ASTRONOMY 535

OBSERVATIONAL TECHNIQUES

CLASS NOTES

Fall 2023

Instructor: Jon Holtzman

1 Class introduction

- Class content
  - Goal: how to understand and optimally collect astronomical data and extract information, in principle and in practice. Position you to take advantage of NMSU facilities if you are interested.
  - What do we mean by observational techniques?
    - Recover intrinsic information coming from sky via measurement as best as possible: requires understanding of instrumental signatures and how to remove them, calibration of data
    - All measurements have uncertainties and, often, understanding uncertainty on a measurement is as important as the measurement itself.
    - In some cases, uncertainties can be minimized by how data are collected, and how data are analyzed/calibrated.
  - class content:
    - knowledge
      - need to understand nature of light and errors and error analysis involved with its collection
      - need to understand how instruments work to understand information, accuracy, and to get most from the data, not to mention being in a position to understand/design instruments of the future. In many cases, instrument development drives scientific knowledge! Consequently, we will discuss telescopes, instruments, and detectors. Principally we will discuss imaging instruments and spectrographs. We may briefly discuss some other types of instruments (imaging spectroscopy, interferometry). We will cover instrumentation/methodology for UV/optical/IR observations; radio may be covered in a separate class, because many of the techniques are different. Also, high energy techniques often differ.
      - Organization: from astronomical object to measurement: properties of light/photons, effect of earth's atmosphere, telescope, instruments, detectors, data reduction, data analysis
    - practical tools
      - using telescopes and data reduction packages may be necessary skills, but perhaps are not best taught in a class environment. Observing/data reduction may best be learned by taking on a research project, e.g. with your advisor, and being forced to learn by doing. Be aware

that different people use different tools! But critical to understand what is being done, as opposed to the operational set of commands to do it! We'll try to include some of this in class sessions, with possible periodic "work-only" sessions.

- · NOTE: to get most from data (and often, to stand apart from the "crowd"), you need to develop refined or new techniques of data analysis. Beware of canned procedures, or at least, remember that you need to fully understand what they are doing. Also, all instruments generally present slightly new problems which demand different techniques.
- · NOTE: hardly anything (telescopes, instruments, software) works flawlessly all of the time. Understanding how things work is your responsibility, not someone elses, even if it's someone elses responsibility to make them work.
- · potentially at a transition time in software for data reduction. Note issues involving IRAF, astropy, and the ARC DAWG. Some opportunities here!

∗ inquiry

- · it is not possible to teach you everything you need to know, or to know how to do.
- · you need to learn how to ask questions (increasing in depth) and answer them, and especially develop the skill to question yourself

– Beware that, while observational techniques are a critical part of doing science, they are tools to use towards an end: collecting and reducing data may be necessary, but is not sufficient, to do science!

– Note changing face of astronomy, e.g. from traditionally-scheduled telescopes to automated telescopes, HST, SDSS. Traditional mode observing, remote observing, service observing/ queue scheduling, dedicated projects, virtual observatory. Big instruments with dedicated software. Many science users no longer reduce their own data. On the flip side, many jobs have to do with development of data pipelines, and expertise along these lines seems to be getting more scarce.

- Logistics

  – We will attempt for some topics to flip the classroom, with reading assignments that significant time/effort should be dedicated to. Come to class with questions. Some simple quizzes will be occasionally be imposed to ensure that reading is taken seriously. Importance of reading.

    * For this work, must spend sufficient time reading for full comprehension. Remember, the idea is that the material will not be repeated in class; we will try to probe the understanding and ability to apply it.
    * Strategies for reading: small bits, multiple times per day and multiple days
    * Question your understanding
    * Look at and work problems
    * Formulate questions: when possible, send them to me via email
- We will attempt to spend some class time working problems and/or developing techniques. Some of this will be in shorter blocks to break up discussion/presentation, but we will have occasional full periods of practical sessions to develop some familiarity with working with data. While we will attempt to couple these sessions with material that we have been discussing, occasionally the practical sessions may follow a separate thread. If we do not complete practical work during a class period, the work be finished outside of class.
- Will attempt to assign frequent short exercises, and less frequent somewhat more involved problems. Note there will be different expectations for the graduate and undergraduates in the class. Some problems may have been given, along with answers, in previous years: you are implicitly expected NOT to query older graduate students about these.
- One midterm (early October?), one final.
- class info/notes accessible through Canvas on web: http://astronomy.nmsu.edu/holtz/a535
    * web notes are not a textbook, and often just provide a framework
    * Steve Majewski (UVa) class notes
- Texts/references:
    * Chromey, *To Measure the Sky*, covers much of the material from the class, at moderately low level
    * Rieke, *Measuring the Universe*. Relatively new book that covers much of the material we will discuss – although not always from the same perspective.
    * Schroeder, *Astronomical Optics*. Good reference on optics relevant for astronomy. Goes into significantly more detail than we will cover in class.
    * *Astronomical CCD Observing and Reduction Techniques* (ed., S. Howell). Out of print.
    * Sutton, Observational Astronomy
    * Birney et al, Observational Astronomy
    * Kitchin, Astrophysical Techniques, and Optical Astronomical Spectroscopy

    - Howell, Handbook of CCD Astronomy
    - Bevington, Data Reduction and Error Analysis for the Physical Sciences: error analysis and propagation, least squares fitting, etc.
    - Press et al., Numerical Recipes, lots of numerical techniques: fitting, interpolation, lots more
  - trip to APO - two weekends in October ??. Housing. Training.

- Pre-class assessment

- Introduction to NMSU telescopes(APO):

  - ARC 3.5m: consortium with U. Wash (25%), NMSU (15.625%), U. Colorado (12.5%), Johns Hopkins University (8%), U. Virginia (6.25%), Georgia State University (6.25%), University of Oklahoma (6.25%) University of Wyoming(6.25%) several leasing partners. Instruments:
    - ARCTIC: "wider-field" optical imager (SPICAM - old imager)
    - AGILE - high speed imager
    - DIS - double imaging spectrograph: slit spectrograph
    - KOSMOS - coming soon! slit spectrograph
    - ARCES - echelle spectrograph
    - NICFPS: IR imager (+grism)
    - TRIPLESPEC: IR spectrograph
  - SDSS 2.5m: wide field telescope. Projects:
    - SDSSI (-2005): Sloan galaxy survey: image 1/4 of northern sky, identify galaxies, get spectra of all galaxies (about a million) brighter than some limit, derive redshifts/distances to produce a 3D map. Successful in this, but perhaps even more successful in helping to develop a firmer understanding of the distribution of quantitative galaxy properties, and also to provide a homogeneous data set of photometry and galaxy spectra across a large region of the sky.
    - SDSSII (2005-2008): finish Sloan galaxy survey, Sloan supernove survey, SEGUE: Sloan Extension for Galactic Understanding and Exploration. SN survey images a strip of the sky (stripe 82) repeatedly to look for time variable objects. SEGUE extends spectrscopy to target Milky Way stars, mostly at higher Galactic latitudes
    - SDSSIII (2008-2014): BOSS, MARVELS, SEGUE-2, APOGEE. BOSS gets reshifts of luminous red galaxies and quasars to study baryon acoustic

osciallation signature in the spatial distribution of galaxies. MARVELS uses novel technology to detect planets. APOGEE targets MW stars, focussing on lower Galactic latitudes, using a new multi-object near-IR spectrograph

  * SDSSIV (2014-2020): eBOSS, MaNGA, APOGEE (inc. APOGEE-S). eBOSS extends BAO observations to higher redshift. APOGEE continues map the Milky Way; APOGEE starts southern hemisphere observations at las Campanas Observatory with a second spectrograph. MaNGA program to get integral field spectroscopy of 10,000 "nearby" galaxies.
  * SDSSV (2020-): Milky Way Mapper (MWM), Black Hole Mapper (BHM), Local Volume Mapper (LVM).

- NMSU 1m. Instruments:
  * 2048x2048 camera
  * feed to SDSS APOGEE (but no longer usable after eBOSS completion, since APOGEE spectrograph is used all the time with the 2.5m)
  * capability: high speed multichannel photometer
  * ARCSAT: 0.5m with two imaging cameras, FLARECAM and SURVEY-CAM
- NMSU Tortugas mountain observatory : 24" with imaging camera. Introduction sessions next week.
- NMSU campus observatory. Training session Monday 8/26.

- Observing opportunities: APO, DST, NOAO, HST. Proposal procedure. Existing data opportunities: SDSS, HST archive, etc.
- Student experience/interests

2 Light, magnitudes, and the signal equation

2.1 Light

- astronomy - learn about objects outside of our atmosphere. need information from them to do so: light!
- light, by quantum mechanics, is photons, has characteristics of both waves and particles. Wavelength/frequency corresponds to energy:

$$E = h\nu = \frac{hc}{\lambda}$$

- electromagnetic spectrum: gamma rays - X rays - UV - optical - IR - mm - radio. Different units often used for wavelength in different parts of spectrum: $1\mathring{A} = 1 \times 10^{-10}$ m (used in UV, optical), $1nm = 1 \times 10^{-9}$ m (used in UV, optical), $1\mu = 1 \times 10^{-6}$ m (used in IR), $1mm = 1 \times 10^{-3}$ m

  Numerical wavelengths of different parts of spectrum (roughly, there is no established strict vocabulary!): far-UV ($0.01 - 0.1\mu$, 100-1000 Å), near-UV ($.1 - 0.35\mu$, 1000-3500 Å), optical ($0.35 - 1\mu$, 3500-10000 Å), near-IR ($1 - 10\mu$), mid-IR ($10 - 100\mu$), far-IR ($100 - 1000\mu$). Of course, some people used frequency instead of wavelength! And others, especially for high energy radiation, use energy!

- We can describe the amount of light an object emits or that we receive by three fundamental quantities: intensity or surface brightness, I, flux F, or luminosity, L.
- The surface brightness is defined as the amount of energy received in a unit surface element per unit time per unit frequency (or wavelength) from a unit solid angle in the direction (θ, ϕ), where θ is the angle away from the normal to the surface element, and ϕ the azimuthal angle. The solid angle is related to the physical size of an object and its distance: $d\Omega = dA/d^2$. Note units often used for solid angle: sterradian, square degree, square arcsec.

  The flux is the amount of energy passing through a unit surface element in all directions, defined by

$$F_\nu = \int I_\nu \cos\theta d\Omega$$

  where $d\Omega$ is the solid angle element, and the integration is over the entire solid angle. Usually, our detectors are pointed such that the light is received perpendicular to the collecting area and the angle subtended by an object is very small, so the $\cos\theta$ term is well approximated by unity.

  The luminosity is the *intrinsic* energy emitted by the source per second. For an isotropically emitting source,

$$L = 4\pi d^2 F$$

where $d =$ distance to source.

- What do we measure for sources? For resolved sources we can directly measure their surface brightness (intensity) distribution on the sky, usually over some bandpass or wavelength interval; for unresolved sources, we measure the flux. We can only calculate the luminosity to a source if we know the distance.

- There is an additional property of light that we haven't discussed: polarization. As a transverse wave, the electromagnetic field associated with a photon is oriented in a particular direction. For many sources of emission, all orientations are produced, and the result is unpolarized light. However, for some emission mechanisms, light can be polarized, with a fraction of it all oriented in the same plane (linear polarized) or with a plane that rotates as it propagates (circularly polarized). In particular, polarization can arise from emission that is reflected and also from emission in regions of magnetic fields. Generally, the polarization is characterezed by the Stokes paramteres, I, Q, U, and V, which give the polarization intensity, two components of linear polarization, and circular polarization.

- Note, there are variations in terminology, especially between disciplines (e.g. astronomy and engineering) and you should make sure you understand what is adopted by someone presenting data. While astronomers are reasonably consistent in the use of these terms, you may run into things like radiance(surface brightness), irradiance (flux), radiant flux (luminosity), spectral intensity, and others, which are all related to intensity or flux.

