$$\mu \in \mathbb{R}^n, \mathbf{\Sigma} \in \mathbb{R}^{n \times n} \text{(and is postive definite)}$$

Mean	μ
Median	μ
Mode	μ
Standard deviation	$\mathbf{\Sigma}$



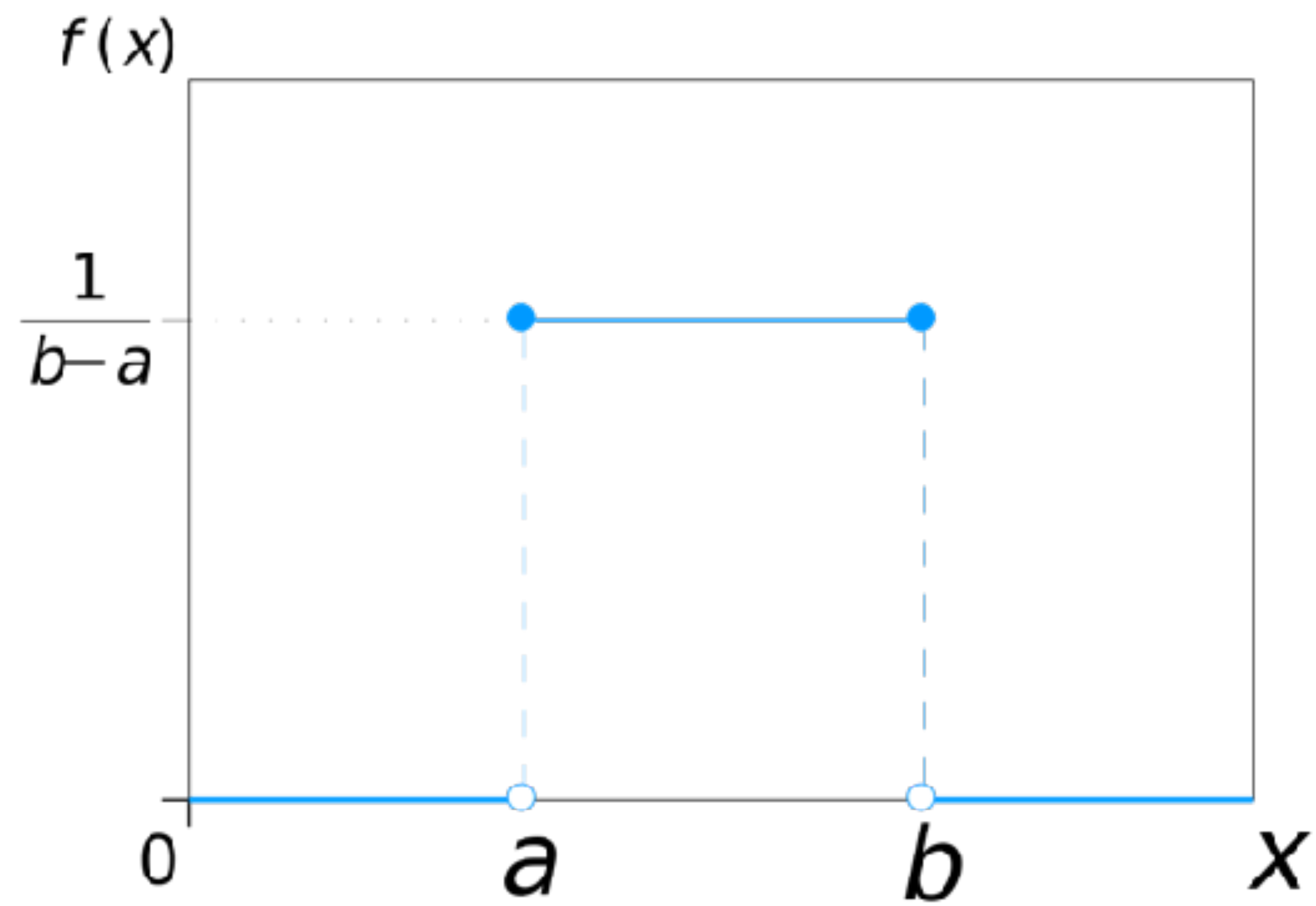
Frequently used as a prior when you don't know anything about the hypothesis

Used to generate random values from any distribution via probability integral transform (coming up)

THE UNIFORM DISTRIBUTION

$$p(x) = \frac{1}{b-a} \text{ for } x \in [a, b] \text{ else } 0$$

Mean	$\frac{1}{2}(a+b)$
Median	$\frac{1}{2}(a+b)$
Mode	Any value between a and b
Standard deviation	$\sqrt{\frac{1}{12}(b-a)^2}$



Distribution derived from Normal distribution
1) Chi square distribution

Modified from Maria Suveges, Lauren

If $X_i \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$ $\Longrightarrow$ $\sum_{i=1}^{k} X_i^2 \sim \chi_k^2$

iid= Independent identically distributed

$$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} \exp\left(-x/2\right)$$

mean: k

variance: 2 k

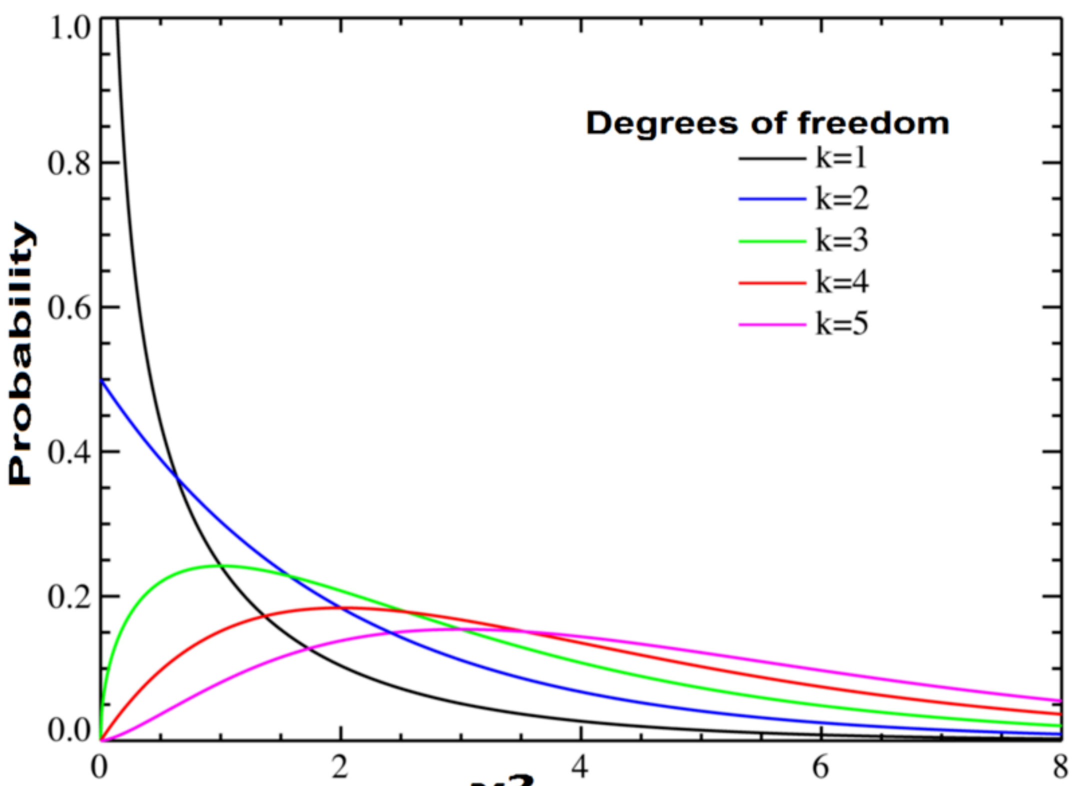
skewness: √(8/k)

kurtosis: 12/k

$$X_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma)$$

$$\sum_{i=1}^{k} (X_i - \bar{X})^2/\sigma^2 \sim \chi_{k-1}^2$$

When k is large χ_k^2 approximates a $\mathcal{N}(k, 2k)$



from wikipedia

Central limit theorem

The distribution of the mean of a sufficiently large number of random variables can be approximated by a Gaussian distribution!

$$X_i, \ i = 1, \ldots, n \ \ \text{iid} \ \ \text{with} \ \ \mathrm{E(X_i)} = \mu \ \ \mathrm{Var(X_i)} = \sigma^2$$

iid= Independent identically distributed

$$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \quad \text{follows approximately} \quad \mathcal{N}(0, 1)$$

⇓

One reason why
the Gaussian distribution is so important

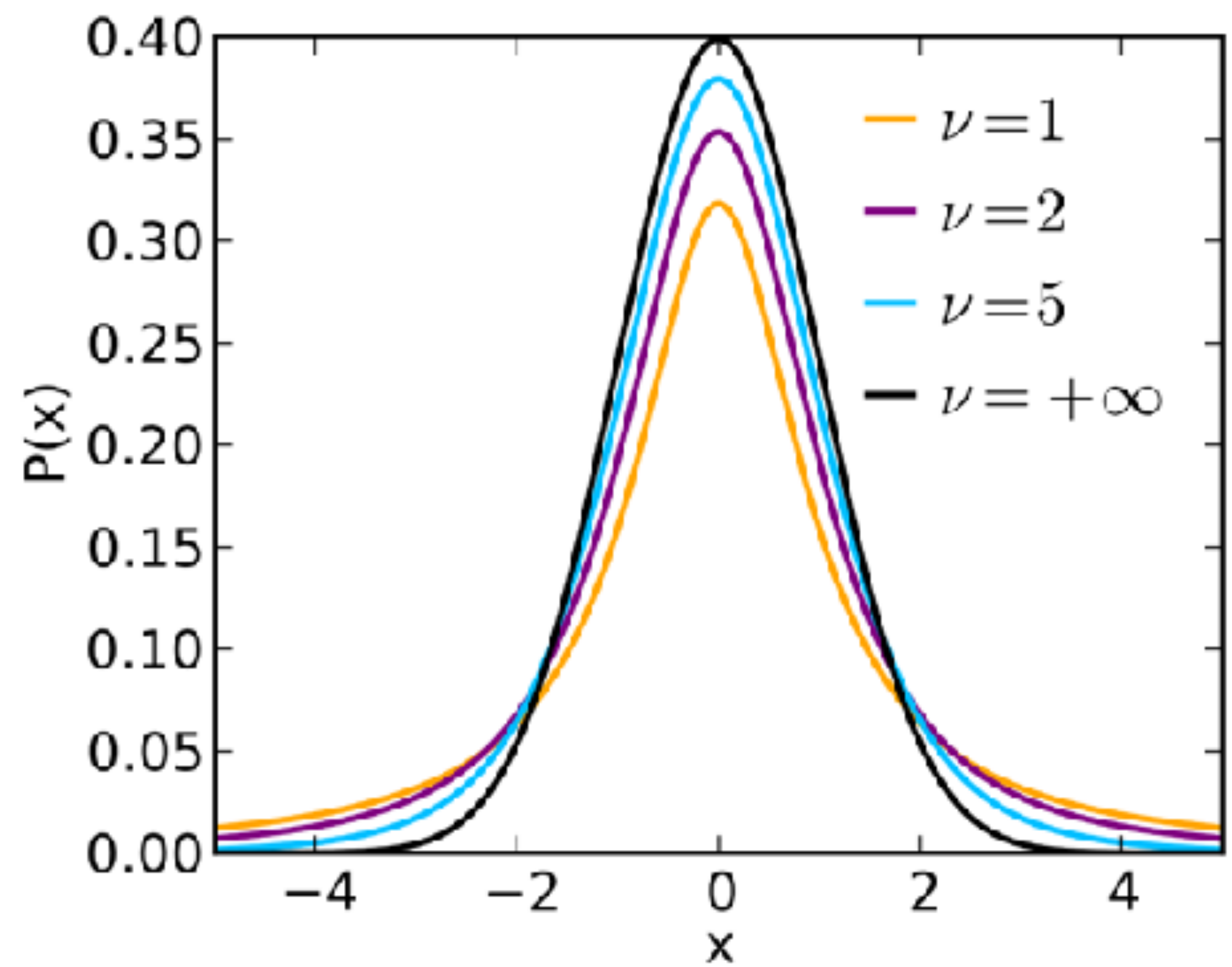
Useful for tests of significance of difference in means
Tests for significance of slope in linear regression