- Amount of light emitted is a function of wavelength, so we actually are often interested in estimates of the monochromatic flux/intensity/luminosity, sometimes known as flux/intensity/luminosity density = flux / unit wavelength (or unit frequency), also sometimes known as specific flux/intensity/luminosity.

  F_ν : flux per unit frequency. F_λ : flux per unit wavelength.

$$\int_{nu1}^{nu2} F_\nu d\nu = -\int_{lambda1}^{lambda2} F_\lambda d\lambda$$

$$F_\nu = -F_\lambda d\lambda / d\nu$$
$$= F_\lambda \frac{c}{\nu^2} = F_\lambda \frac{\lambda^2}{c}$$

Similarly, intensity and luminosity can be given per unit wavelength (or frequency). Note that a constant F_λ implies a non-constant F_ν and vice versa!

Integral of flux/brightness over all wavelengths/frequencies gives the *bolometric* flux/brightness.

- Units: astronomers often (not always) work in CGS units, although, as discussed below, they most often work in a dimensionless unit ... magnitudes.
  - You may also run into the Jansky, a flux density unit corresponding to $10^{-26} W/m^2/Hz$
- In practice, with most detectors, we measure photon flux (photon counting devices), rather than energy flux (bolometers). The photon flux is given by

$$\text{photonflux} = \text{flux/energyperphoton}$$

$$= \int F_\lambda \frac{\lambda}{hc} d\lambda$$

- Terminology of measurements:
  - photometry (broad-band flux measurement),
  - spectroscopy (relative measurement of fluxes at different wavelengths),
  - spectrophotometry (absolute measurement of fluxes at different wavelengths).
  - astrometry: concerned with positions of observed flux;
  - morphology: intensity as a function of position; often, absolute measurements are unimportant

Know the basic characteristics of light: energy, frequency and wavelength, and how they are related. Know the different regimes in the electromagnetic spectrum, the units used to characterize the different regimes, and characteristic wavelengths for each regime. Have a complete understanding of the difference between intensity, flux, and luminosity and their units. Recognize and understand that these are generally a function of wavelength/frequency andcan be specified per unit wavelength or per unit frequency and how to convert between the two.

2.2 Magnitudes and photometric systems

In astronomy, however, magnitude units are often used instead of measuring the basic quantities in energy or photon flux. Magnitudes are a dimensionless quantities, and are related to flux (same holds for surface brightness or luminosity) by:

$$m = -2.5 \log \frac{F}{F_0}$$

or

$$m = -2.5 \log F + 2.5 \log F_0$$

where the coefficient of proportionality, F_0, depends on the definition of photometric system; the quantity $-2.5 \log F_0$ may be referred to as the photometric system zeropoint. This defining equation is sometimes referred to as the Pogson equation, after Pogson (1856). Inverting, one gets:

$$F = F_0 10^{-0.4m}$$

Note that since magnitudes are logarithmic, the *difference* between magnitudes corresponds to a ratio of fluxes; ratios of magnitudes are generally unphysical! If one is just doing relative measurements of brightness between objects, this can be done without knowledge of F_0 (or, equivalently, the system zeropoint); objects that differ in brightness by ΔM mag have the same ratio of brightness rgardless of what photometric system they are in:

$$m_1 - m_2 = -2.5 \log \frac{F_1}{F_2}$$

$$\frac{F_1}{F_2} = 10^{-0.4(m_1 - m_2)}$$

The photometric system definitions and zeropoints are only needed when converting between calibrated magnitudes and fluxes. However, the utility of a system when doing astrophysics generally requires an understanding of the actual fluxes.

Luminosities are represented as absolute magnitudes, i.e., the magnitude a star would have if it were at a distance of 10 parsec; as before, you need a distance to get a luminosity. The inverse square law expressed in magnitudes leads to the *distance modulus*:

$$m_0 - M = 5 \log d - 5$$

(derive it!), where m_0 is the apparent magnitude corrected for interstellar extinction: $m_0 = m - A$.

Just as fluxes can be represented in magnitude units, flux densities can be specified by monochromatic magnitudes:

$$F_\lambda = F_0(\lambda) 10^{-0.4m(\lambda)}$$

although spectra are more often given in flux units than in magnitude units. Note that it is possible that F_0 is a function of wavelength!

Know how magnitudes are defined, be able to work with them very comfortably, and recognize that relative fluxes can be represented as magnitudes independent of the magnitude system.

There are three main types of magnitude systems in use in astronomy. We start by describing the two simpler ones:the STMAG and the ABNU mag system. In these simple system, the reference flux is just a constant value in F_λ or F_ν. However, these are not always the most widely used systems in astronomy, because no natural source exists with a flat spectrum.

In the STMAG system, $F_{0,\lambda} = 3.63E-9 ergs/cm^2/s/\mathring{A}$, which is the flux of Vega at $5500\mathring{A}$; hence a star of Vega's brightness at $5500\mathring{A}$ is defined to have m=0. Alternatively, we can write

$$m_{STMAG} = -2.5 \log F_\lambda - 21.1$$

(for F_λ in cgs units).

In the ABNU system, things are defined for F_ν instead of F_λ, and we have

$$F_{0,\nu} = 3.63 \times 10^{-20} erg/cm^2/s/Hz 10^{-0.4 m_\nu}$$

or

$$m_{ABNU} = -2.5 \log F_\nu - 48.6$$

(for F_ν in cgs units). Again, the constant comes from the flux of Vega.

Usually, when using magnitudes, people are talking about flux integrated over a spectral bandpass. In this case, F and F_0 refer to fluxes integrated over the bandpass. The STMAG and ABMAG integrated systems are defined relative to sources of constant F_λ and F_ν systems, respectively.

$$m_{STMAG} = -2.5 \log \frac{\int F_\lambda \lambda d\lambda}{\int 3.63 \times 10^{-9} \lambda d\lambda}$$

(the factor of λ comes in for photon counting detectors).

$$m_{ABNU} = -2.5 \log \frac{\int (F_\nu/\nu) d\nu}{\int (3.63 \times 10^{-20}/\nu) d\nu}$$

(where the units are implicitly cgs with these numerical fluxes for Vega).

Note that these systems differ by more than a constant, because one is defined by units of F_λ and the other by F_ν, so the difference between the systems is a function of wavelength. They are defined to be the same at $5500\mathring{A}$. (Question: what's the relation between m_{STMAG} and m_{ABNU}?)

Note also that, using magnitudes, the measured magnitude is nearly independent of bandpass width (a broader bandpass does not imply a brighter (smaller) magnitude), which is not the case for fluxes!

The standard UBVRI broadband photometric system, as well as several other magnitude systems, however, are not defined for a constant F_λ or F_ν spectrum; rather, they are defined relative to the spectrum of an A0V star. Most systems are defined (or at least were originally) to have the magnitude of Vega be zero in all bandpasses (VEGAMAGS); if you ever get into this in detail, note that this is not exactly true for the UBVRI system.

For the broadband UBVRI system, we have

$$m_{UBVRI} \approx -2.5 \log \frac{\int_{UBVRI} F_\lambda(object) \lambda d\lambda}{\int_{UBVRI} F_\lambda(Vega) \lambda d\lambda}$$

(as above, the factor of λ comes in for photon counting detectors).

Figure 1 shows a plot to demonstrate the difference between the different systems.

Why do the different systems exist? While it seems that STMAG and ABNU systems are more straightforward, in practice it is difficult to measure absolute fluxes, and much easier to measure relative fluxes between objects. Hence, historically observations were tied to observations of Vega (or to stars which themselves were tied to Vega), so VEGAMAGs made sense, and the issue of determining physical fluxes boiled down to measuring the physical flux of Vega. Today, in some cases, it may be more accurate to measure the absolute throughput of an instrumental system, and using STMAG or ABNU makes more sense.

Know that there are several different magnitude systems in use, and understand how they differ. Know when it is important to know what the magnitude system is, and when it isn't.

2.2.1 Colors

Working in magnitudes, the difference in magnitudes between different bandpasses (called the color index, or simply, color) is related to the flux ratio between the bandpasses, i.e., the color. In the UBVRI system, the difference between magnitudes gives the ratio of the fluxes in different bandpasses *relative to the ratio of the fluxes of an A0V star in the*

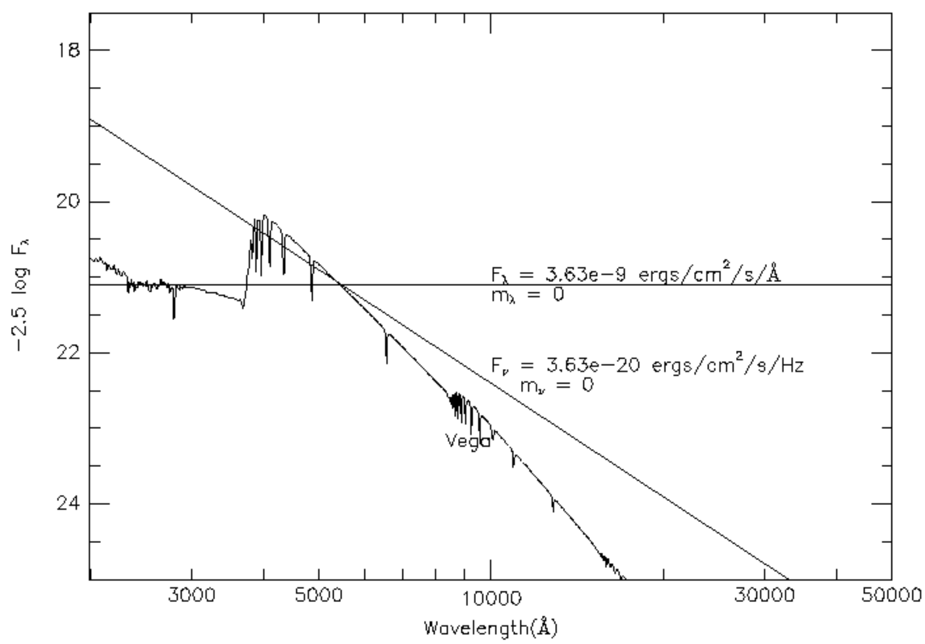


Figure 1: Differences between different magnitude systems

different bandpasses (for VEGAMAG). Note the typical colors of astronomical objects – which are different for the different photometric systems!

Which is closer to the UBVRI system, STMAG or ABNU?

What would typical colors be in an STMAG or ABNU system?

Understand how colors are represented by a difference in magnitude, and recognize how colors expressed in magnitudes are related to the shape of the underlying spectrum, with differences for different magnitude systems.

2.2.2 Magnitude-flux conversion

How would one go about converting Vega-based magnitudes to fluxes? Roughly, just look up the flux of Vega at the center of the passband (e.g., here (from Bessell et al 1998 or here (see references within), or here; note, however, if the spectrum of the object differs from that of Vega, this won't be perfectly accurate (see, e.g. discussion of WISE photometry) Given UBVRI magnitudes of an object in the desired band, filter profiles (e.g. Bessell 1990, PASP 102,1181), and absolute spectrophotometry of Vega (e.g., Bohlin & Gilliland 2004, AJ 127, 3508, one can determine the flux.

If one wanted to estimate the flux of some object in arbitrary bandpass given just the V magnitude of an object (a common situation used when trying to predict exposures times, see below), this can be done if an estimate of the spectral energy distribution (SED) can be made (e.g., from the spectral type, or more generally, the stellar parameters T_{eff}, $\log g$, and metallicity). Given the filter profiles, one can compute the integral of the SED over the V bandpass, determine the scaling by comparing with the integral of the Vega spectrum over the same bandpass, then use the normalized SED to compute the flux in any desired bandpass. Some possibly useful references for SEDs are: Pickles atlas, MILES library, Bruzual, Persson, Gunn, & Stryker; Hunter, Christian, & Jacoby; Kurucz).

Things are certainly simpler in the ABNU or STMAG system, and there has been some movement in this direction: the STScI gives STMAG calibrations for HST instruments, and the SDSS photometric system is close to an ABNU system.