THE STUDENT'S T-DISTRIBUTION

$$p(x;\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\cdot\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

$$\nu \in \mathbb{R}^+, x \in \mathbb{R}$$

Mean	0 if $\nu > 1$
Median	0
Mode	0
Standard deviation	$\frac{\nu-2}{\nu}$ if $\nu > 2$ ∞ if $1 < \nu < 2$



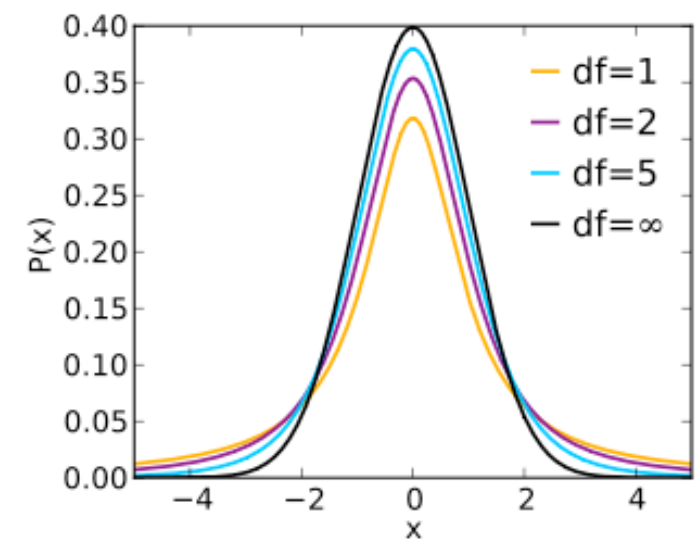
Distribution derived from Normal distribution
2) Student distribution

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$$

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$$

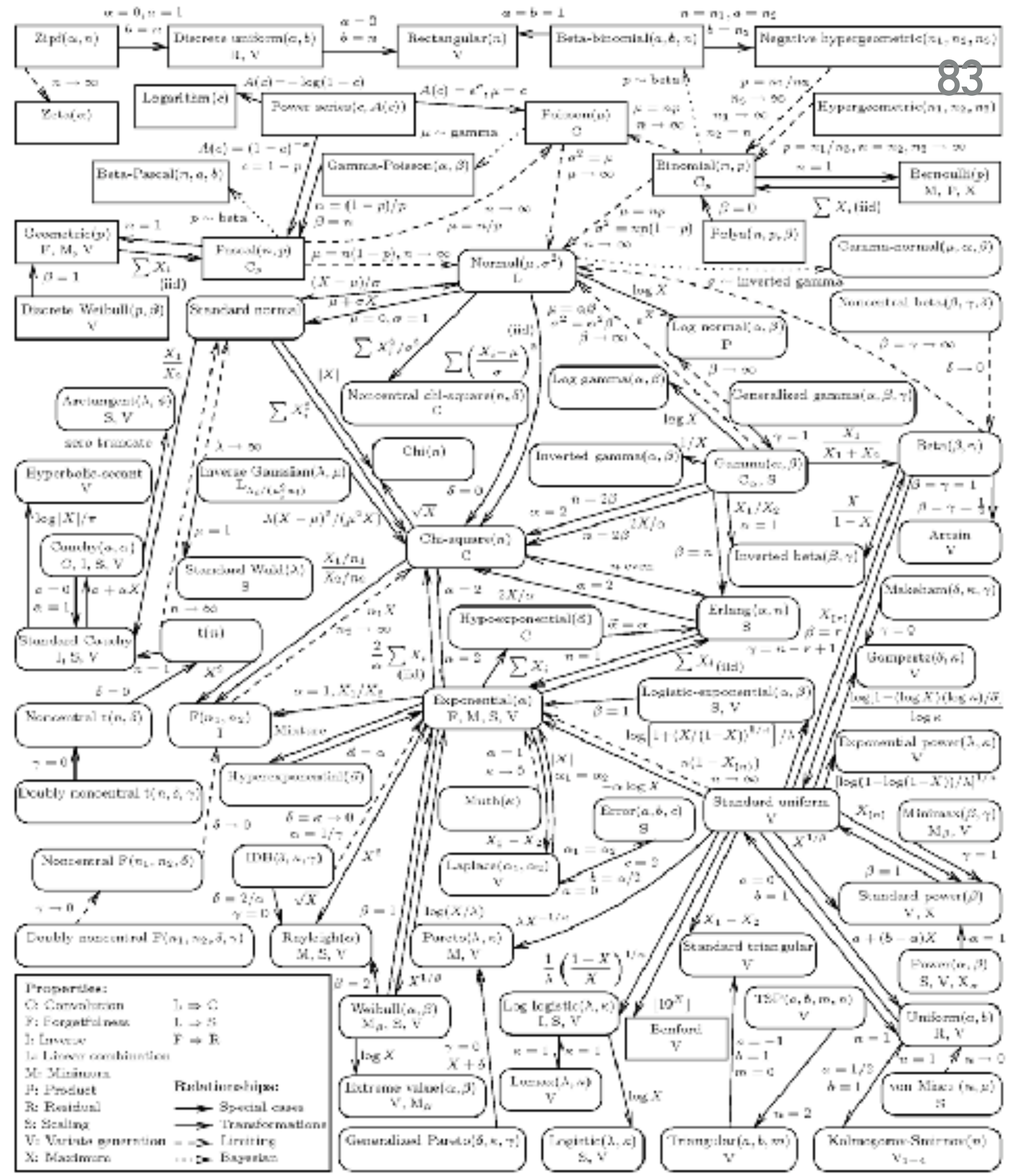
$$f(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\,\Gamma(n/2)}\left(1+\frac{x^2}{n}\right)^{-(n+1)/2}$$

Note $t_\infty = \mathcal{N}(0,1)$



mean: 0 $n>1$
NaN $n=0,1$
variance: $n/(n-2)$ $n>2$
∞ $1<n\leq 2$
otherwise NaN
skewness: 0 $n>3$
kurtosis: $6/(n-4)$ $n>4$

Lawrence M Leemis & Jacquelyn T McQueston (2008) "Univariate Distribution Relationships", The American Statistician, 62:1, 45-53, DOI: [10.1198/000313008X270448](https://doi.org/10.1198/000313008X270448)



Chapter 2 of the textbook has a bunch of other distributions you'll encounter frequently in astrophysics (and other domains) - definitely skim this

Mean and Variance of Distributions

Distribution	Mean	Variance
Binomial	Mp	Mp(1-p)
Poisson	μ	μ
Gaussian	μ	σ
Uniform [a,b)	(a+b)/2	(b-a)/12

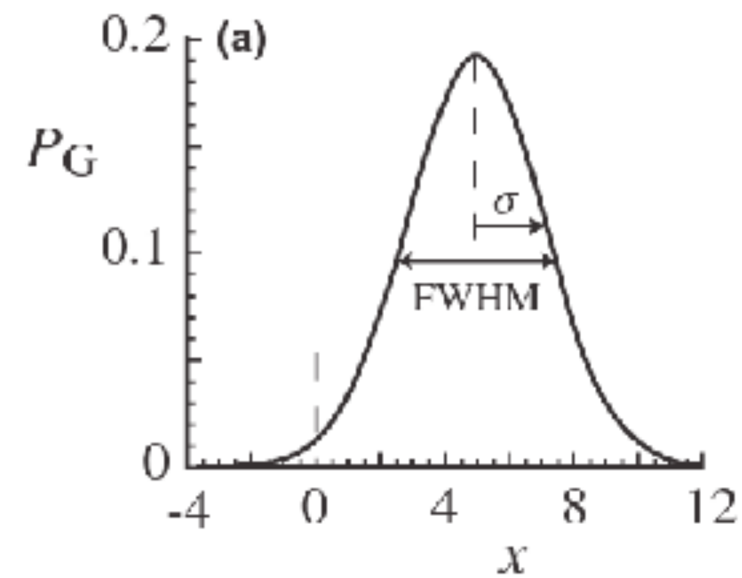
18

Recap

Standard deviation

$$\sigma = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(x_i - \mu)^2}$$

Gaussian distribution
for continuous values

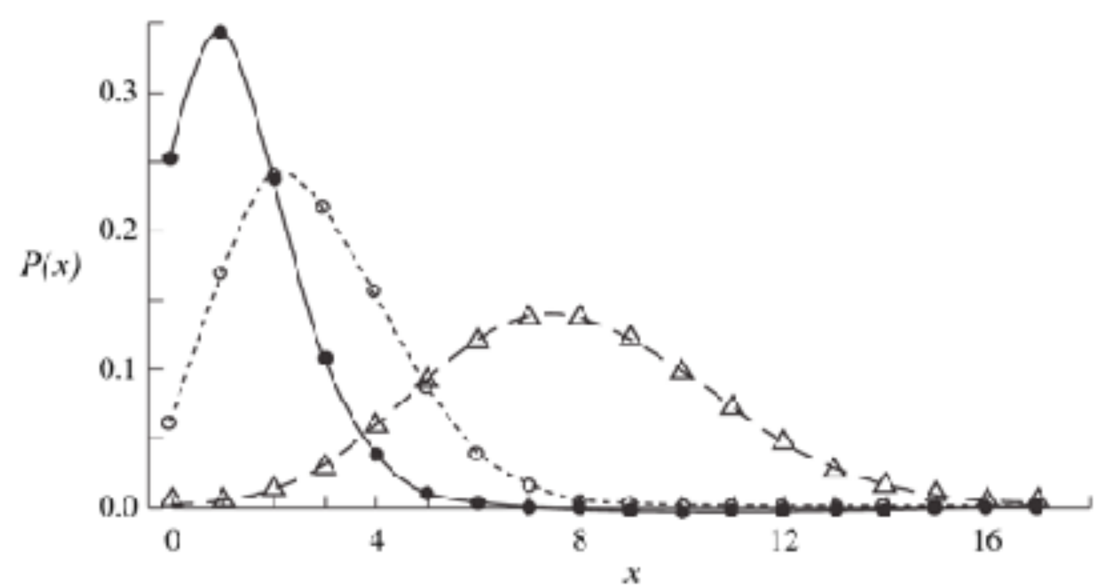


$$P_G(x, \mu, \sigma)\mathrm{d}x = \frac{\mathrm{d}x}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

$$\mathrm{FWHM}_{\mathrm{Gaussian}} = 2.354\sigma$$

Poisson distribution
for discrete events

$$P_p(x, \mu) = \frac{\mu^x}{x!}\mathrm{e}^{-\mu}$$



measurements

Imagine we want to check the length of rod.
The rod is supposed to be 6 mm long.

We measure:

$$L = 6.131826 \text{ mm}$$

measurements

Imagine we want to check the length of rod.
The rod is supposed to be 6 mm long.

We measure:

$$L = 6.131826 \text{ mm}$$

Is the rod exactly 6 mm long ?

measurements

Imagine we want to check the length of rod.
The rod is supposed to be 6 mm long.

We measure:

L = 6.131826 mm

Is the rod exactly 6 mm long ?

Let’s measure it some more

measurements

Imagine we want to check the length of rod.
The rod is supposed to be 6 mm long.

We measure:

$$L = 6.131826 \text{ mm}$$

Is the rod exactly 6 mm long ?

Let’s measure it some more

6.23149
5.78408
5.93843
6.07046
5.98985
5.87304
5.95400
6.03581
6.16218
5.87393

mean = 5.96638
median = 5.96715
stddev = 0.104616

6.23149
5.78408
5.93843
6.07046
5.98985
5.87304
5.95400
6.03581
6.16218
5.87393

mean = 5.96638
median = 5.96715
stddev = 0.104616

length of rod
L = 6.0 +/- 0.1 mm

significant digits !

measurements

Low let's say I have many 6 mm rods.
I want to measure the spread in lengths to check for uniformity.

How can I do this?

6.13579
5.71718
6.28263
5.74564
5.92986
5.98274
5.94347
6.13205
6.18098
5.13283

mean = 5.91832
median = 5.98274
stddev = 0.331139

6.13579
5.71718
6.28263
5.74564
5.92986
5.98274
5.94347
6.13205
6.18098
5.13283

mean = 5.9
median = 6.0
stddev = 0.3

one of these is an outlier ?

redo it

6.13579
5.71718
6.28263
5.74564
5.92986
5.98274
5.94347
6.13205
6.18098
6.20577

mean = 6.0
median = 6.0
stddev = 0.2

are we done?

6.13579
5.71718
6.28263
5.74564
5.92986
5.98274
5.94347
6.13205
6.18098
6.20577

mean = 6.0
median = 6.0
stddev = 0.2

are we done?

We know from our previous measurement of a single rod that we have a measurement error of 0.1mm

6.13579
5.71718
6.28263
5.74564
5.92986
5.98274
5.94347
6.13205
6.18098
6.20577

mean = 6.0
median = 6.0
stddev = 0.2

are we done?

We know from our previous measurement of a single rod that we have a measurement error of 0.1mm

so spread of rod values is :

$$\sigma = \sqrt{(\sigma_{sample})^2 - (\sigma_{measurement})^2}$$

$$\sigma = \sqrt{(0.2)^2 - (0.1)^2} = 0.17 \text{ mm}$$

$\sqrt{n}$

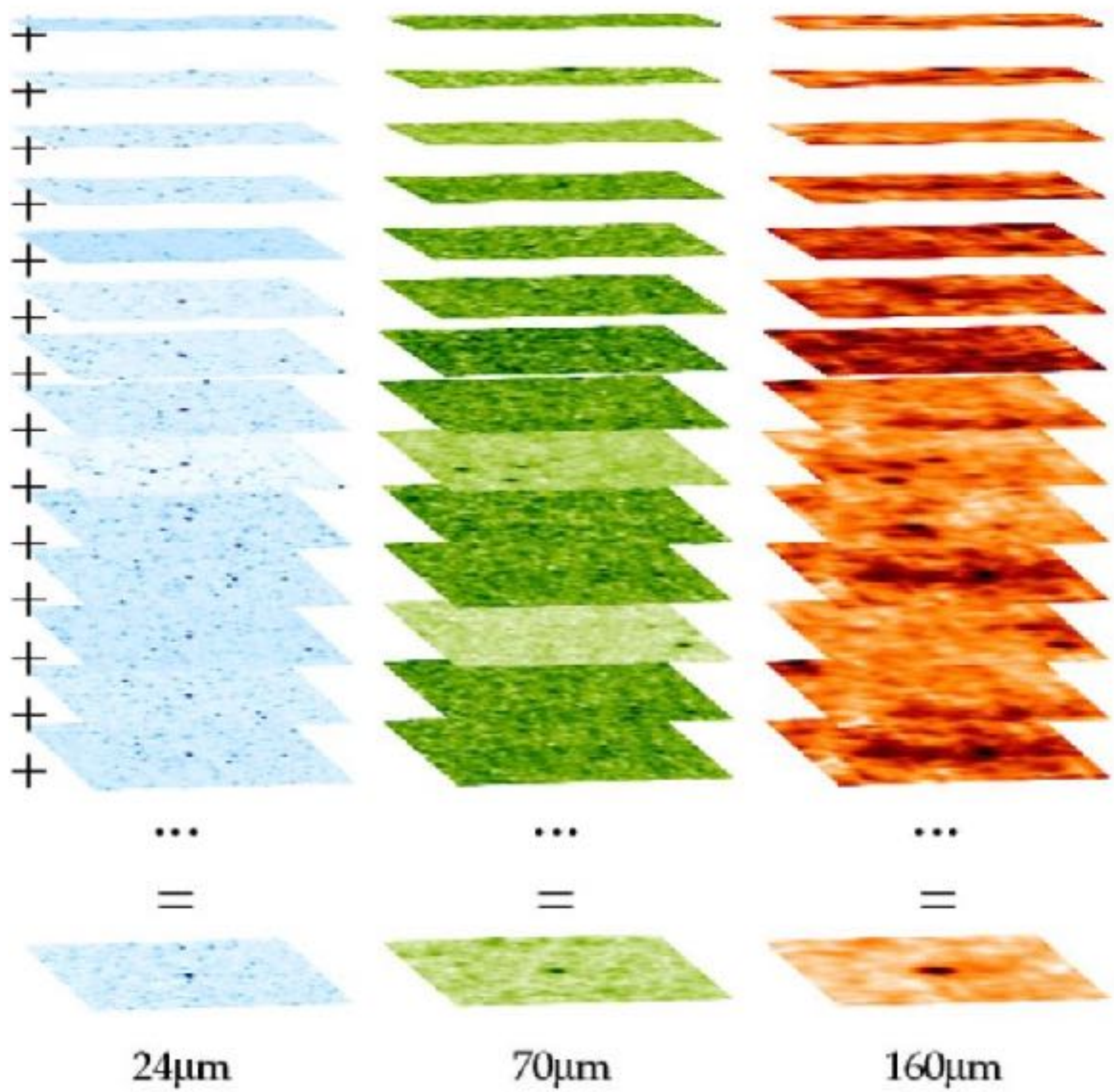
- For typical errors in a counting experiment, error is $\sqrt{n}$
- central limit theorem says your measurement error will average down as $\sqrt{n}$
- Most detectors in astronomy are photon counters. So errors go down with $\sqrt{n}$. But rate is average, so we can say $\sqrt{\boldsymbol{\mu}\Delta t}$ and so it does down as $\sqrt{t}$

$\sqrt{n}$

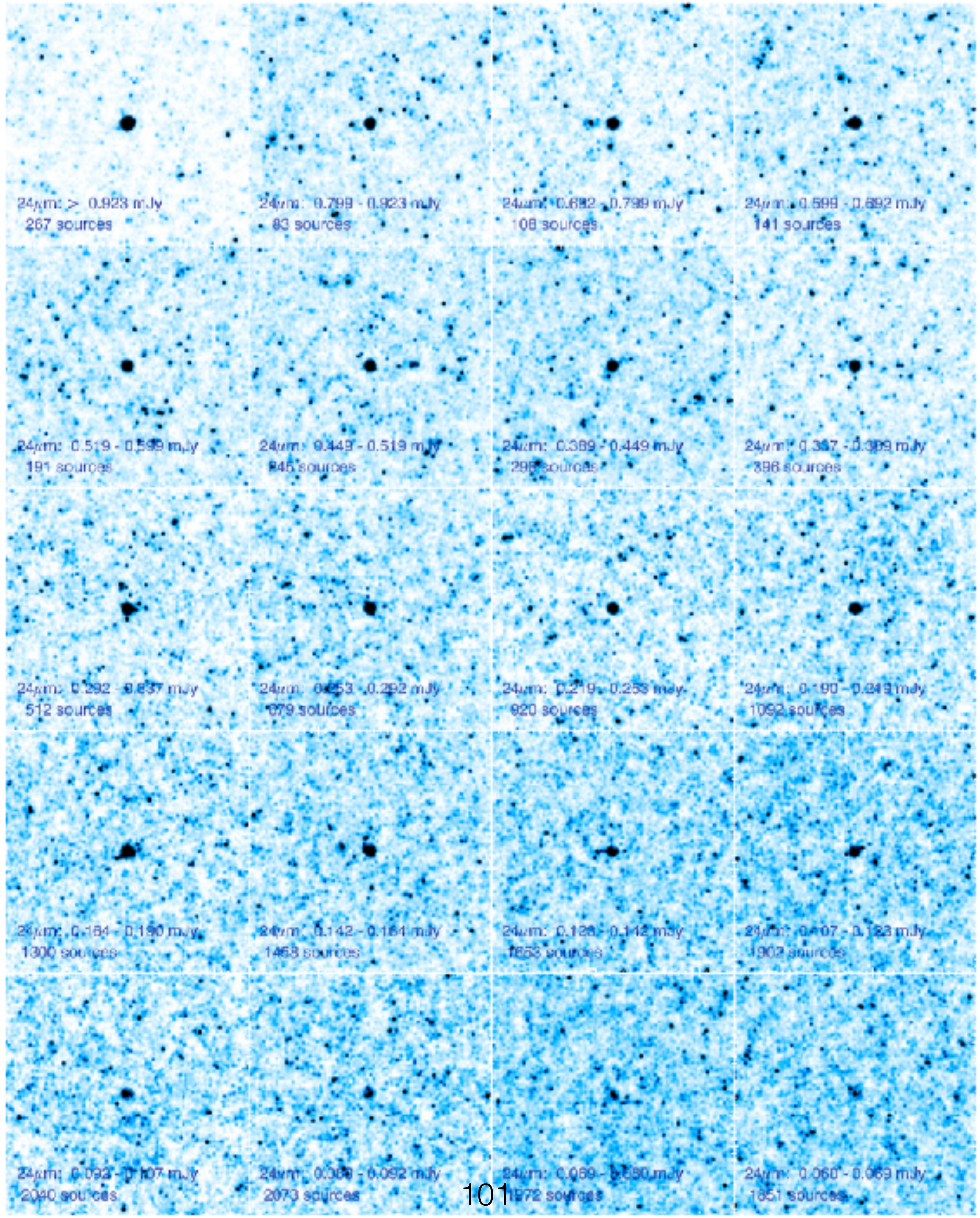
Q: Say you measure the flux of a star. You measure the error. Now say you want the error to go down by 1/2. How much longer do you need to integrate?