Note, however, that even when the systems are conceptually well defined, determining the absolute calibration of any photometric system is very difficult in reality, and determining absolute fluxes to the 1% level is very challenging.

As a separate note on magnitudes themselves, note that some people, in particular, the SDSS imaging survey, have adopted a modified type of magnitudes, called asinh magnitudes, which behave like normal (also known as Pogson) magnitude for brighter objects, but have different behavior for very faint objects (near the detection threshold); see Lupton, Gunn, & Szalay 1999 AJ 118, 1406 for details.

2.3 Observed fluxes, the signal equation, and photometry

What if you are measuring flux with an actual instrument, i.e. counting photons? The intrinsic photon flux from the source is not trivial to determine from the observed photon flux, i.e., the number of photons that you count. The observed flux depends on the area of your photon collector (telescope), photon losses and gains from the Earth's atmosphere (which changes with conditions), and the efficiency of your collection/detection apparatus (which can change with time). Generally, the astronomical *signal* (which might be a flux or a surface brightness, depending on whether the object is resolved) can be written

$$S = Tt \int \frac{F_\lambda}{\frac{hc}{\lambda}} a_\lambda tel_\lambda inst_\lambda filt_\lambda det_\lambda d\lambda \equiv TtS'$$

where S is the observed photon flux (the *signal*), T is the telescope collecting area, t is the integration time, a_λ is the atmospheric transmission (more later) and the other terms refer to the efficiency of various components of the system (telescope, instrument, filter, detector). S' is an observed flux rate, i.e. with all of the real details of the observing system included. I refer to this as the *signal equation*.

Usually, however, one doesn't use this information to go backward from S to F_λ because it is very difficult to measure all of the terms precisely, and some of them (e.g. a, and perhaps some of the system efficiencies) are time-variable; a is also spatially variable.

While the signal equation isn't usually used for calibration, it is very commonly used for computing the approximate number of photons you will receive from a given source in a given amount of time for a given observational setup. This number is critical to know in order to estimate your expected errors and exposure times in observing proposals, observing runs, etc. Understanding errors in absolutely critical in all sciences, and maybe even more so in astronomy, where objects are faint, photons are scarce, and errors are not at all insignificant. The signal equation provides the basis for exposure time calculator (ETC) programs, because it gives an expectation of the number of photons that will be received by a given instrument as a function of exposure time. As we will see shortly, this provides the information we need to calculate the uncertainty in the measurement as a function of exposure time.

2.3.1 Photometry

So if we don't use the signal equation for calibration, how do we go about determining calibrated brightnesses from measurements? To do this, most observations are performed differentially to a set of other stars of known brightness. If one or more stars of known brightness are observed in the same observation, then the atmospheric term is (approximately) the same for all stars; this is known as *differential photometry*. From the photon flux of the object with known brightness, one can calculate an *instrumental magnitude*:

$$m = -2.5 \log(S/t)$$

and then determine the *zeropoint* that needs to be added to give the calibrated magnitude (M, make sure you recognize that this is still an apparent magnitude!):

$$M = m + z$$

Note that the zeropoint gives a measure of the system sensitivity: it is the magnitude of an object which produces 1 count/s, so a larger zeropoint indicates a more sensitive system (i.e., from larger aperture, throughput, etc.); alternatively, one can calculate an "effective area" for an exposure. The normalization by the exposure time in the instrumental magnitude to get counts/sec is not strictly necessary, but it is useful if you are using the zeropoint from one exposure to calibrate another exposure of a different exposure time.

Note that in the real world, one has to also consider sensitivity differences (e.g., slightly different filter profiles) between a given experimental setup and the setup used to measure the reference brightnesses. If the experimental system differs in response details to the standard system, the zeropoint will be different for objects with different spectral energy distributions. Usually, at attempt is made to calibrate this using so-called tranformation coefficients and parametrizing the SED differences by the color of the objects. The relation between the instrumental magnitude and the standard magnitude is given by:

$$M = m + t(color) + z$$

where capital letters are the magnitude on the standard system, z is the zeropoint, and t is the *transformation coefficent*. There is a separate such relation for each filter in which observations are made.

The color is generally parameterized by the ratio of the flux at two different wavelengths, or, in magnitudes, the difference between the magnitudes. The two wavelengths should be measured near in wavelength to the wavelength of the filter being corrected; generally, one uses the bandpass being corrected as one of the wavelenghts and an adjacent bandpass as the other. For example, when correcting V magnitudes, people usually use $B - V$, $V - R$, or $V - I$ for the color term, e.g.:

$$V = m_V + t_V(B - V) + z_V$$

Clearly, to do this solution, you need more than one standard star, since there are two unknowns (t_V and z_V), and to get a meaningful estimate of t_V, you want the standard stars to cover as wide a range of color as possible. While you can solve for the coefficients with two stars, one generally would like to have more than this, and solve for the coefficients using, e.g., least squares.

There are two ways to define the color, either in terms of the observational system or in terms of the standard system. The latter is slightly preferred for using least-squares (small errors on the independent variable), and also because it allows observations from different nights to be combined. Note that this formulation does not require you to know the colors of your objects a priori, it's just algebra to figure them out as long as you have observations in both filters, e.g., once you have the transformation coefficients and the zeropoints for two filters, you can solve:

$$B = m_B + t_B(B - V) + z_B$$

$$V = m_V + t_V(B - V) + z_V$$

for both B and V given m_B, m_V, t_B, t_V, z_B, and z_V.

The use of these *first-order* transformation coefficients is accurate as long as your filter system does not differ much from the standard system, and additionally, that the spectrum of your program objects does not differ significantly from the spectrum of the standard objects. The more these conditions are not met, the less accurate the results. Some additional accuracy in the case of differing systems can be achieved by using higher order transformation coefficients. However, even in this case, it is always important to remember that if the spectrum of the program object differs significantly from the standards, derived fluxes can be significantly in error.

Certainly, you get to a point when the response of one system is so different than the response of another system that no transformation can be determined. In this case, you have two different photometric systems. In fact, there are several different photometric systems at use in astronomy today, and each has advantages and disadvantages.

If there are no stars of known brightness in the same observation, then calibration must be done against stars in other observations. This then requires that the different effects of the Earth's atmosphere in different locations in the sky be accounted for. This is known as *all-sky*, or absolute, photometry. To do this requires that the sky is "well-behaved", i.e. one can accurately predict the atmospheric throughput as a function of position. This requires that there be no clouds, i.e. *photometric* weather. Differential photometry can be done in non-photometric weather, hence it is much simpler! Of course, it is always possible to obtain differential photometry and then go back later and obtain absolute photometry of the reference stars. We will discuss later how to incorporate the effects of the Earth's atmosphere. However, all-sky photometry is becoming less and less common as catalogs of well calibrated stars are becoming available across the entire sky (e.g., SDSS or PanSTARRS).

Of course, at some point, someone needs to figure out what the fluxes of the calibrating stars really are, and this requires understanding all of the terms in the signal equation. It is challenging, and often, absolute calibration of a system is uncertain to a couple of percent!

It is also common to stop with differential photometry, even if there are no stars of known brightness in your field, if you are studying variable objects, i.e. where you are just interested in the *change* in brightness of an object, not the absolute flux level. In this case, one only has to reference the brightness of the target object relative some other object (or ensemble of objects) in the field that are non-variable. One has to be careful that the reference object is itself not a variable, and this becomes more challenging if you are trying to measure small variations in brightness.

Understand the signal equation and the terms in it. Understand the distinction between estimating count rates from an understanding of all of the terms in the signal equation vs. measuring the overall throughput (zeropoint) by observing stars of known brightness. Know what instrumental magnitudes and zeropoints are. Understand the ideas behind the use of transformation coefficient..

3 Uncertainties and error analysis

For a given rate of emitted photons, there's a probability function which gives the number of photons we detect, even assuming 100% detection efficiency, because of *statistical* uncertainties. In addition, there may also be *instrumental* uncertainties. Consequently, we now turn to the concepts of probability distributions, with particular interest in the distribution which applies to the detection of photons.

3.1 Distributions and characteristics thereof

- concept of a distribution : define $p(x)dx$ as probability of event occuring in $x + dx$:

$$\int p(x)dx = 1$$

Some definitions relating to values which characterize a distribution:

$$\text{mean} \equiv \mu = \int xp(x)dx$$

$$\text{variance} \equiv \sigma^2 = \int (x-\mu)^2 p(x)dx$$

$$\text{standarddeviation} \equiv \sigma = \sqrt{variance}$$

median : mid-point value.

$$\frac{\int_{-\infty}^{x_{median}} p(x)dx}{\int_{-\infty}^{\infty} p(x)dx} = \frac{1}{2}$$

mode : most probable value

Note that the geometric interpretation of above quantities depends on the nature of the distribution; although we all carry around the picture of the mean and the variance for a Gaussian distribution, these pictures are not applicable to other distributions, but the quantities are still well-defined.

Also, note that there is a difference between the *sample* mean, variance, etc. and the *population* quantities. The latter apply to the true distribution, while the former are estimates of the latter from some finite sample (N measurements) of the population. The sample quantities are derived from:

$$\text{sample mean} : \bar{x} \equiv \frac{\sum x_i}{N}$$

$$\text{sample variance} \equiv \frac{\sum(x_i - \bar{x})^2}{N-1} = \frac{\sum x_i^2 - (\sum x_i)^2/N}{N-1}$$

The sample mean and variance approach the true mean and variance as N approaches infinity. But note, especially for small samples, your estimate of the mean and variance may differ from their true (population) values (more below)!

Understand the concept of probability distribution functions and basic quantities used to describe them: mean, variance, standard deviation, median, and mode. Understand the difference between population quantities and sample quantities.

3.1.1 The binomial distribution

Now we consider what distribution is appropriate for the detection of photons. The photon distribution can be derived from the *binomial* distribution, which gives the probability of observing the number, x, of some possible event, given a total number of events n, and the probability of observing the particular event (among all other possibilities) in any single event, p, under the assumption that all events are independent of each other:

$$P(x, n, p) = \frac{n! p^x (1-p)^{n-x}}{x!(n-x)!}$$

For the binomial distribution, one can derive:

$$\text{mean} \equiv \int x p(x) dx = np$$

$$\text{variance} \equiv \sigma^2 \equiv \int (x-\mu)^2 p(x) dx = np(1-p)$$

3.1.2 The Poisson distribution

In the case of detecting photons, n is the total number of photons emitted, and p is the probability of detecting a photon during our observation out of the total emitted. We don't know either of these numbers! However, we do know that $p << 1$ and we know, or at least we can estimate, the mean number detected:

$$\mu = np$$

.

In this limit, the binomial distribution asymtotically approaches the *Poisson* distribution:

$$P(x, \mu) = \frac{\mu^x \exp^{-\mu}}{x!}$$

From the expressions for the binomial distribution in this limit, the mean of the distribution is μ, and the variance is

$$\text{variance} = \sum_x [(x - \mu)^2 p(x, \mu)]$$

$$\text{variance} = np = \mu$$

$$\sigma = \sqrt{\mu}$$

This is an *important result*.

Note that the Poisson distribution is generally the appropriate distribution not only for counting photons, but for *any* sort of counting experiment where a series of events occurs with a known average rate, and are independent of time since the last event.

Figure 2 shows some plots of the Poisson distribution for $\mu = 2, \mu = 10$, and $\mu = 10000$.

Understand the Poisson distribution and when it applies. Know how the variance/standard deviation of the Poisson distribution is related to the mean of the distribution.

3.1.3 The normal, or Gaussian, distribution

Note, for large μ, the Poisson distribution is well-approximated around the peak by a *Gaussian*, or *normal* distribution:

$$P(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

This is important because it allows us to use "simple" least squares techniques to fit observational data, because these generally assume normally distributed data. However, beware that in the tails of the distribution, and at low mean rates, the Poisson distribution can differ significantly from a Gaussian distribution. In these cases, least-squares may not be appropriate to model observational data; instead, one might need to consider maximum likelihood techniques instead.