A. 0.5x
B. 2x
C. 3x
D. 4x

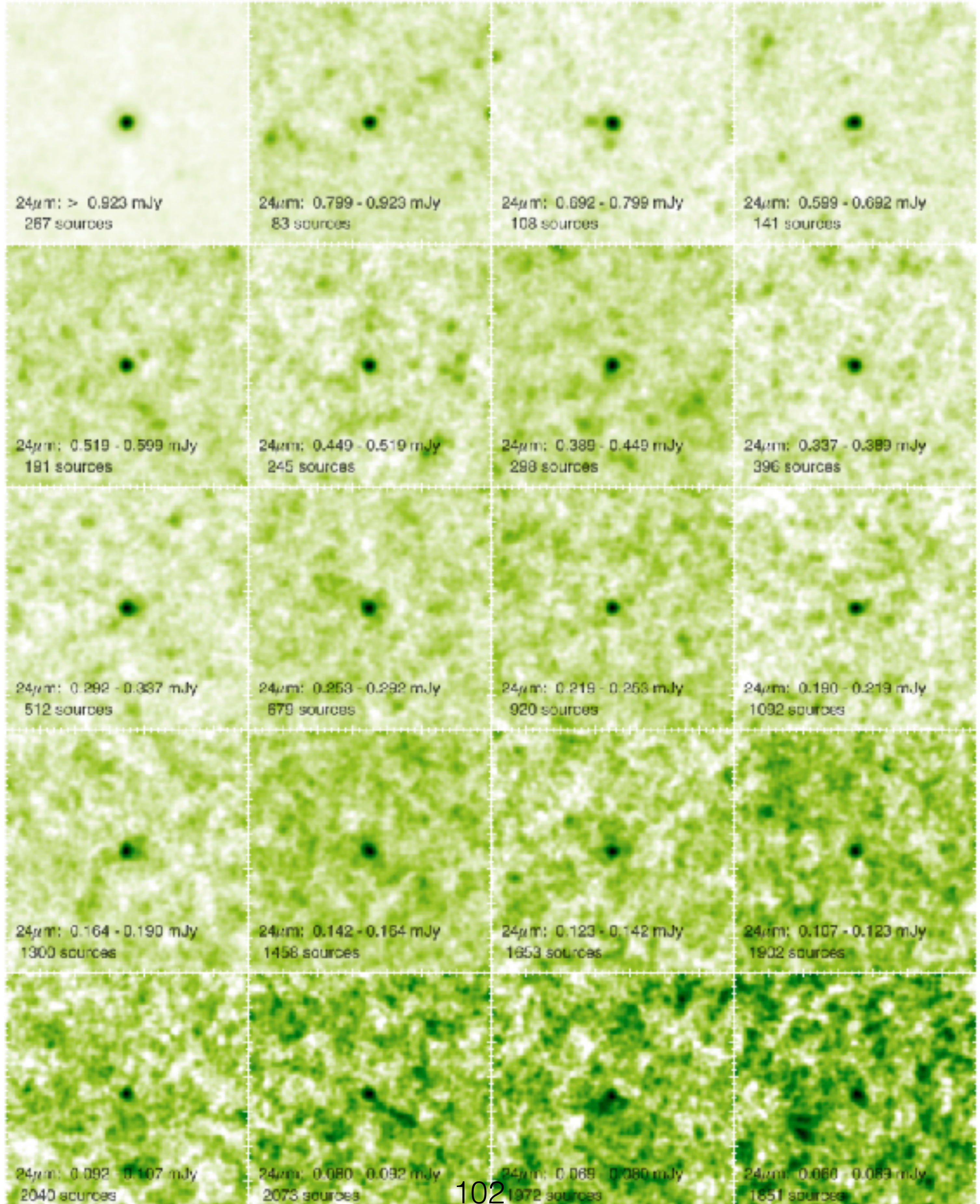
Example: stacking



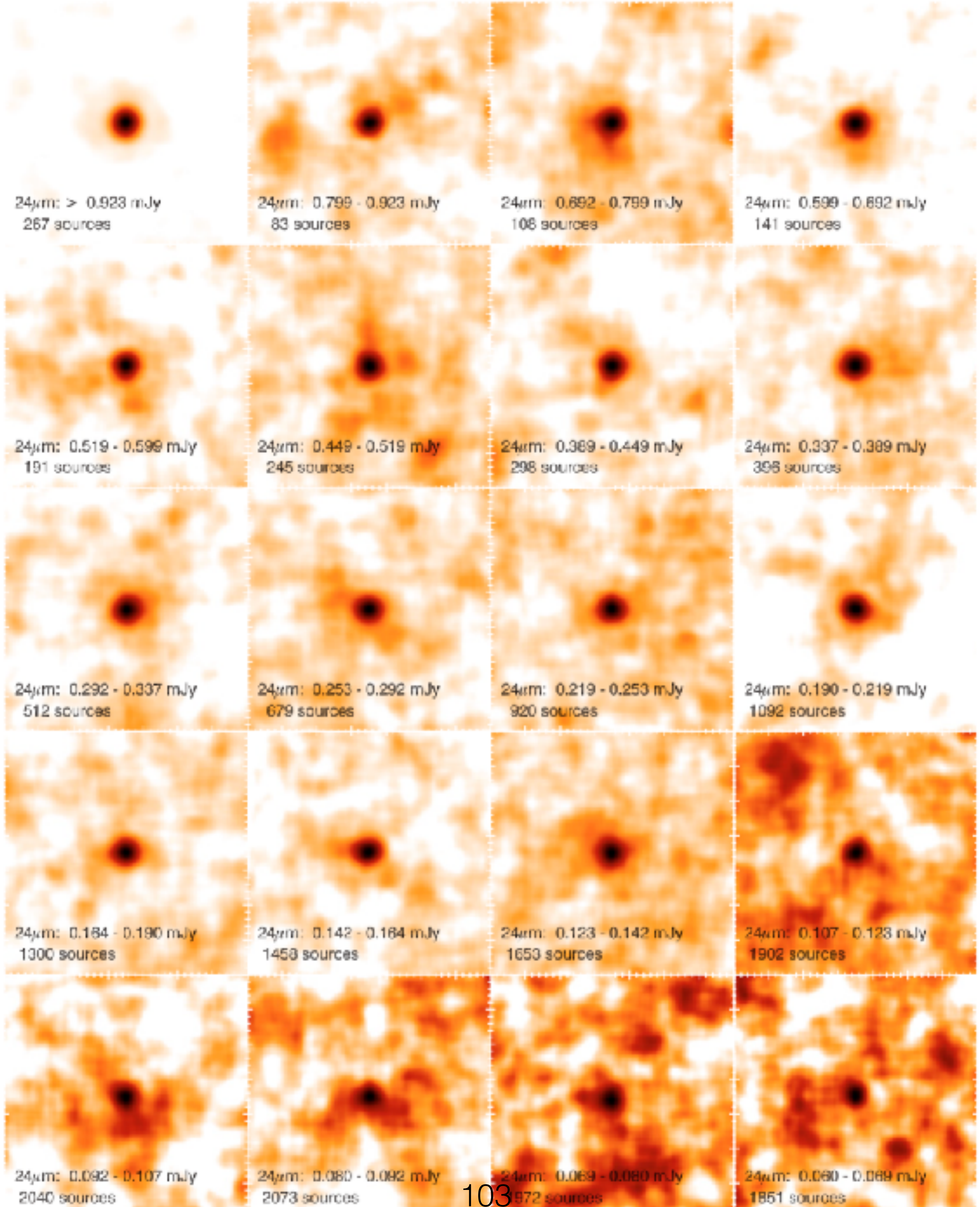
MIPS 24 μm Stacked Images

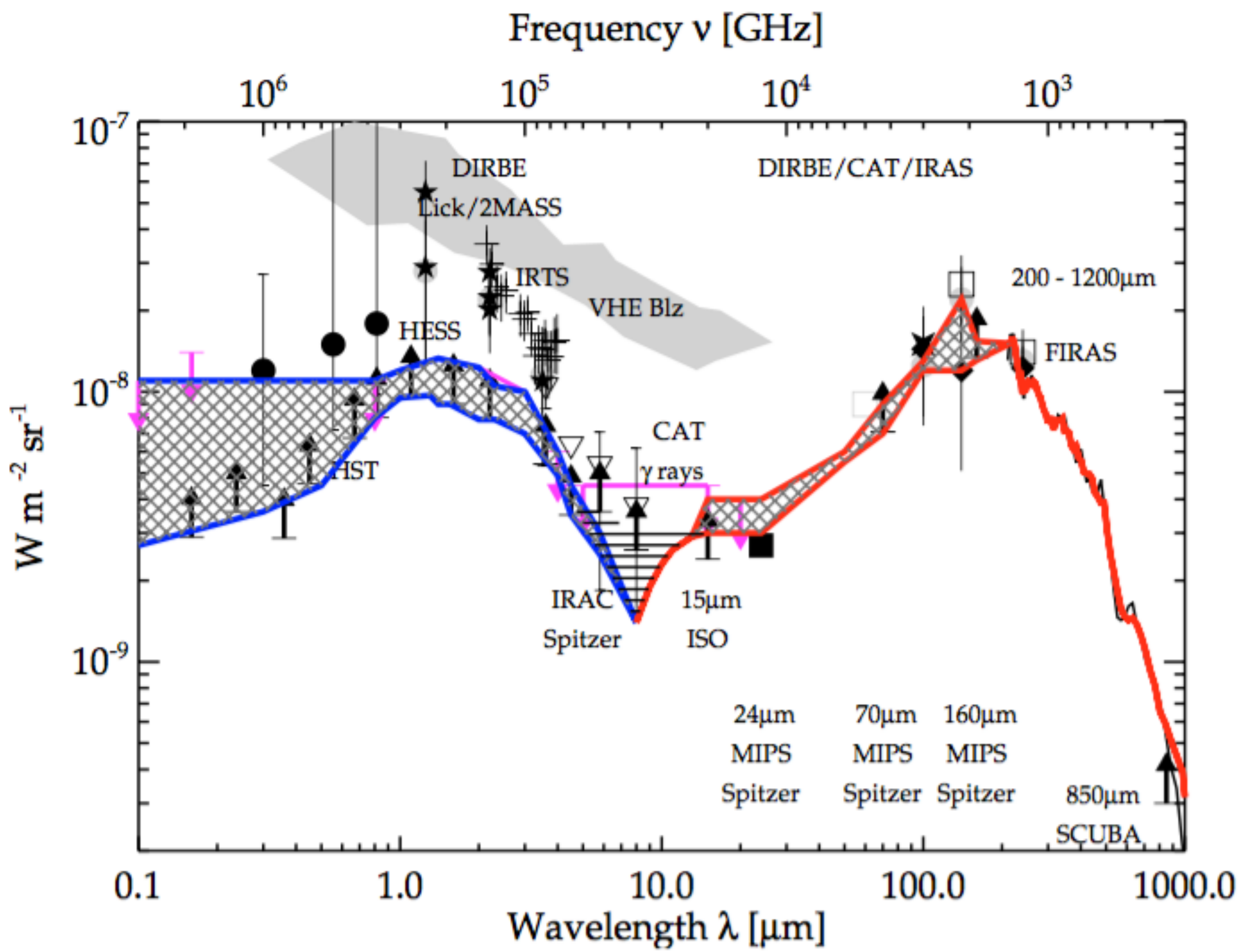


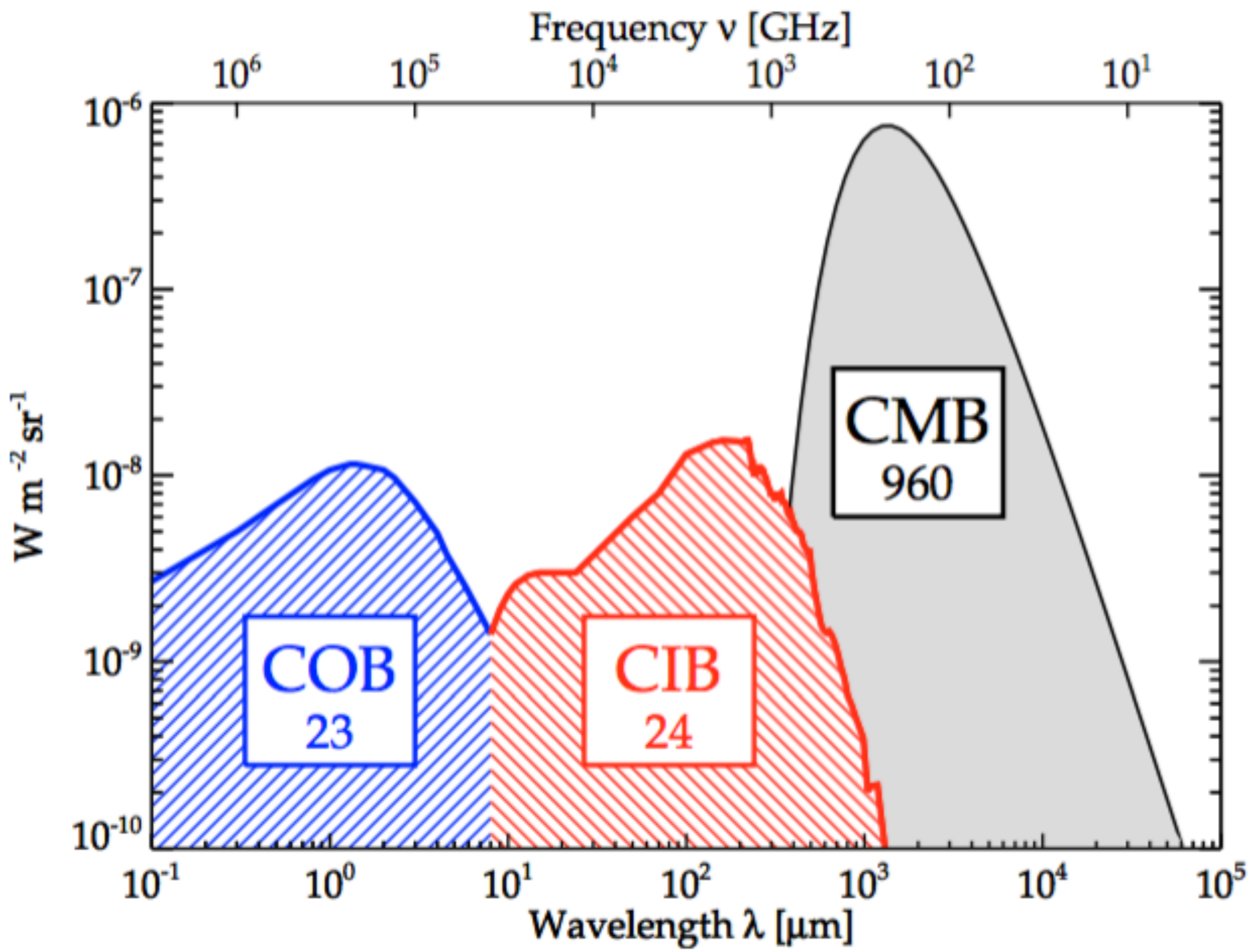
MIPS 70 μm Stacked Images



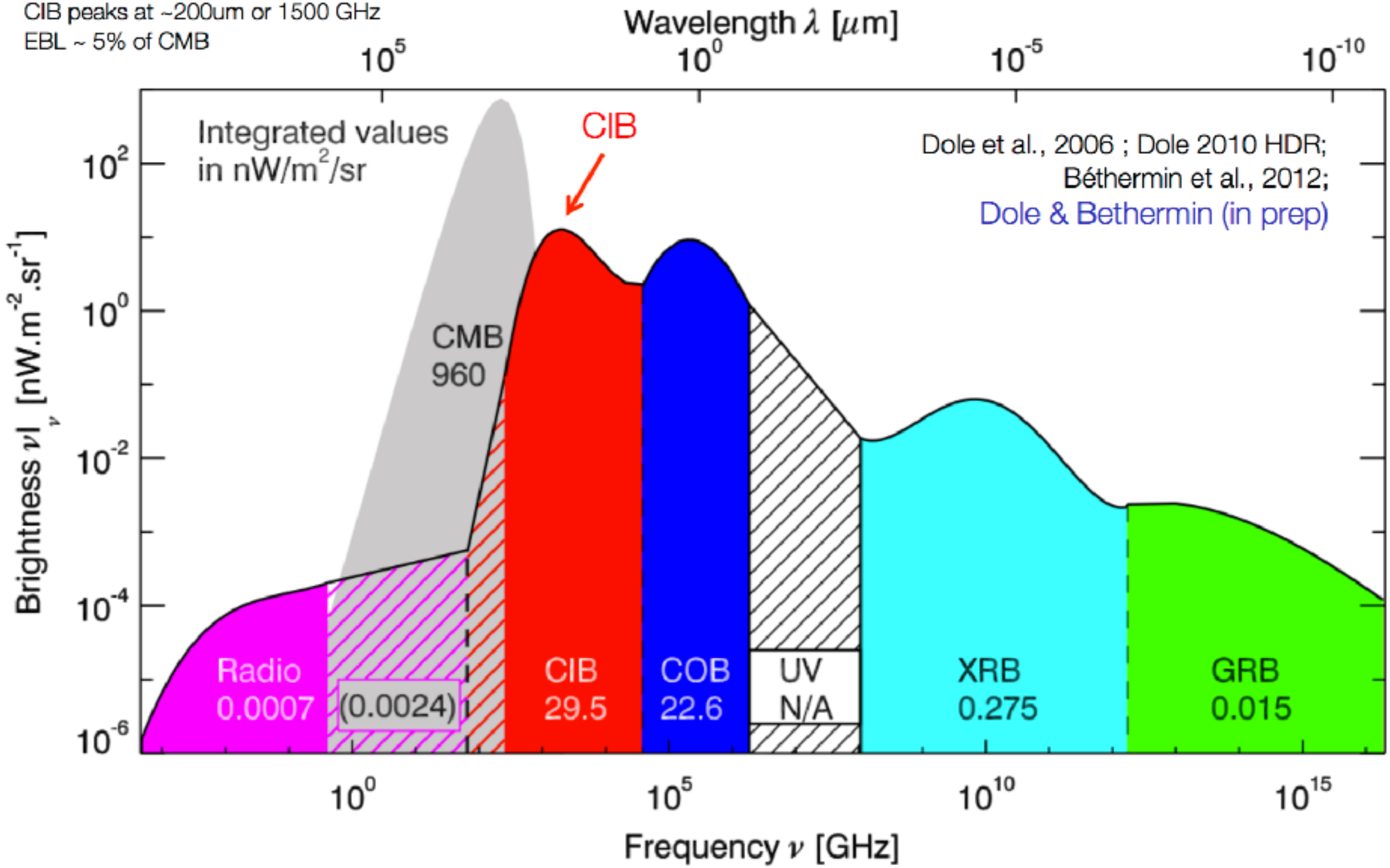
MIPS 160 μm Stacked Images




Frequency ν [GHz]
10^6
10^5
10^4
10^3
10^-7
10^-8
10^-9
W m^-2 sr^-1
DIRBE
Lick/2MASS
IRTS
VHE Blz
HESS
HST
CAT
γ rays
IRAC
Spitzer
15μm
ISO
DIRBE/CAT/IRAS
200 - 1200μm
FIRAS
24μm
MIPS
Spitzer
70μm
MIPS
Spitzer
160μm
MIPS
Spitzer
850μm
SCUBA
0.1
1.0
10.0
100.0
1000.0
Wavelength λ [μm]


Frequency ν [GHz]
10^6
10^5
10^4
10^3
10^2
10^1
10^-6
10^-7
10^-8
10^-9
10^-10
W m^-2 sr^-1
CMB
960
COB
23
CIB
24
10^-1
10^0
10^1
10^2
10^3
10^4
10^5
Wavelength λ [μm]

CIB > COB
CIB peaks at ~200um or 1500 GHz
EBL ~ 5% of CMB



Power Law Distribution

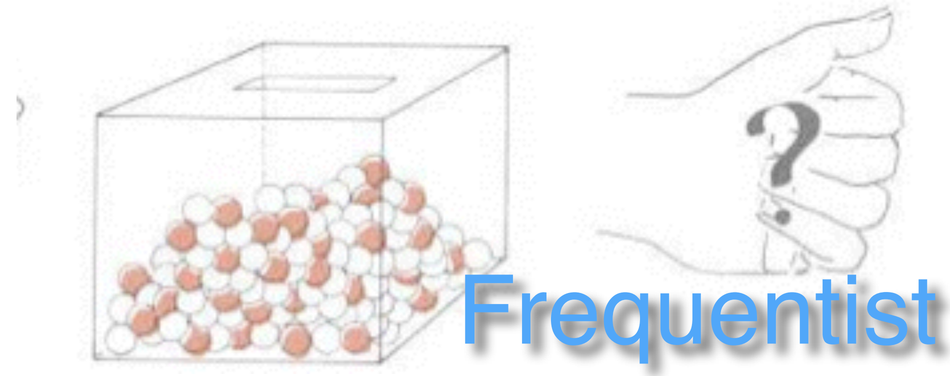
Integral Form: $$N(> L) = KL^{\gamma+1}$$

Differential Form: $$dN = (\gamma + 1)KL^{\gamma}dL$$

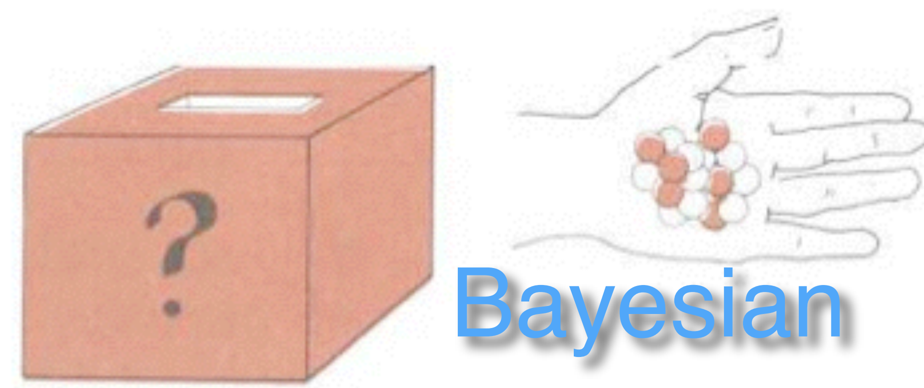
- Common distribution in astronomy:
  - Stars brighter than magnitude m on the sky.
    - What is mean magnitude of a star in the night sky?
    - What is the standard deviation?
- Problem lies in it not being a true probability distribution.
  - If limits are defined, the mean and variance can be derived.
  - This is a good area where simulation is useful.

Two approaches to discussing the problem:

- Knowing the distribution allows us to predict what we will observe.



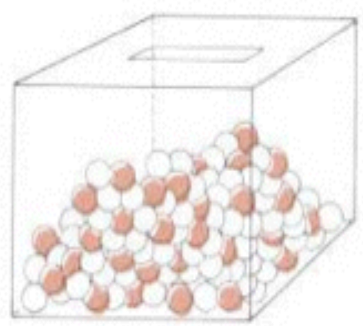
- We often know what we have observed and want to determine what that tells us about the distribution.



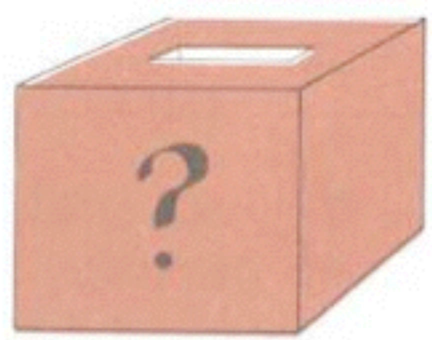
20


WHAT CAN THESE FISH...
...TELL US ABOUT THOSE FISH?

Frequentist Approach

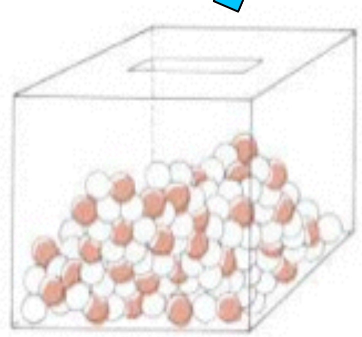


“I hypothesize that there are an equal number of red and white balls in a box.”



“I see I have drawn 6 red balls out of 10 total trials.”