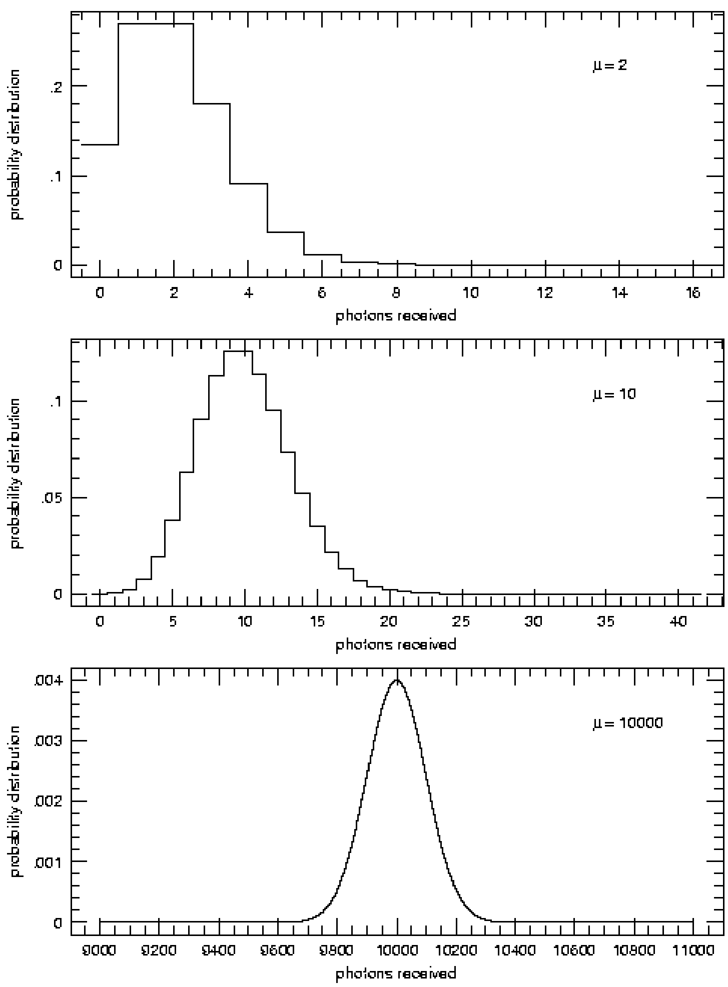


Figure 2: The Poisson distribution for three values of μ

.

The normal distribution is also important because many physical variables seem to be distributed accordingly. This may not be an accident because of the *central limit theorem*: if a quantity depends on a number of independent random variables with ANY distribution, the quantity itself will be distributed normally (see statistics texts).

In observational techniques, we encounter the normal distribution because one important source of instrumental noise, *readout noise*, is distributed normally.

Know what a Gaussian (normal) distribution is, including the full functional form of it. Understand under what circumstances the Poisson distribution is similar to a normal distribution.

3.2 Importance of error distribution analysis

You need to understand the expected uncertainties in your observations in order to:

- predict the amount of observing time you'll need to get uncertainties as small as you need them to do your science,
- answer the question: is scatter in observed data consistent with expected uncertainties? If the answer is no, they you know you've either learned some astrophysics or you don't understand something about your observations. This is especially important in astronomy where objects are faint and many projects are pushing down into the noise as far as possible. Really we can usually only answer this probabilistically. Generally, tests compute the probability that the observations are consistent with an expected distribution (the null hypothesis). You can then look to see if this probability is low, and if so, you can reject the null hypothesis.
- interpret your results in the context of a scientific prediction

3.3 Confidence levels

For example, say we want to know whether some single point is consistent with expectations, e.g., we see a bright point in multiple measurements of a star, and want to know whether the star flared. Say we have a time sequence with known mean and variance, and we obtain a new point, and want to know whether it is consistent with known distribution?

If the form of the probability distribution is known, then you can calculate the probability of getting a measurement more than some observed distance from the mean. In the case where the observed distribution is Gaussian (or approximately so), this is done using the *error function* (sometimes called *erf(x)*), which is the integral of a gaussian from some starting value.

Some simple guidelines to keep in mind follow (the actual situation often requires more sophisticated statistical tests). First, for Gaussian distributions, you can calculate that 68% of the points should fall within plus or minus one sigma from the mean, and 95.3% between plus or minus two sigma from the mean. Thus, if you have a time line of photon fluxes for a star, with N observed points, and a photon noise σ on each measurement, you can test whether the number of points deviating more than 2σ from the mean is much larger than expected. To decide whether any single point is really significantly different, you might want to use more stringent criterion, e.g., 5σ rather than a 2σ criterion; a 5σ has much higher level of significance. On the other hand, if you have far more points in the range $2 - -4\sigma$ brighter or fainter than you would expect, you may also have a significant detection of intensity variations (provided you really understand your uncertainties on the measurements!).

Also, note that your observed distribution should be consistent with your uncertainty estimates given the above guidelines. If you have a whole set of points, that all fall within 1σ of each other, something is wrong with your uncertainty estimates (or perhaps your measurements are correlated with each other)!

For a series of measurements, one can calculate the χ^2 statistic, and determine how probable this value is, given the number of points.

$$\chi^2 = \sum \frac{(observed(i) - model(i))^2}{\sigma_i^2}$$

For a given value of χ^2 and number of measurements, one can calculate the probability that the measurements are consistent with the model (and the uncertainties are correctly predicted). A quick estimate of the consistency of the model with the observed data points can be made using reduced χ^2, defined as χ^2 divided by the *degrees of freedom* (number of data points minus number of parameters), which should be near unity if the measurements are consistent wth the model.

4 Noise equation: how do we predict expected uncertainties?

4.1 Signal-to-noise

Astronomers often describe uncertainties in terms of the fractional error, e.g. the amplitude of the uncertainty divided by the amplitude of the quantity being measured; often, the inverse of this, referred to as the *signal-to-noise ratio* is used. Given an estimate the number of photons expected from an object in an observation, we can calulate the *signal-to-noise* ratio:

$$\frac{S}{N} = \frac{S}{\sqrt{\sigma^2}}$$

which is the inverse of the predicted fractional error (N/S).

Consider an object with observed photon flux (per unit area and time, e.g. from the signal equation above), S', leading to a signal, $S = S'Tt$ where T is the telescope area and t is the exposure time. In the simplest case, the only noise source is Poisson statistics from the source, in which case:

$$\sigma^2 = S = S'Tt$$

$$\frac{S}{N} = \sqrt{S} = \sqrt{S'Tt}$$

In other words, the S/N increases as the square root of the object brightness, telescope area, efficiency, or exposure time. Note that S is directly observable, so one can calculate the S/N for an observation without knowing the telescope area or exposure time! We've just broken S down so that you can specifically see the dependence on telescope area and/or exposure time.

Understand the concept of S/N and fractional error. Know how S/N depends on the signal for the Poisson-limited case.

4.1.1 Background noise

A more realistic case includes the noise contributed from Poisson statistics of "background" light (more on the physical nature of this later), B', which has units of flux per area on the sky (i.e. a surface brightness); note that this is also usually given in magnitudes.

$$B' = \int \frac{B_\lambda}{\frac{hc}{\lambda}} q_\lambda d\lambda$$

The amount of background in our measurement depends on how we choose to make the measurement (how much sky area we observe). Say we just use an aperture with area, A, so the total observed background counts is

$$AB = AB'Tt$$

Again, $B'Tt$ is the directly observable quantity, but we split it into the quantities on which it depends to understand what factors are important in determining S/N.

The total number of photons observed, O, is

$$O = S + AB = (S' + AB')Tt$$

The variance of the total observed counts, from Poisson statistics, is:

$$\sigma_O^2 = O = S + AB = (S' + AB')Tt$$

To get the desired signal from the object only , we will need to measure separately the total signal and the background signal to estimate:

$$S \equiv S'Tt = O - A < B >$$

where $< B >$ is some estimate we have obtained of the background surface brightness. The noise in the measurement is

$$\sigma_S^2 \approx \sigma_O^2 = S + AB = (S' + AB')Tt$$

where the approximation is accurate if the background is determined to high accuracy, which one can do if one measures the background over a large area, thus getting a large number of background counts (with correspondingly small fractional error in the measurement).

This leads to a common form of the *noise equation*:

$$\frac{S}{N} = \frac{S}{\sqrt{S + AB}}$$

Breaking out the dependence on exposure time and telescope area, this is:

$$\frac{S}{N} = \frac{S'}{\sqrt{S' + AB'}}\sqrt{T}\sqrt{t}$$

In the *signal-limited* case, $S' >> B'A$, we get

$$\frac{S}{N} = \sqrt{S} = \sqrt{S'tT}$$

In the *background limited* case, $B'A >> S'$, and

$$\frac{S}{N} = \frac{S}{\sqrt{AB}} = \frac{S'}{\sqrt{B'A}}\sqrt{tT}$$

As one goes to fainter objects, the S/N drops, and it drops faster when you're background limited. This illustrates the importance of dark-sky sites, and also the importance of good image quality.

Consider two telescopes of collecting area, T_1 and T_2. If we observe for the same exposure time on each and want to know how much fainter we can see with the larger telescope at a given S/N, we find:

$$S_2 = \frac{T_1}{T_2} S_1$$

for the signal-limited case, but

$$S_2 = \sqrt{\frac{T_1}{T_2}} S_1$$

for the background-limited case.

Understand how Poisson uncertainties in the background contributes to the S/N of an object, and how the background contribution depends on both the brightness of the background but also on the image quality because the amount of background included in the measurement.

4.1.2 Instrumental noise

In addition to the uncertainties from Poisson statistics (statistical noise), there may be additional terms from instrumental uncertainties. A common example of this that is applicable for CCD detectors is readout noise, which is additive *noise* (with zero mean!) that comes from the detector and is independent of signal level. For a detector whose readout noise is characterized by σ_{rn},

$$\frac{S}{N} = \frac{S}{\sqrt{S + BA_{pix} + \sigma_{rn}^2}}$$

if a measurement is made in a single pixel. If an object is spread over N_{pix} pixels, then

$$\frac{S}{N} = \frac{S}{\sqrt{S + BA + N_{pix}\sigma_{rn}^2}}$$

For large σ_{rn}, the behavior is the same as the background limited case. This makes it clear that if you have readout noise, image quality (and/or proper optics to keep an object from covering too many pixels) is important for maximizing S/N. It is also clear that it is critical to have minimum read-noise for low background applications (e.g., spectroscopy).

There are other possible additional terms in the noise equation, arising from things like dark current, digitization noise, uncertainties in sky determination, uncertainties from photometric technique, etc. (we'll discuss some of these later on), but in most applications, the three sources discussed so far – signal noise, background noise, and readout noise – are the dominant noise sources.

Note the applications where one is likely to be signal dominated, background dominated, and readout noise dominated.

Understand readout noise and how it represented by a normal distribution with zero mean. Know under what circumstances readout noise is an important contributor to the total noise.

4.2 Error propagation

Why are the three uncertainty terms in the noise equation added in quadrature? The measured quantity (S) is a sum of $S + B - <B> + <R>$, where $<R> = 0$ since readout noise has zero mean. The uncertainty in a summed series is computed by addding the individual uncertainties in quadature; in the equation above, we have neglected the uncertainty in $<B>$. To understand why they add in quadrature, let's consider general error propagation.

More reasons to consider error propagation: let's say we want to make some calculations (e.g., calibration, unit conversion, averaging, conversion to magnitudes, calculation of colors, etc.) using these observations: we need to be able to estimate the uncertainties in the calculated quantities that depend on our measured quantities.

Consider what happens if you have several known quantities with known error distributions and you combine these into some new quantity: we wish to know what the uncertainty is in the new quantity.

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

The question is what is σ_x if we know σ_u, σ_v, etc.?