Prediction

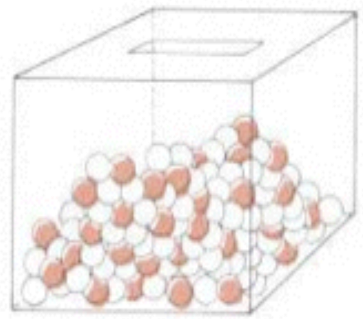


“A box with equal number of balls will have a mean of 5 red balls with a standard deviation of 1.6.”

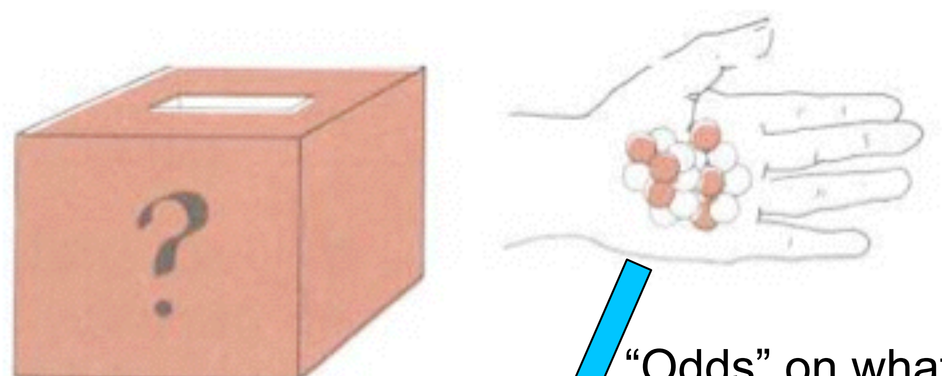
“Based on this I cannot reject my original hypothesis.”

21

Bayesian Approach

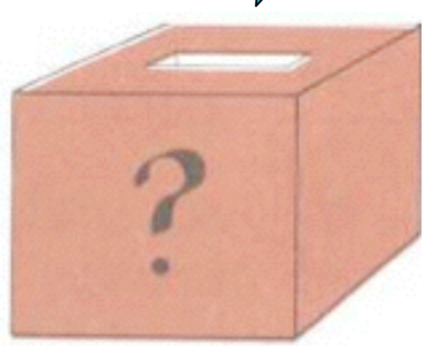


"I hypothesize that there are an equal number of red and white balls in a box."



"I see I have drawn 6 red balls out of 10 total trials."

"Odds" on what is in the box.



"There is a 24% chance that my hypothesis is correct."

22

Approaches to Statistics

- “Frequentist” approaches will calculate statistics that a given distribution would have produced, and confirms or rejects a hypothesis.
  - These are computationally easy, but often solve the inverse of the problem we want.
  - Locked into a distribution (typically Gaussian)
- Bayesian approaches use both the data and any “prior” information to develop a “posterior” distribution.

  Allows calculation of parameter uncertainty more directly.

  More easily incorporates outside information.

23

Conditional Probability

If two events, A and B, are related, then if we know B the probability of A happening is:

$$P(A|B) = \frac{P(A\,and\,B)}{P(B)}$$

Reversing the events, we get:

$$P(B|A) = \frac{P(B\,and\,A)}{P(A)}$$

P(B | A) should be read as "probability of B given A"

Now, P(A and B) = P (B and A) which gives us the important equality:

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}$$

This is Bayes' Formula.

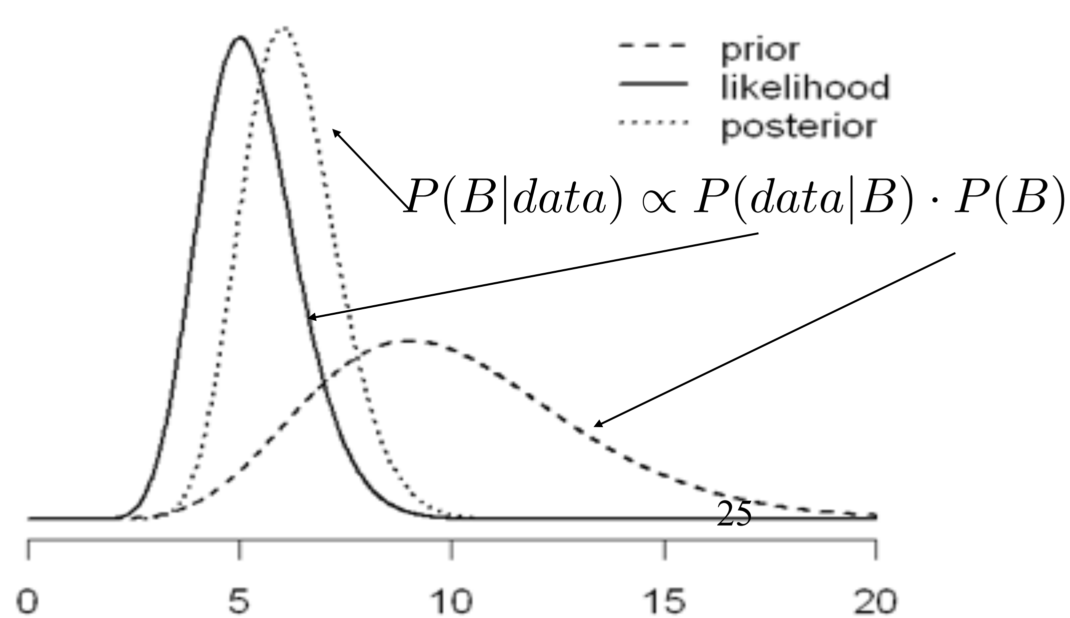
Bayes' Theorem

- Bayes' formula is used to merge data with prior information.

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}$$

  - A is typically the data, B the statistic we want to know.
  - P(B) is the "prior" information we may know about the experiment.

  - P(data) is just a normalization constant

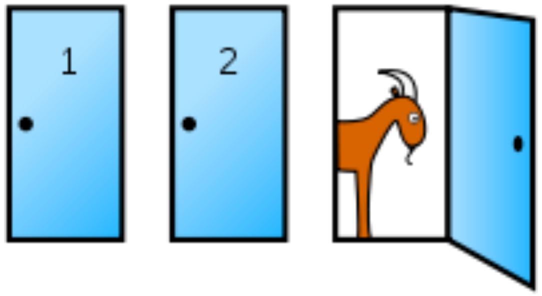


posterior probability ∝ Likelihood x prior probability

Example of Bayes' Theorem

A game show host invites you to choose one of three doors for a chance to win a car (behind one) or a goat (behind the other two).

After you choose a door (say, door 1), the host opens another door (say, door 3) to reveal a goat. Should you switch your choice?



26

Using Bayes' theorem

- Assume we are looking for faint companions, and expect them to be around 1% of the stars we observe.
- From putting in fake companions we know that we can detect objects in the data 90% of the time.
- From the same tests, we know that we see “false” planets 3% of the observations.
- What is the probability that an object we see is actually a planet?

$$P(planet) = 0.01$$
$$P(noplanet) = 0.99$$

$$P(+det.|planet) = 0.9$$
$$P(-det.|planet) = 0.1$$

$$P(+det.|noplanet) = 0.03$$

$$P(planet| + det.) = \frac{P(+det|planet)P(planet)}{P(+det)}$$

$$P(+det.) = P(+det|planet)P(planet) + P(+det|noplanet)P(noplanet)$$

$$P(planet| + det.) = \frac{0.9 \cdot 0.01}{0.9 \cdot 0.01 + 0.03 \cdot 0.99} = 0.23$$

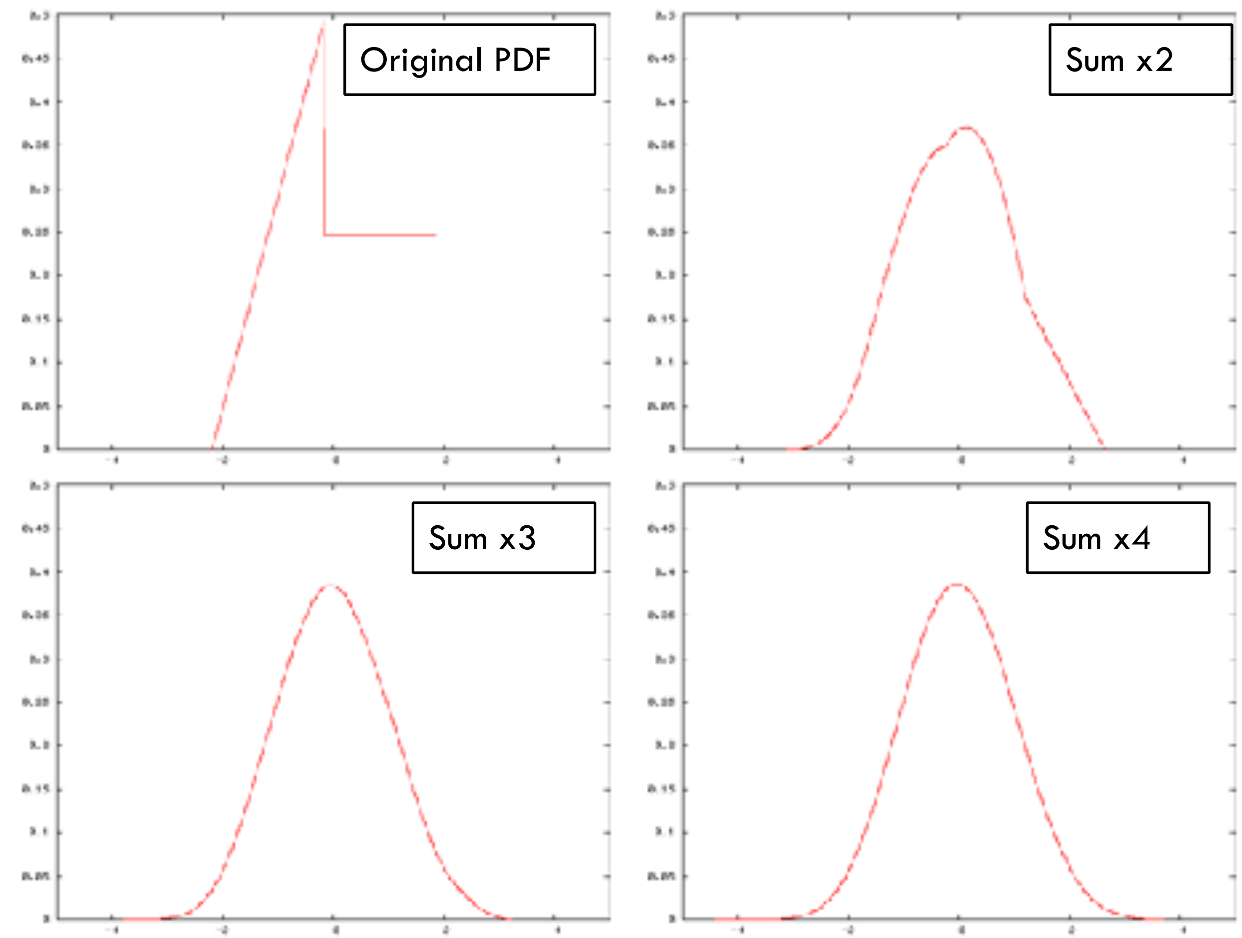
27

General Bayesian Guidance

- Focuses on probability rather than accept/reject.
- Bayesian approaches allow you to calculate probabilities the parameters have a range of values in a more straightforward way.
- A common concern about Bayesian statistics is that it is subjective. This is not necessarily a problem.
- Bayesian techniques are generally more computationally intensive, but this is rarely a drawback for modern computers.

Why are Gaussian statistics so pervasive? —> Central Limit Theorem

The distribution of an average tends to be Normal, even when the distribution from which the average is computed is decidedly non-Normal.