As long as uncertainties are small:

$$x_i - <x> \sim (u_i - <u>) \left(\frac{\partial x}{\partial u}\right) + (v_i - <v>) \left(\frac{\partial x}{\partial v}\right) + ...$$

$$\sigma_x^2 = lim(N- > \infty)\frac{1}{N}\sum(x_i - < x >)^2$$
$$= \sigma_u^2\left(\frac{\partial x}{\partial u}\right)^2 + \sigma_v^2\left(\frac{\partial x}{\partial v}\right)^2 + 2\sigma_{uv}^2\frac{\partial x}{\partial u}\frac{\partial x}{\partial v} + ...$$

The last term is the *covariance*, which relates to whether uncertainties are *correlated*.

$$\sigma_{uv}^2 = lim(n- > \infty)\frac{1}{N}\sum(u_i - < u >)(v_i - < v >)$$

If uncertainties are uncorrelated, then $\sigma_{uv} = 0$ because there's equal chance of getting opposite signs on v_i for any given u_i. When working out uncertainties, make sure to consider whether there are correlated errors! If there are, you *may* be able to reformulate quantities so that they have independent errors: this can be very useful!

Examples for *uncorrelated* errors:

- addition/subtraction:

  $$x = u + v$$
  $$\sigma_x^2 = \sigma_u^2 + \sigma_v^2$$

  In this case, errors are said to *add in quadrature*.

- multiplication (or division, if you formulate the division problem as a multiplication!):

  $$x = uv$$
  $$\sigma_x^2 = v^2\sigma_u^2 + u^2\sigma_v^2$$

- natural logs:

  $$x = \ln u$$
  $$\sigma_x^2 = \sigma_u^2/u^2$$

  For base-10 logs, note that $\log x = \log e \ln x$

  Note that when dealing with logarithmic quantities, uncertainties in the log correspond to *fractional* uncertainties in the raw quantity.

4.2.1 Distribution of resultant uncertainties

When propagating errors, even though you can calculate the variances in the new variables, the distribution of uncertainties in the new variables is not, in general, the same as the distribution of uncertainties in the original variables, e.g. if uncertainties in individual variables are normally distributed, uncertainties in output variable are not necessarily.

When two variables are added, however, the output is normally distributed.

Know the error propagation formula and how to apply it.

4.2.2 Determining sample parameters: averaging measurements

We've covered uncertainties in single measurements. Next we turn to averaging measurements. Say we have multiple observations and want the best estimate of the mean and variance of the population, e.g. multiple measurements of stellar brightness. Here we'll define the best estimate of the mean as the value which maximizes the likelihood that our estimate equals the true parent population mean.

For equal uncertainties, this estimate just gives our normal expression for the sample mean:

$$\bar{x} = \frac{\sum x_i}{N}$$

Using error propagation, the estimate of the uncertainty in the sample mean is given by:

$$\sigma_{\bar{x}}^2 = \sum \frac{\sigma_i^2}{N^2} = \frac{\sigma^2}{N}$$

But what if uncertainties on each observation aren't equal, say for example we have observations made with several different exposure times? Then the optimal determination of the mean is using a:

$$\mathrm{weightedmean} = \frac{\sum \frac{x_i}{\sigma_i^2}}{\sum \frac{1}{\sigma_i^2}}$$

and the estimated uncertainty in this value is given by:

$$\sigma_\mu^2 = \sum \frac{\frac{\sigma_i^2}{\sigma_i^4}}{(\sum \frac{1}{\sigma_i^2})^2} = \frac{1}{\sum \frac{1}{\sigma_i^2}}$$

where the σ_i's are individual weights/errors (people often talk about the *weight* of an observation, i.e. $\frac{1}{\sigma_i^2}$: large weight means small uncertainty.

This is a standard result for determining sample means from a set of observations with different weights.

However, there can sometimes be a subtlety in applying this formula, which has to do with the question: how do we go about choosing the weights/errors, σ_i? We know we can *estimate* σ using Poisson statistics for a given count rate, but remember that this is a sample variance (which may be based on a single observation!) not the true population variance. This can lead to biases.

Consider observations of a star made on three nights, with measurements of 40, 50, and 60 photons. It's clear that the mean observation is 50 photons. However, beware of the being trapped by your undertainty *estimates*. From each observation alone, you would estimate uncertainties of $\sqrt{40}$, $\sqrt{50}$, and $\sqrt{60}$. If you plug these uncertainty estimates into a computation of the weighted mean, you'll get a mean rate of 48.64!

Using the individual estimates of the variances, we'll bias values to lower rates, since these will have lower estimated variances.

Note that it's pretty obvious from this example that you should just weight all observations equally. However, note that this certainly isn't always the right thing to do. For example, consider the situation in which you have three exposures of different exposure times, and you are calculating the photon rate (counts/s). Here you probably want to give the longer exposures higher weight (at least, if they are signal or background limited). In this case, you again don't want to use the individual uncertainty estimates or you'll introduce a bias. There is a simple solution here also: just weight the observations by the exposure time. However, while this works fine for Poisson uncertainties (variances proportional to count rate), it isn't strictly correct if there are instrumental uncertainties as well which don't scale with exposure time. For example, the presence of readout noise can have this effect; if all exposures are readout noise dominated, then one would want to weight them equally, if readout noise dominates the shorter but not the longer exposures, once would want to weight the longer exposures even higher than expected for the exposure time ratios! The only way to properly average measurements in this case is to estimate a sample mean, then use this value scaled to the appropriate exposure times as the input for the Poisson uncertainties. See the next subsection for a more comprehensive discussion.

Another subtlety: averaging counts and converting to magnitudes is not the same as averaging magnitudes!

4.2.3 Weighted mean of a series of exposures

Say you have a bunch of measurements of the brightness of an object. Under the assumption that the object is not variable, you wish to get the best estimate of the count rate of the object. Your measurements are not all at the same exposure time. So you have a bunch of measurements:

$$c_i = rt_i$$

where each of the c_i has noise associated with it.

For each measurement, you can calculate a rate:

$$r_i \equiv \frac{c_i}{t}$$

Now you wish to average all of the measured rates to get the best estimate of the true rate. Because the measurements were taken with different exposure times, they have different amount of uncertainty, so you want to use a weighted mean to get the best estimate.

To calculate a weighted mean, you need to determine the variances of the observations. Given the true rate, r, the variances of the measurements are

$$\sigma^2 = rt_i + n\sigma_{rn}^2$$

where n is the number of pixels you've summed over to get the total counts, and σ_{rn} is the readout noise.

To get the variances of the *rates*, $r_i = \frac{c_i}{t}$, we need to divide the variances of the counts by t^2 (i.e. divide the standard deviations by t, so

$$\sigma^2 = r/t_i + n\sigma_{rn}^2/t_i^2$$

The problem is that you don't know what r is. Normally, you might just take

$$\sigma_i^2 = r_i + n\sigma_{rn}^2/t_i^2$$

but the problem with this is that you'd be weighting each measurment as if it had a different true rate, when in fact they all have the same true rate; this can lead to biases.

Let's go back to the correct formulation, which gives for the weighted mean:

$$< r >= \frac{\sum \frac{r_i}{(r/t_i + n\sigma_{rn}^2/t_i^2)}}{\sum \frac{1}{(r/t_i + n\sigma_{rn}^2/t_i^2)}}$$

In the case of all equal exposure times, then the σ^2 come out of the sums and cancel between the numerator and denominator, leaving

$$< r >= \sum \frac{r_i}{N}$$

i.e., an unweighted mean, just as we'd expect

In the case of zero readout noise we have

$$< r >= \frac{\sum \frac{r_i}{r/t_i}}{\sum \frac{1}{r/t_i}}$$

The r can now come out of the sums and cancel in the numerator and denominator, so the weighted mean is now independent of the true weight and becomes

$$< r >= \frac{\sum t_i r_i}{\sum t_i}$$

i.e., calculate the mean by weighting by the exposure times.

For the general case of unequal exposure times in the presence of readout noise, you can avoid biases by choosing any reasonable estimate of r for use in calculating the weights. For example, you could use the observed rate with the highest S/N as the basis for calculating weights. Or you could even iterate the solution a couple of times. If you were to do this, you'd find that the weighted mean hardly depends at all on what you choose for the estimate of r so long as it is close to the true r.

Understand how the distinction between sample variance and true variance can lead to biases when calculating a weighted mean, and how to overcome this.

4.2.4 Can you split exposures?

Although from S/N considerations, one can determine the required number of counts you need (exposure time) to do your science, when observing, one must also consider the question of whether this time should be collected in single or in multiple exposures, i.e. how long individual exposures should be. There are several reasons why one might imagine that it is nicer to have a sequence of shorter exposures rather than one single longer exposure (e.g., tracking, monitoring of photometric conditions, cosmic ray rejection, saturation issues), so we need to consider under what circumstances doing this results in poorer S/N.

Consider the object with photon flux S', background surface brightness B', and detector with readout noise σ_{rn}. A single short exposure of time t has a variance:

$$\sigma_S^2 = S'Tt + B'ATt + N_{pix}\sigma_{rn}^2$$

If N exposures are summed, the resulting variance will be

$$\sigma_N^2 = N\sigma_S^2$$

If a single long exposure of length Nt is taken, we get

$$\sigma_L^2 = S'TNt + B'ATNt + N_{pix}\sigma_{rn}^2.$$

The ratio of the noises, or the inverse ratio of the S/N (since the total signal measured is the same in both cases), is

$$\frac{\sigma_N}{\sigma_L} = \sqrt{\frac{S'TNt + B'ATNt + NN_{pix}\sigma_{rn}^2}{S'TNt + B'ATNt + N_{pix}\sigma_{rn}^2}}$$

The only difference is in the readout noise term! In the signal- or background-limited regimes, exposures can be added with no loss of S/N. However, if readout noise is significant, then splitting exposures leads to reduced S/N.

Understand under what circumstances you can split exposures without a noise penalty, and under what circumstances you cannot. Understand some of the reasons why you might want to split an exposure into shorter pieces.

4.3 Random errors vs systematic errors

So far, we've been discussing *random* errors. There is an additional, usually more troublesome, type of errors known as *systematic* errors. These don't occur randomly but rather are correlated with some, possibly unknown, variable relating to your observations, and can have the effect of not just adding spread around the true value that you are trying to measure, but actually measuring the wrong mean.

EXAMPLE : flat fielding

EXAMPLE : WFPC2 CTE

Note also that in some cases, systematic errors can masquerade as random errors in your test observations (or be missing altogether if you don't take data in exactly the same way), but actually be systematic in your science observations.

EXAMPLE: flat fielding, subpixel QE variations.

Note that error analysis from expected random errors may be the only clue you get to discovering systematic errors. To discover systematic errors, plot residuals vs. everything!

Understand the distinction between random and systematic errors. Understand how comparing observed scatter with expected scatter may be critical to discovering systematic errors.

5 Observing principles and tools

5.1 Time systems

Systems of time: see Naval observatory reference for a full listing of different types of time!

5.1.1 Solar time

Time tied to position of Sun. Note the distinction between *mean* solar time and *apparent* solar time (the "equation of time" and the analemma).

Most used solar time is Universal time. UT = local mean solar time at Greenwich = UT = Greenwich time = "Zulu". Tied to location of Sun, but averaged to "mean sun".

Local time: accounts for longitude of observer. For practically, legal time is split into time zones.

In detail, official time is kept by atomic clocks (International Atomic Time, or TAI), and coordinated UT (UTC) is atomic time with leap seconds added to compensate for changes in earth's rotation, where these are added to keep UTC within a second of solar time (UT1). See here for some details.

5.1.2 Sidereal time

Time based on position of stars, i.e. Earth's sidereal rotation period $\sim$ 23h 56m 4s. Local sidereal time is GMST (Greenwich mean sidereal time) minus longitude. At the vernal equinox (time in sky when Sun crosses the celestial equator as its declination is increasing), sidereal time = UT. Difference between UT and GMST is one rotation (day) over the course of a year, so about 2 hours per month.

Sidereal is relevant for position of stars: stars come back to the same position every sidereal day. As we'll see below, a given star crosses the meridian when the local sidereal time equals the right ascension of the star.