If x is the average of n independent random variables of expected value μ and variance σ^2, then:

(All scaled to have μ=1, σ=1; CC)

$$\text{As } n \to \infty \quad p(x) \to \frac{1}{\sqrt{2\pi \frac{\sigma^2}{n}}} e^{-\frac{(x-\mu)^2}{2\frac{\sigma^2}{n}}}$$

Least Squares Fitting

Assume that you have a set of data, and wish to compare it with a model.

Two basic questions:
(1)What are the best fit model parameters?
(2)Is the model a good fit to the data?

A standard approach to the first part is "least-squares" fitting. The basic idea is that for a function f(x), the parameters that minimize the square of the difference between the model and each data point will be the best model parameters. (Squaring leads to positive and negative deviations being treated equally.)

The equation is:

$$S = \sum_{i=1}^{N} (y_i - f(x_i))^2$$

With the best solution coming from the parameters that minimize S. If you are fitting a line,
f(x)=ax+b, then this is called linear least-squares fitting. The method though is general.

Chi-Squared Fitting

Chi-squared minimization, or chi-squared fitting, is a refinement of the least squares method that takes into account the error bar associated with each data point. This additional information lets you answer both questions:

(1) What are the best fit model parameters?

(2) Is the model a good fit to the data?

If each data point is now represented by 3 numbers (x_i, y_i, σ_i), where σ_i is the error bar, then the quantity to be minimized is:

$$\chi^2 = \sum_{i=1}^{N} \left(\frac{y_i - y}{\sigma_i} \right)^2$$

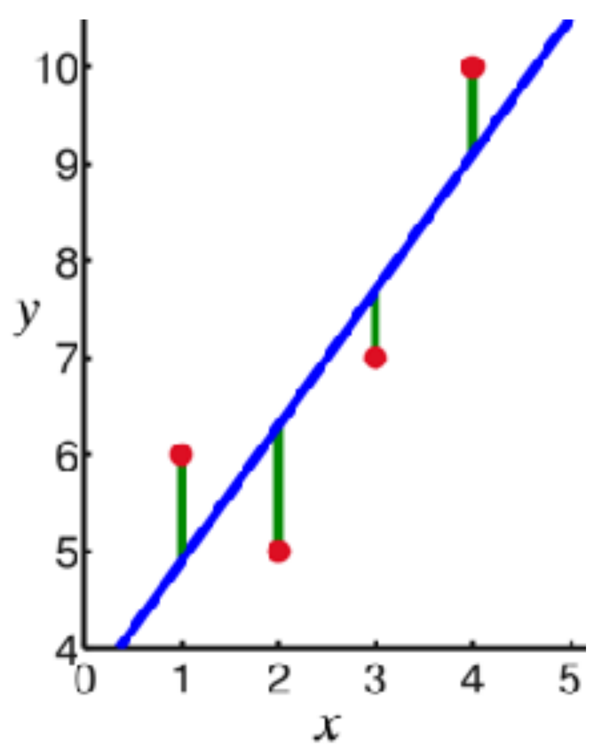
Note that now what you are fundamentally doing is taking the ratio of the scatter between the data and model to the error bar and summing this for all the data points. If the model is a good fit to the data, then these should on average be equal to ~1.

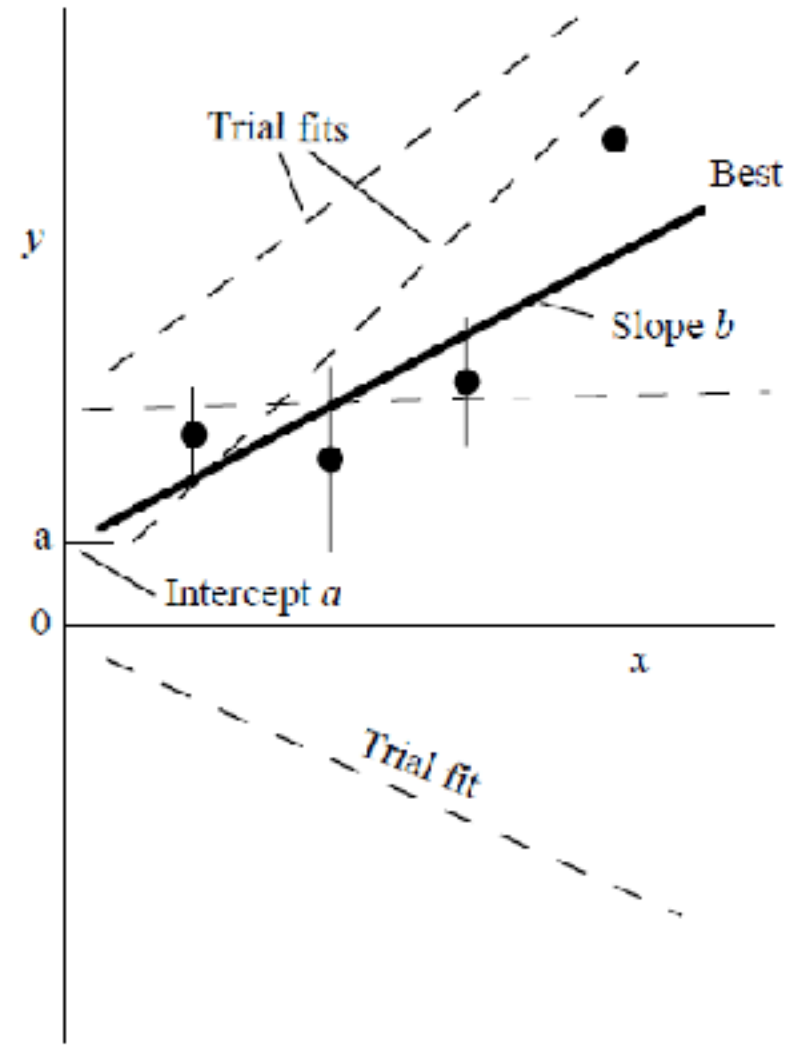
Least-squares minimization

- Least-squares minimization wrt ***Q:***

$$\chi^2 = \sum_t \frac{(x_t - f(t,Q))^2}{2\sigma^2}$$

- Best fit minimizes ***χ²***

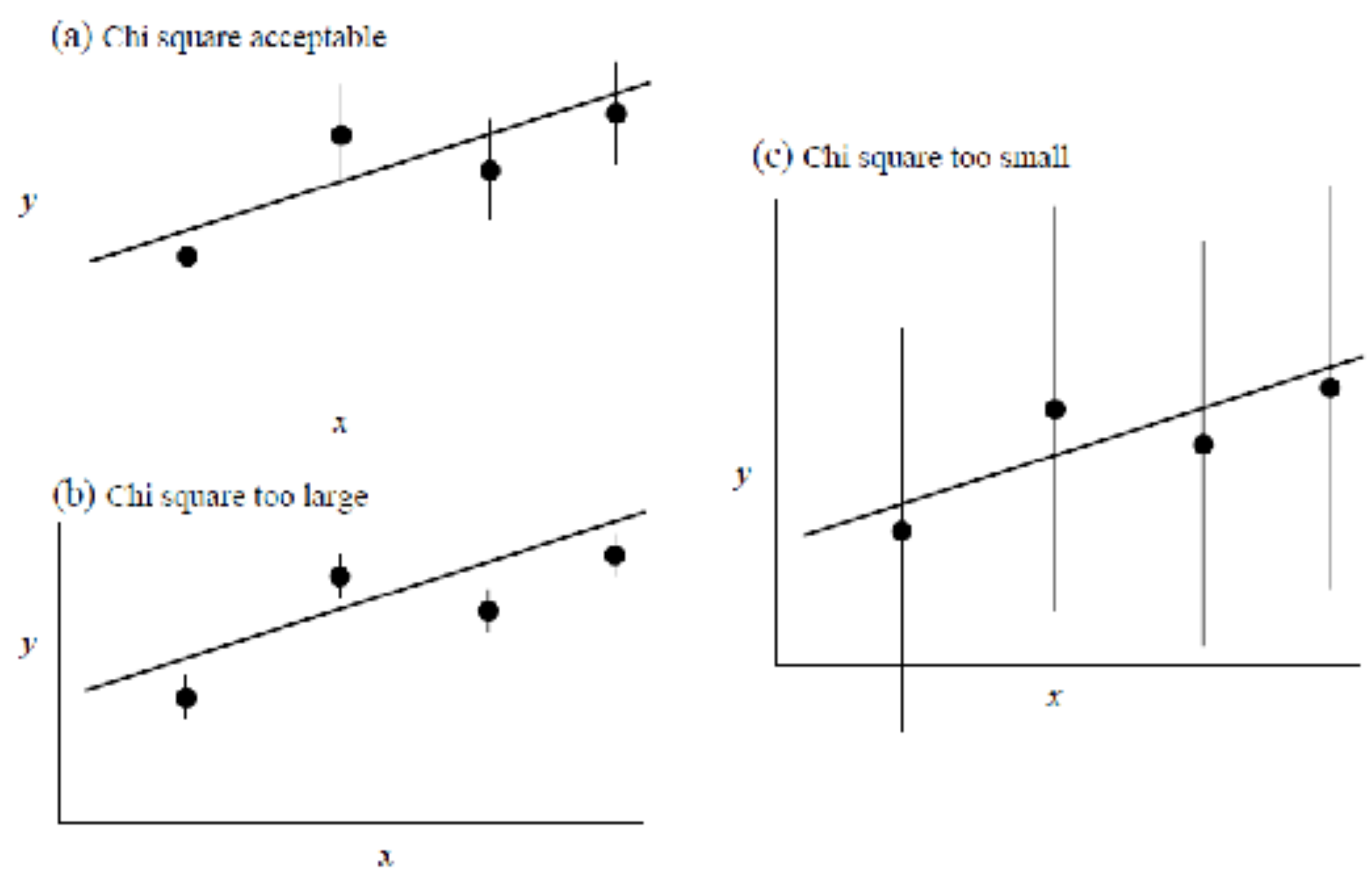




Chi-squared statistic

- How to assess the quality of the fit ?
- Useful to ask the question:

  "If the data model function is the true function, and I made another similar set of measurements, what is the probability that I would measure a greater χ^2 than the current value ?"



Chi-Squared Fitting

Quantitatively, the way that you answer the question of whether the model is a good fit to the data is by computing a quantity called the “reduced” chi-squared,

$$\chi^2_\nu = \frac{\chi^2}{N - M - 1} = \frac{\sum_{i=1}^{N}(y_i - f(x_i))^2}{N - M - 1}$$

where N is the number of data points and M is the number of tunable parameters in your function.

Example: f(x) = a*x+b , and you are fitting 10 data points which each have (x_i, y_i, σ_i)

In this case N=10, M=2.

If $\chi^2_\nu \sim 1$ → The model is a good fit to the data.

If $\chi^2_\nu >> 1$ → The model is a bad fit to the data, or your errors are underestimated.

If $\chi^2_\nu << 1$ → Your errors are overestimated.