5.1.3 Calendars

Standard calendar is Gregorian, with leap years, etc. etc.

For astronomy, simpler to keep track of days rather than year/month/day. Most dates given by the Julian date (number of days since UT noon, Monday Jan 1 4713 BC). Variations include modified Julian date (JD - 2400000.5, fewer digits and starts at midnight), heliocentric Julian date (JD adjusted to the frame of reference of the Sun, so can differ by up to 8.3 minutes).

Note that repeating events are often described an an event *ephemeris*: $t_i(event) = t_0 + i(period)$.

The term ephemeris is also used to describe how the position of an object changes over time, e.g. planetary ephemerides.

5.2 Coordinate systems

SEDS website on astronomical coordinate systems

5.2.1 Celestial coordinate systems

- RA-DEC : tied to Earth rotation, longitude and latitude. Zero RA at vernal equinox.
- ecliptic: tied to plane of Earth rotation around the Sun. Zero ecliptic longitude tied to vernal equinox.
- galactic: tied to plane of the Milky Way

At vernal equinox, RA=12h crosses the meridian at midnight.

Note that for a celestial coordinate system tied to the Earth's rotation, coordinates of an object change over time because of the changing direction of the earth's axis: precession and notation. Because of this, coordinates are always specified for some reference equinox: J2000/FK5, B1950, ect.; if using coordinates to point a telescope, you need to account for this (but generally, telescope software does!). Note distinction between equinox and epoch, where the latter is relevant for objects that move (which everything does at some level!).

Transformations between systems straightforward from spherical trigonmetry.

Note the common usage of an Aitoff projection (equal areas) of the sky in celestial coordinates, with location of ecliptic and galactic plane. Software tools (Python, projection="aitoff" in subplot, IDL: aitoff and aitoff_grid in Astronomy users library).

5.2.2 Local coordinate systems

- Equatorial: HA-dec. $HA = LST - \alpha$. $LST = GMST - \mathrm{longitude}$. Note normal convention for HA is to get larger to the W, i.e. opposite of RA. Objects at zenith have $\delta =$ latitude of observer.
- Horizon: alt-az, or zd-az

Local coordinates are important for pointing telescopes! Note that there are various other effects that one has to consider for pointing a telescope at a source of known celestial position: proper motion, precession, nutation, "aberration of light", parallax, atmospheric refraction.

5.3 Finding positions of celestial objects

- SIMBAD: look up coordinates of many objects outside solar system by name, etc., also provides much other reference information
- VizieR catalog database. Database of astronomical catalogs, with search and download possibilities.
- NED: NASA extragalactic database: galaxies, etc.
- solar system ephemerides: JPL HORIZONS

5.4 Orientations of objects in the sky

Usually specified by *position angle*: angle of object in degrees from NS line, measured counterclockwise.

An important observational position angle for spectroscopy: *parallactic angle*, the position angle of the line from zenith to horizon

5.5 Observability

Obviously, to observe an object, one requires that it is visible above the horizon. In general, one would like to observe objects through the shortest possible path through the Earth's atmosphere, i.e., when they are *transiting* (crossing the meridian, HA=0). The more atmosphere the light goes through, the more losses due to atmospheric absorption/scattering (more severe at shorter wavelengths), and the more image degradation from atmospheric seeing. Of course, it doesn't make sense to wait for an object to transit if you don't have anything else to do in the meantime; efficient use of telescope time is primary concern. Generally, most observers attempt to observe at airmasses (*ask if you don't know what this means*!) less than 2, i.e. within 60 degrees of the zenith. Once you hit an airmass of 3, the object is rapidly setting (except at very high declination). Of course, for some solar system objects (objects near the Sun), one has no choice but to observe at high airmass.

Note that HA gives some indication of observability, but that higher declination objects can be observed to higher HA than lower declination objects. Roughly, at the celestial equator, an HA of 3 hrs is about an airmass of 2, and in many cases, one doesn't want to go much lower in the sky.

Another issue with observability has to do with the Moon, since it is harder to see fainter objects when the sky is brighter (do you understand in detail why?). Moon brightness is related to its phase, and to a lesser extent, to distance from your object. Of course, if the Moon is below the horizon, it does not have an effect. So for planning observations of faint objects, one also has to consider Moon phase and rise/set times. Note that the sky brightness from the Moon is a function of wavelength, and at IR wavelengths, it is not a very signficant contributor to the total sky brightness: so often, telescopes spend bright time working in the IR.

5.6 Tools

Here are some useful software tools to do tasks related to coordinate systems and observability. There are certainly other tools out there. Anything that accomplishes the desired tasks adequately is fine to use! Just make sure you're not limited by the tools that you choose. These are available on the Astronomy Linux cluster; you can probably install them on your laptop, but they will probably not be there by default.

- skycalc/skycalendar: text based programs, installed on our Linux cluster (link is to source code if you wish to install on your laptop). skycalendar gives daily almanac, position of moon, etc. skycalc allows you to enter coordinates of an object and obtain observability information for any specified date. Other features included as well: coordinate transformation, position of planets
- JSkyCalc (java -jar /home/local/java/JSkyCalc.jar): JAVA implementation of skycalc, also installed on the Astronomy cluster (and available for download)
- Python: astropy.coordinates: tools for coordinate systems and transformations, astroplan: python observation planning tools
- IDL: euler in Astronomy users library
- WCSTOOLS: full set of useful coordinate system programs, e.g. coordinate system transformation (command skycoor). Largely useful for use with coordinate system information in image headers (more later!). Installed on the astronomy cluster.

5.7 Using telescopes

Generally it is usually fairly straightforward to use an astronomical telescope. Most of the time after arrival at an observatory can be spent checking the instrument and detector performance rather than checking the telescope performance. You should carefully consider, however, how to maximize your efficiency at the telescope; telescope time is expensive and hard to come by.

Before going to a telescope, you might consider the following checklist of things to do:

- Learn how the telescope is commanded. In particular, you may wish to find out whether scripts can be written (or are already available) to do routine motions on the sky (e.g. dither back and forth between positions, especially important for IR observing)

- Find out whether user object catalogs can be used, and if so, find out the format and prepare files. It can save a lot of time and headaches to have your coordinates preloaded in a file and save the agony of typing numbers in the middle of the night and getting them wrong.

- Find out the pointing accuracy of the telescope. You may not necessarily be able to count on the telescope pointing exactly at the coordinates that you tell it to go to. For fainter objects, it is highly recommended that you bring finding charts with you to the telescope. Note that it is now fairly easy to make finding charts from the Digitized Sky Survey (from Palomar and ESO plates). The program getimage is available for your use; this program can extract FITS images of arbitrary size from the digitized sky survey, which we have on CDROM. Any image processing package should be capable of reading these and producing hardcopy pictures. Another easy interface available over the WWW can be found at http://skyview.gsfc.nasa.gov.

- Find out the guiding performance of the telescope. For what exposure times is guiding required? If you will be taking exposures which need to be guided, how does one find guide stars? Can time be saved by finding guide stars in advance?

- Plan your observations. You generally want to observe objects at the lowest possible airmass. In combination with your scientific priorities, this will set the order in which you can plan to make observations. You can figure out transit times by considering the sidereal times for your night and the right ascension of your targets using

  $$HA = LST - \alpha$$

  You can get observing calendars using the program skycalendar and compute airmass tables, etc., for specific targets using the program skycalc (i'm sure lots of other programs are also available for these tasks). If you are doing spectroscopy, consider whether differential refraction will be important and whether you can mitigate its effects by observing objects around the parallactic angle (which, of course, changes over the course of the night!).

  Remember to consider the calibration observations you will need to take (we'll talk more about this later). Also remember to plan for different possible conditions. For example, if you program requires photometric weather, what will you do if it's not photometric? what if the seeing is horrible, etc.?

- Be prepared to be able to analyze image quality (e.g., FWHM) and focus the telescope. One of the first things to be done just after dark is focussing the telescope. This generally involves taking images at a range of focus settings and comparing them to determine the best focus. One should be prepared with software for analyzing image quality to make this determination - also, perhaps, software for looking at all images simultaneously. Generally, the focus position is encoded somehow so one gets a quantitative measure of the secondary location. One should be aware, however, of the possibility of slack in the gears controlling the focus mechanism, which can make the focus not repeat even when the readout position is the same; because of this, it is generally wise to always move to a focus position from one direction.

  While focussing, one can generally also get an idea of the quality of the seeing of the night. Remember that seeing varies from frame to frame, and because of this, multiple exposures even at the same focus can look very different. To minimize seeing effects, one may wish to choose a focus star on which exposures of several seconds can be made: for a brighter star with very short exposures, seeing changes may confuse you. Clearly, however, one doesn't want to use a very faint star because one would like to get the focussing procedure over as quickly as possible so you can get on with your science. Remember, however, the signal-to-noise gains are substantial for a more concentrated image, so it will be worth your while to do a good job: if you rush it, you may regret it later when you have more time to notice how blurred your images are!

  You also need to remember that the focus is likely to change throughout the night as the temperature changes. So continue to inspect your images as you take them, and if the quality appears to be degrading, you should redo a focus run. Most likely, the telescope focus will consistently change in one direction (as it gets colder) and you may even be able to get a good estimate of how much it changes as a function of the temperature with experience. Which direction focus goes is a good thing to write down at a telescope, as you can save significant time during mid-night focus changes if you already know which direction you need to go (but always beware of someone coming and rewiring the focus motor/control since the time of your last run!). Determining the correct balance of time spent optimizing (i.e. focussing) vs. taking science data can be tricky, and likely depends on the nature of your program.

  One may wish to quickly inspect an out-of-focus image for signs of large aberrations in the system. Almost certainly, nothing will be done about these immediately, but if the image quality is poor enough, it may be possible to have something (e.g., alignment) done the next day, so you still may possibily help your observing run, or certainly, you will help subsequent observers. At least, if something seems strange, you should let someone know so they can judge for themselves if there is really a problem.

Overall, this is a key point; you need to be vigilant to look for peculiarities in your data, and if you see something that hasn't previously been documented or that you don't understand, you need to ask someone about it rather than just assume that it is "normal"!

5.8 Plannning observing

1. Prepare targets: coordinate files, finding charts, etc
2. Understand S/N requirements per target, and implications for desired number of counts. What exposure times do you expect, and how will you check for each exposure to make sure you are not exposing for too long or too short?
3. Understand calibration requirements, and plan for calibrations exposures
4. Observing logs: summary program informtion, weather information, calibration data, seeing information, exposure information. COMMENTS are critical. READABILITY is critical (no superwide rows!).

5.9 Digital imaging and display

5.10 Image display

Much astronomical data is in the form of 2D images. It is critical to understand how to display such data and be able to see all of the information it contains. This is an issue because in most cases, the data will contain more information that can be displayed on a screen at any one time. There are two issues: spatial resolution and, probably more importantly, *dynamic range*.

5.10.1 Spatial resolution

Note that many modern detectors have larger pixel dimensions than many computer displays. This means that it's not possible to see all of the pixels at one time; you can either see a subframe of the entire image at full spatial resolution, or the entire image at reduced spatial resolution; generally software does reduced spatial resolution by displaying every other, every 3rd, every 4th, etc. pixel value, so it is possible to miss features!

5.10.2 Dynamic range: brightness and contrast

Most image displays provide only 8-bits of display range in intensity, giving only 256 possible intensities; I don't think the human eye can distinguish many more (or even that many) with any reliability. Most astronomical images can have up to 16-bits of dynamic range, 256 **times** more levels! Any image with more dynamic range must somehow be compressed into 8-bits before it can be displayed. This can be done either by sampling the true image coarsely (in intensity), which allows viewing of the whole dynamic range but can lead to the apparent loss of intensity detail, or by fully sampling only a part of the true image range, which leads to the loss of ability to view detail outside the chosen range, or by something in between. Most packages will use some default algorithm to make this choice automatically, so you have to be careful to understand what is being done, and what information might be lost in what you are looking at. Any decent display package will give you control over how to display the image, and you need to understand in detail how you can see different things in images when you display them in different way. To be able to choose reasonable display parameters, you will need to know something about the intensity values in your images, so most display packages will allow you to directly see pixel intensities. This is also useful so you can make sure that the values are somewhere around the levels that you expect!

Image scaling parameters are generally specified by a low and a high data value (or a low value and a range) which give the limits in the true data which will be scaled into 8-bits. In old-imaging parlance, the *brightness* is set by the choice of value that will correspond to the darkest pixel, and the *contrast* is set by the difference between the darkest and lightest pixel.

Common choices for automatic scaling might be to display an image such that the pixel with the lowest data value in an image will appear black, and the pixel with the highest data value will appear white; this is sometimes called 100% scaling. However, many images can have defects which might appear as very low or very high data values, so often this choice will set display parameters non-optimally. Alternative autoscaling might be determined from the low and high data values of the middle 99% of the data values (i.e. exclude the 0.5% lowest and highest data values), or 98%, etc., etc.

To change the display scaling factors, the data values must be rescaled and the image redisplayed. On modern machines, this is generally still quite fast. However, there is a faster way to *partially* get the same result, see below.

5.10.3 nonlinear scalings

Note you can also use a nonlinear scaling to sample a larger (or smaller) range. Example: logarithmic, square root scaling, asinh scaling.

5.10.4 color maps and pseudocolor

Once an intensity subsection is chosen, it can be displayed with any choice of "color map", which specifies the colors to be assigned to each of the display levels. These can be various shades of grey (greyscale) or some other color, or some arbitrary color scheme (pseudo-color). Note that most packages allow the user to manipulate the color table, allowing users to change the contrast and brightness of a displayed subsection; for this reason, it is usually reasonable to chose a range with a significantly larger range than 256 data values.

Most packages will allow the user to inspect individual data values based on a cursor location. Beware, however, of packages which give data readout based on scaling parameters and 8-bit display number only: these are unable to give correct values outside of the scaled region of the image.

The color map is implemented at a lower level and can generally be changed very rapidly. One use of this is to "stretch" or "roll" the color map to change the brightness and/or the contrast in the image.

5.10.5 true color images

True color images obviously require information about colors of the objects in the picture, so they cannot be made from an image taken through a single filter. Generally, three independent filters are used to create true color images, e.g. RGB images. The image in each individual filter must be properly scaled if one wants to make the true color image match what would be seen with the eye, i.e. correct white balance.

One can also use images in multiple filters to construct "pseudo-true" color images, e.g. emission line regions in one color, continuum in another, etc.

5.10.6 other display functions

Other useful display tools include zoom, blink, interactive image analysis (peak, valley, fwhm, etc), marking of objects, etc.

5.11 Standalone display tools

- DS9: standalone display tool, but also most commonly used display tool with IRAF (an astronomical image processing package).
- XIMTOOL: another display tool that can be used with IRAF.
- GAIA: also includes image processing routines

These are all installed on the Astronomy cluster; you should be able to install them on your laptop via the links above if you want to.

Of course, any image processing package will generally include a display tool as part of the package, and we will use these extensively. But, in dicussing principles of image display, perhaps it's best to start with standalone display tools. These can be very useful for quick-look analysis.

5.12 Digital imaging file formats

FITS format. Two parts in one file: header plus data. Header contains ASCII information, data is in binary format. Beware that headers must conform to specific lengths: don't use an editor on a FITS file! Headers have a small amount of required information, plus there are lots of possibilities for optional information.

5.13 Gain and readout noise

5.13.1 Digitization

In array detectors, after the charge is collected and read out, it is sent through a chain of electrons which digitizes the signal, often after amplifying it. The digitization is made by a device known as an A/D convertor; these work by comparing an input signal with a set of reference voltages which successively differ by factors of two. Thus an input signal is translated into a series of bits depending on whether the input voltage exceeds a series of reference voltages. Typical A/D correctors in use in astronomy consider 16-bits. The digital signal which comes out of the CCDs is variously referred to as counts, digital numbers (DN), or analog-to-digital units (ADU). The number of output counts is related to the number of input counts by a constant which depends on the amplification in the electronics. The amplification factor is known by most people as the gain, but astronomers define the gain of a device by the number of input electrons divided by the number of output counts (i.e., the inverse gain); this "astronomers" gain is specified in units of e^-/DN. Because the number which we receive from the electronics chain differs from the number of input electrons (i.e, the number of detected photons), the calculation of noise must take this into account. The photon counting noise (rms) is given by the square root of the number of detected photons. The number of detected photons is given by GC, where G is the (inverse) gain and C is the number of detected counts. Consequently, the noise *in electrons* is $\sqrt{GC}$, and in units of counts is given by $\sqrt{C/G}$. This is apart from readout noise; the latter is usually specified in units of electrons, giving a total noise in electrons of $\sqrt{GC + \sigma_{rn}^2}$, or, in units of counts, by $\sqrt{C/G + \sigma_{rn}^2/G^2}$.

A/D converters can only measure a positive incoming signal. At low light levels, the true input signal can be negative in the presence of readout noise. To avoid trucation of the negative signals, a constant voltage, called the bias, is added to the signal before it passes through the A/D. This bias must later be removed to preserve the correct count ratios between different sources; this is generally accomplished in CCDs by using the overscan region of the image.

A/D convertors can introduce small systematic errors in recorded count rates if the reference voltages are not carefully controlled.

Understand how digitzation works. Know what the gain is. Understand how to calculate noise in both units of detected electrons and in counts.

5.13.2 Determining gain and readout noise

Consider an ensemble of measurements taken at a light level L. The noise in this ensemble should be $\sigma^2 = LG + \sigma_{rn}^2$, where G is the gain and rn is the readout noise in electrons, and σ is the noise in electrons.

If you do this at a lot of different light levels, then you can plot σ^2 vs L, and the slope should give you G and the intercept rn^2. However, remember that if you compute σ from the images, this gives σ in DN, so the slope will give you $1/G$. This test is also excellent for checking the basic performancer of a detector. Deviations from nonlinearity can also usually be seen on such a plot.

However, in its most straightforward application the test is very time consuming and hard to analyze: you have to take many exposures at each different light level, and then determine a gain and readout noise for each pixel and look at them all. It is much easier just to use the set of all pixels as your ensemble at each light level. However, you can't do this directly, because each pixel may have a different sensitivity and different fixed pattern noise, so you're not measuring a true ensemble. If there is significant variation of sensitivity than you can't use a whole area at all, because the noise properties will vary across the area. You can avoid these problems by working with the difference between pairs of observations: if the light level is the same in the two images, then you'll be left with an image that only has noise.

Specifically, take a pair of images and form the difference. The expected noise is

$$\sigma^2 = 2(LG + \sigma_{rn}^2)$$

where L is the light level, G is the gain, and σ is the noise in electrons. Since $\sigma(electrons) = G\sigma_{DN}$, we have

$$\sigma_{DN}^2 = 2(\frac{L}{G} + \frac{\sigma_{rn}^2}{G^2})$$

. You can directly measure $\sigma(DN)$ from your difference image. Make sure to do it over a region which doesn't vary significantly in light level. Now repeat the measurement at a variety of light levels and plot $\sigma^2(DN)$ vs L. If you fit a line through this, the slope is $2/G$ and the intercept is $2\sigma_{rn}^2/G^2$. If a straight line doesn't fit the points, then there is some sort of problem, which you should probably track down.

You can abbreviate this test if you just want to get an estimate of the gain and readout noise. First, take a pair of bias frames. These have light level of zero, so the noise from the difference just gives you $\sqrt{2}\sigma_{rn}$. (Note that you still need to take a pair in case there is superbias structure). Then take a pair at a high light level; at this level the readout noise is probably negligible, and you can determine the gain from

$$G = 2\frac{L}{\sigma_{DN}^2}$$

Understand how to determine gain and readout noise.

5.14 Basic data reduction

Overscan subtraction.

Flat fielding.

6 Effects of the earth's atmosphere

The earth's atmosphere has several different effects: it emits light, it absorbs light, it shifts the apparent direction of incoming light, and it degrades the coherence of incoming light, leading to degradation of image quality when collecting light over a large aperture.

6.1 Night sky emission

We first consider emission from the sky, which comes from several sources:

- air glow
- zodiacal light (outside of our atmosphere!)
- sunlight
- moonlight
- aurorae
- light pollution
- unresolved stars and galaxies (outside of our solar system!)
- thermal emission: sky, telescope and dome!

Many of these source are emission line sources, not contiunuum sources, as shown in Figure 3.

How bright are these sources?

- air glow : strong in lines
- zodiacal light : $m_V \sim 22.2 - 23.5$ mag/square arcsec depending on ecliptic latitude (and to a lesser degree on ecliptic longitude). Note that this exists even in Earth orbit!
- sunlight : small away from twilight, depending on distance sun is below horizon. Note different definitions of twilight: civil (6 degrees), nautical (12 degrees), and astronomical (18 degrees). By astronomical twilight, there is essentially no contribution. However, much useful observing might be done before this!

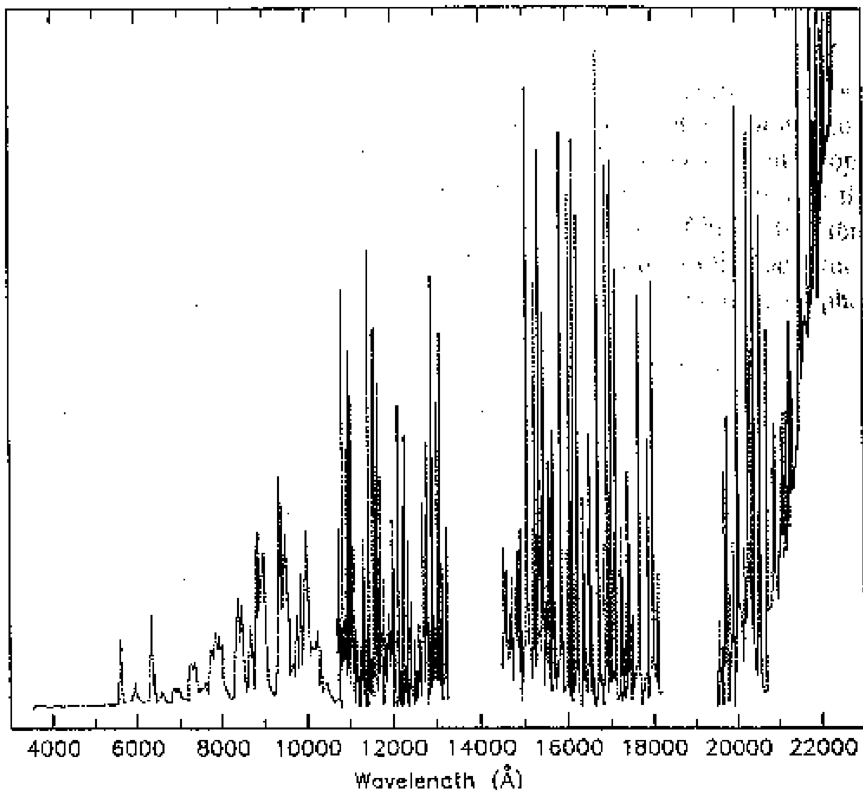


Fig 3. The night sky from 4000 A to 2.2 μm (courtesy D. Crampton), showing the strong night sky emission. The rise in the integrated night sky background is actually less frightening: The sky $(I\text{-}H)_{AB} \sim 3.3$, comparable to the rise between B and I, where the sky $(B\text{-}I)_{AB} \sim 2.5$.

Figure 3: Night sky emission, from article by Lilly

- moonlight : variable, can be very bright! ∼10 times brighter at full moon.

lunar age (days)	U	B	V	R	I (mag/square arcsec)
0	22.0	22.7	21.8	20.9	19.9
3	21.5	22.4	21.7	20.8	19.9
7	19.9	21.6	21.4	20.6	19.7
10	18.5	20.7	20.7	20.3	19.5
14	17.0	19.5	20.0	19.9	19.2

  The significance of moonlight, especially in the optical, gives rise to splitting nights into dark, grey, and bright time, where dark time means no moon above horizon, grey time something like less than 50% moon above horizon, otherwise bright time.