Least-squares: caveats

- Least-square methods are very sensitive to non-Gaussian outliers

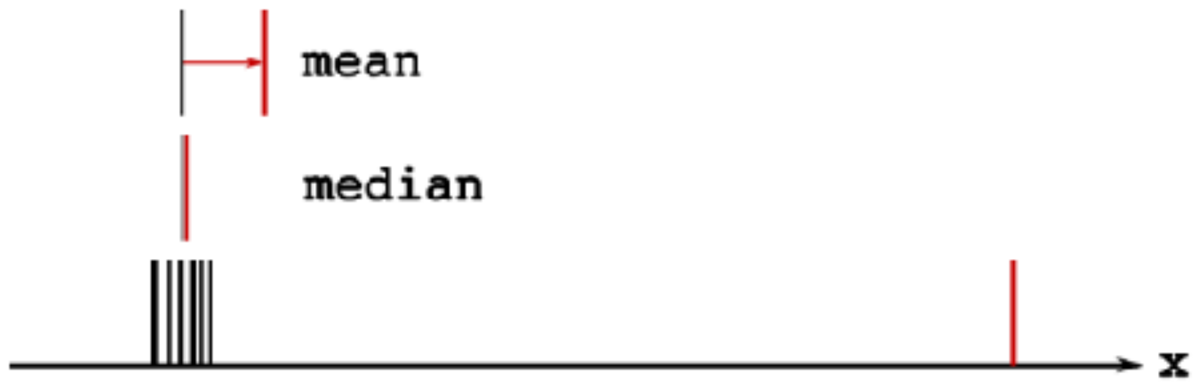


Fig. 6.9.— An outlier can greatly affect the average of a data set. The median is a more robust estimator than the mean.

- More robust (non-parametric) measures of model deviation may help

$$\sum_i \left| \frac{x_i - \bar{x}}{\sigma_i} \right|$$

$$\sum_i \log\left(1 + \frac{1}{2}\left(\frac{x_i - \bar{x}}{\sigma_i}\right)^2\right),$$

The error propagation equation

$x = f(u,v)$

Consider a Taylor-series expansion about mean values $(\bar{u},\bar{v})$:

$$(x_i - \bar{x}) \simeq (u_i - \bar{u})\left(\frac{\partial x}{\partial u}\right) + (v_i - \bar{v})\left(\frac{\partial x}{\partial v}\right)$$

But,

$$\sigma_u^2 = \left\langle (u - \bar{u})^2 \right\rangle = \operatorname*{Lim}_{n \to \infty}\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right]$$

Therefore:

$$\sigma_x^2 = \sigma_u^2\left(\frac{\partial x}{\partial u}\right)^2 + \sigma_v^2\left(\frac{\partial x}{\partial v}\right)^2 + 2\sigma_{uv}\left(\frac{\partial x}{\partial u}\right)\left(\frac{\partial x}{\partial v}\right) + \ldots$$

Covariance and independence

- Covariance $\boldsymbol{\sigma_{uv}}$: measure of the correlation between fluctuations in ***u*** and ***v***:

$$\sigma_{uv} = \left\langle \left(u - \overline{u}\right)\left(v - \overline{v}\right) \right\rangle$$

- Covariance is zero for ***uncorrelated*** statistical errors:

$$\sigma_x^2 = \sigma_u^2 \left(\frac{\partial x}{\partial u}\right)^2 + \sigma_v^2 \left(\frac{\partial x}{\partial v}\right)^2 + 2\sigma_{uv} \left(\frac{\partial x}{\partial u}\right)\left(\frac{\partial x}{\partial v}\right) + \ldots$$

- We will make this assumption in what follows.

Covariance and independence

- Covariance $\boldsymbol{\sigma_{uv}}$: measure of the correlation between fluctuations in ***u*** and ***v***:

$$\sigma_{uv} = \left\langle (u - \bar{u})(v - \bar{v}) \right\rangle$$

- Covariance is zero for ***uncorrelated*** statistical errors:

$$\sigma_x^2 = \sigma_u^2 \left(\frac{\partial x}{\partial u} \right)^2 + \sigma_v^2 \left(\frac{\partial x}{\partial v} \right)^2 + \cancel{2\sigma_{uv} \left(\frac{\partial x}{\partial u} \right) \left(\frac{\partial x}{\partial v} \right)} + \ldots$$

- We will make this assumption in what follows.

$$\sigma_{uv} = 0$$

Error propagation: sums and differences

- Consider **x = *u* + *const.***

$$x = f(u) = u + a$$

$$\sigma_x^2 = \sigma_u^2 \left(\frac{\partial x}{\partial u}\right)^2 = \sigma_u^2$$

- Consider **x = *u* + v**

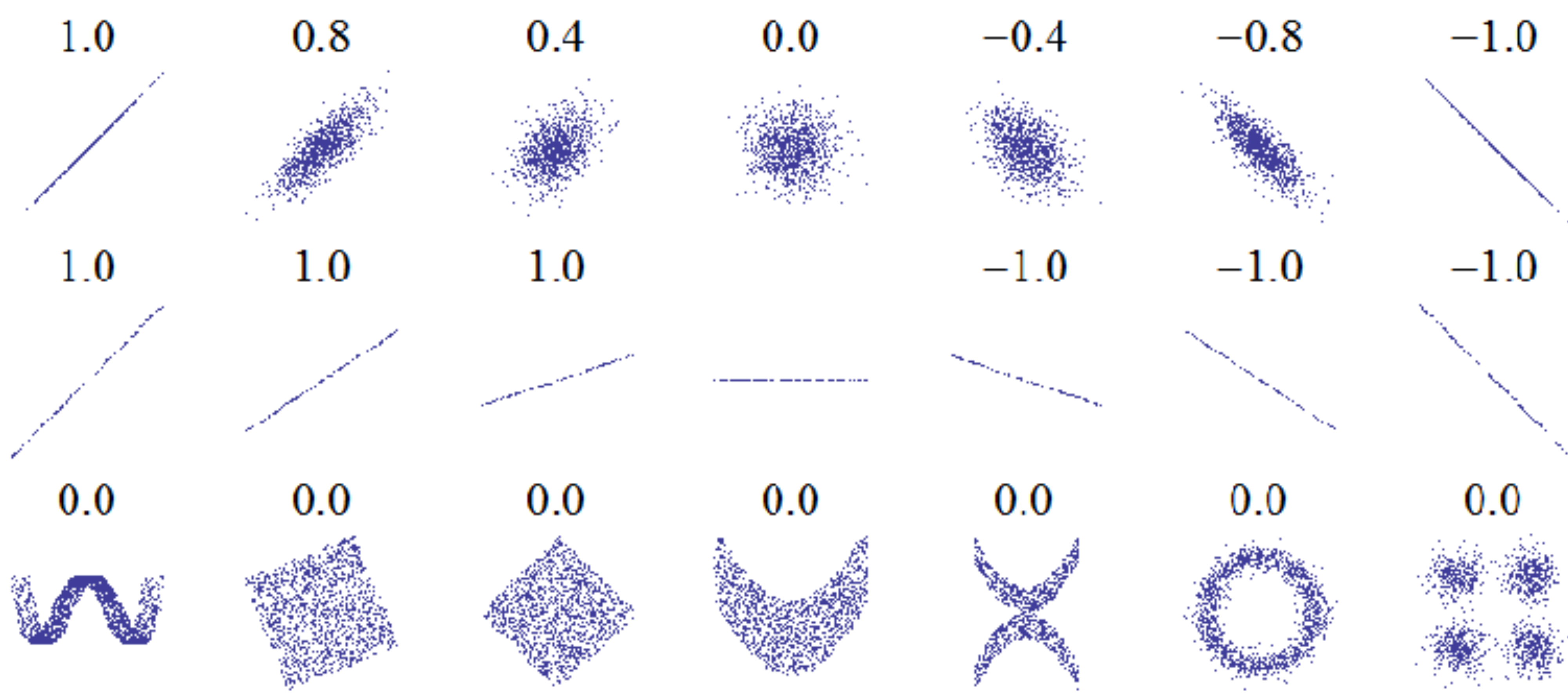
$$x = f(u,v)$$
$$= u + v$$

$$\sigma_x^2 = \sigma_u^2 \left(\frac{\partial x}{\partial u}\right)^2 + \sigma_v^2 \left(\frac{\partial x}{\partial v}\right)^2$$
$$= \sigma_u^2 + \sigma_v^2$$

The *quadrature* rule

$$\mathrm{Cov}(x, y) = E([x - \langle x \rangle])E([y - \langle y \rangle])$$

$$\mathrm{Cor}(x, y) = \frac{\mathrm{Cov}(x,y)}{\sqrt{\mathrm{Var}(x)\mathrm{Var}(y)}}$$



Some useful formulas

Function	Variance	Standard Deviation
$f = aA$	$\sigma_f^2 = a^2\sigma_A^2$	$\sigma_f = a\sigma_A$
$f = aA + bB$	$\sigma_f^2 = a^2\sigma_A^2 + b^2\sigma_B^2 + 2ab\,\mathrm{cov}_{AB}$	$\sigma_f = \sqrt{a^2\sigma_A^2 + b^2\sigma_B^2 + 2ab\,\mathrm{cov}_{AB}}$
$f = AB$	$\sigma_f^2 \approx f^2\left[\left(\frac{\sigma_A}{A}\right)^2 + \left(\frac{\sigma_B}{B}\right)^2 + 2\frac{\mathrm{cov}_{AB}}{AB}\right]$	$\sigma_f \approx \lvert f\rvert\sqrt{\left(\frac{\sigma_A}{A}\right)^2 + \left(\frac{\sigma_B}{B}\right)^2 + 2\frac{\mathrm{cov}_{AB}}{AB}}$
$f = \frac{A}{B}$	$\sigma_f^2 \approx f^2\left[\left(\frac{\sigma_A}{A}\right)^2 + \left(\frac{\sigma_B}{B}\right)^2 - 2\frac{\mathrm{cov}_{AB}}{AB}\right]$ [11]	$\sigma_f \approx \lvert f\rvert\sqrt{\left(\frac{\sigma_A}{A}\right)^2 + \left(\frac{\sigma_B}{B}\right)^2 - 2\frac{\mathrm{cov}_{AB}}{AB}}$
$f = aA^b$	$\sigma_f^2 \approx f^2\left(b\frac{\sigma_A}{A}\right)^2$ [12]	$\sigma_f \approx \lvert f\rvert\left(b\frac{\sigma_A}{A}\right)$
$f = a\ln(bA)$	$\sigma_f^2 \approx \left(a\frac{\sigma_A}{A}\right)^2$ [13]	$\sigma_f \approx \left\lvert a\frac{\sigma_A}{A}\right\rvert$
$f = a\log_{10}(A)$	$\sigma_f^2 \approx \left(a\frac{\sigma_A}{A\ln(10)}\right)^2$ [13]	$\sigma_f \approx \left\lvert a\frac{\sigma_A}{A\ln(10)}\right\rvert$
$f = ae^{bA}$	$\sigma_f^2 \approx f^2\,(b\sigma_A)^2$ [14]	$\sigma_f \approx \lvert f\,(b\sigma_A)\rvert$
$f = a^{bA}$	$\sigma_f^2 \approx f^2\,(b\ln(a)\sigma_A)^2$	$\sigma_f \approx \lvert f\,(b\ln(a)\sigma_A)\rvert$
$f = A^B$	$\sigma_f^2 \approx f^2\left[\left(\frac{B}{A}\sigma_A\right)^2 + (\ln(A)\sigma_B)^2 + 2\frac{B\ln(A)}{A}\mathrm{cov}_{AB}\right]$	$\sigma_f \approx \lvert f\rvert\sqrt{\left(\frac{B}{A}\sigma_A\right)^2 + (\ln(A)\sigma_B)^2 + 2\frac{B\ln(A)}{A}\mathrm{cov}_{AB}}$

From wikipedia.org

Best practice: statistical data analysis

- Understand your data model.
  - Propagation, calibration, and noise.
- Examine your implicit statistical assumptions.
  - E.g. all noise is Gaussian.
  - Errors are uncorrelated
- Always quote errors and uncertainties.
- Explain your error calculations.
  - Be conservative if errors are unknown.
- Keep science conclusions within the uncertainties.
  - Standard is observational honesty, and adherence to best community practice.
  - Absolute certainty isn’t possible – we proceed incrementally from one reproducible result to another.