- light pollution : strong in distinct lines

Optical:

For broadband work, for example in the V band, $m_{sky} \sim 22$ mag/arcsec at good site so we switch over to background limited around m=22 for good image quality, switch over around m=20 for poorer image quality. Consequently, image quality matters for faint objects!! Moonlight is very significant, hence faint optical imaging requires dark time.

Optical spectroscopy: sky emission generally not much of a problem except around lines so long as moon is down (or work on bright objects); low dispersion observations can be background-limited for long exposures, but at higher dispersion or shorter exposures, spectroscopy is often readout-noise limited.

IR:

Most of the emission in the near-IR is from emission lines of OH, the so-called "OH forest." For broadband work in H band, $m \sim 13.5$ mag/arcsec; in K band, $m \sim 12.5$ mag/arcsec. So for all except bright objects, we're background limited. This leads to some fundamental differences in data acquisition and analysis between the near-IR and the optical. For infrared spectra, it's harder to estimate S/N : depends on where your feature is located. Moonlight is not very significant, hence much IR work is done in bright time.

Farther in the IR (5μ+), thermal emission from the sky dominates and is extremely bright. In fact, when working at wavelengths with thermal background, the exposure time is often limited by the time it takes to saturate the detector with background (sky) light... in seconds or less!

Sky brightness from most sources *varies* with time and position in the sky in an *irregular* fashion. Consqeuently, it's essentially impossible to estimate the sky *a priori*: sky must be determined from your observations, and if your observations don't distinguish object from sky, you'd better measure sky close by in location and in time: especially critical in the IR. See some IR movies; spectral movie from ESO/Paranal see here.

Know the spectral energy distribution of the main sources of sky emission. Know roughly how bright these sources are in a broadband measurement. Understand the effect of moonlight on observations in the optical vs. those in the near-infrared.

6.2 Transmission of atmosphere

Earth's atmosphere doesn't transmit 100% of light. Various things contribute to the absorption of light:

- scattering, e.g., Rayleigh scattering off molecules
- aerosols: scattering off larger particles
- absorption by variety of molecules:
  - ozone
  - H_2O
  - O_2
  - CO_2
  - N_2O
  - CH_4

All are functions of wavelength, time to some extent, and position in sky.

6.2.1 Sources of extinction

In the optical part of the spectrum, extinction is a roughly smooth function of wavelength and arises from a combination of ozone, Rayleigh scattering, and aerosols, as shown in Figure 4.

The optical extinction can vary from night to night or season to season, as shown in Figure 5. Of course, this is showing the variation over a set of photometric nights; if there are clouds, then the level of variation is much higher! Because of this variation, you must determine the amount of extinction on each night separately if you want accuracy better than a few percent (even for photometric nights!). Generally, the *shape* of the extinction curve as a function of wavelength probably varies less than the amplitude at any given wavelength. Because of this, one commonly uses *mean extinction coefficients* when doing spectroscopy where one often only cares about relative fluxes. To first order,

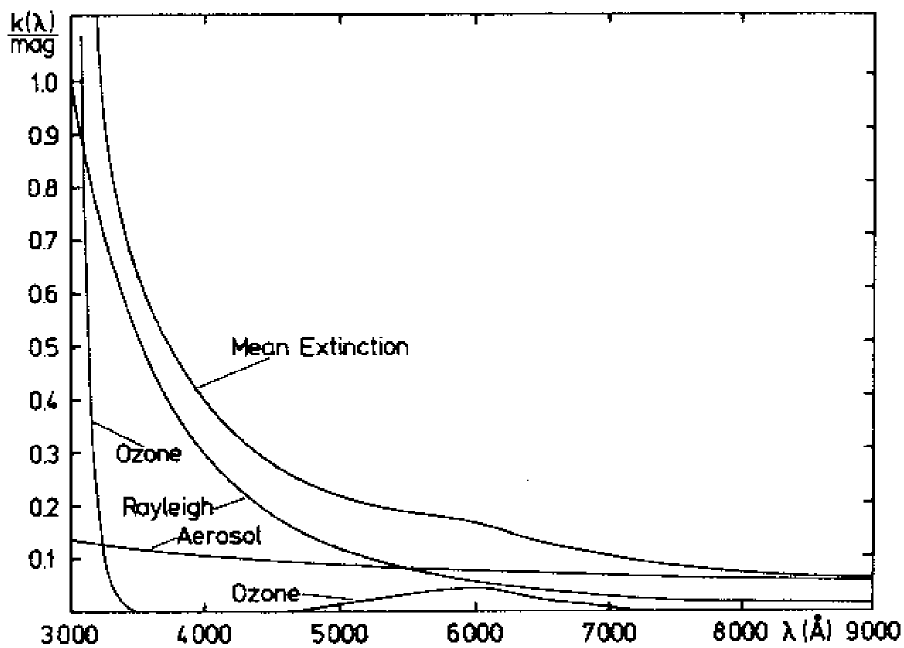


Fig. 1. Mean vertical extinction at Flagstaff, Arizona, in May-June 1976. The assumed ozone and Rayleigh contributions are shown separately

Figure 4: Extinction sources in the optical (Tug, Lockwood, and White,).

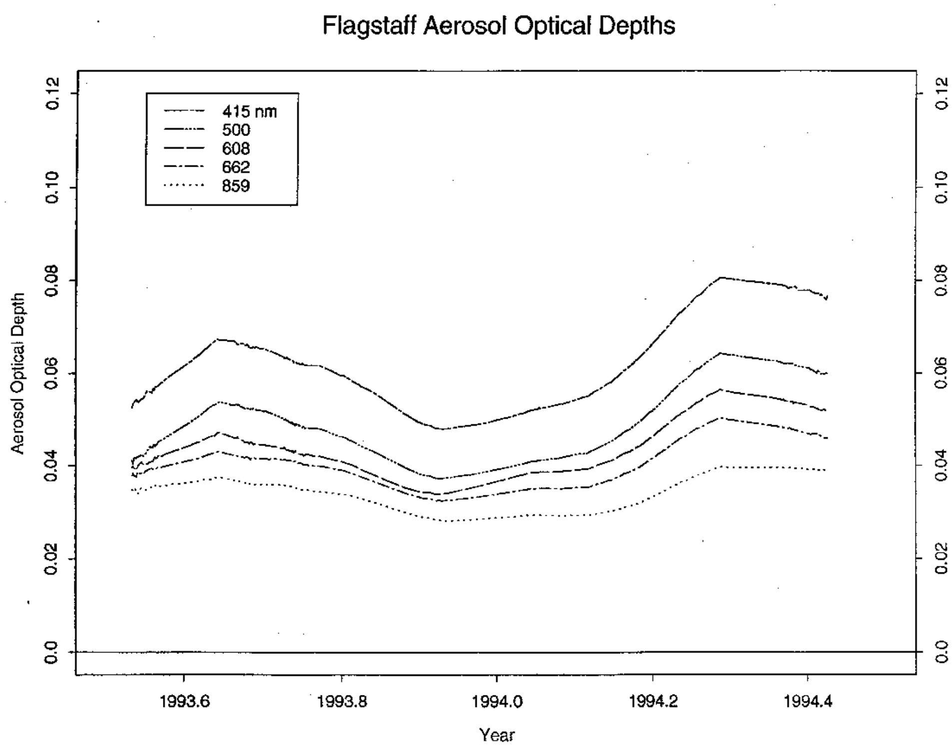


Figure 5: Variation in aerosol optical depth with time (Lockwood)

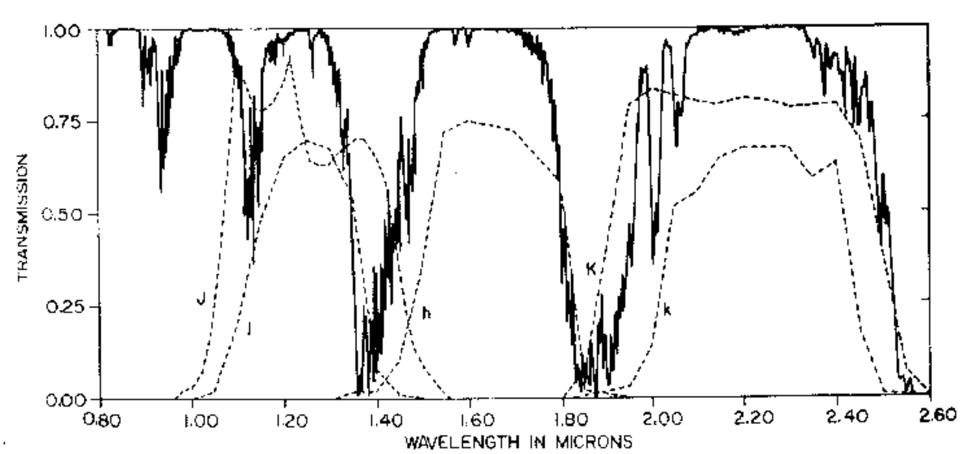


FIG. 1.—The adopted Kitt Peak summer atmospheric transmission is plotted as a function of wavelength. The sensitivity functions of the Johnson J and K filters (Johnson 1965b) and the currently used Kitt Peak J, H, K filters identified as j, h, and k; are also plotted.

Figure 6: Atmospheric transmission in the near-IR (Manduca & Bell ...).

the extinction from clouds is "gray", i.e. not a function of wavelength, so relative fluxes can be obtained even with some clouds present.

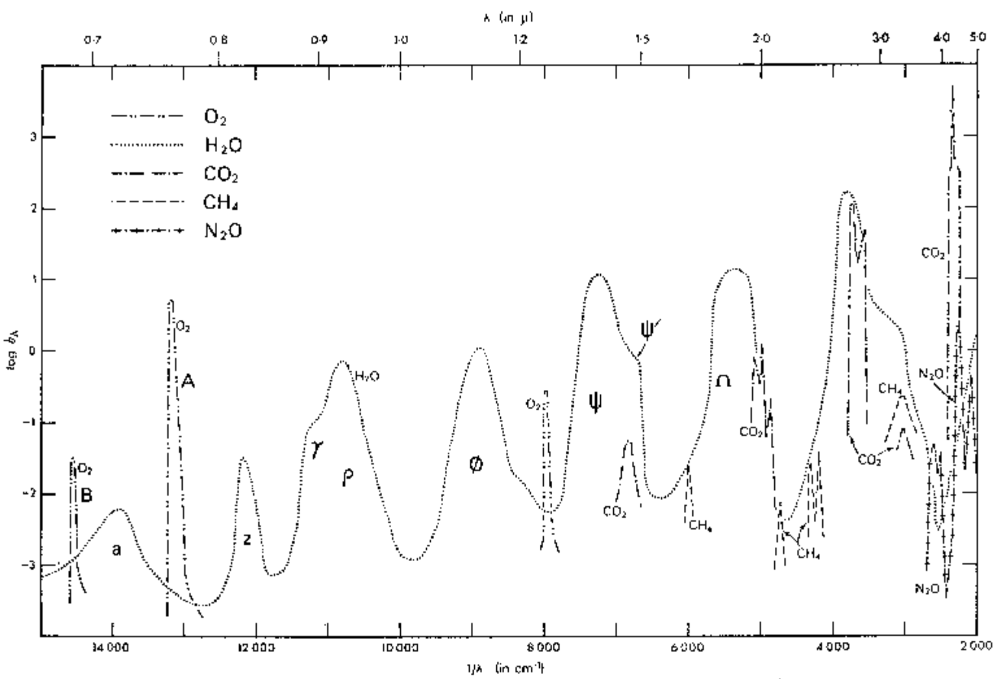
There is significant molecular absorption in the far-red part of the optical spectrum, in particular, note the A (7600) and B (6800) atmospheric bands from O_2.

In the infrared, the extinction does not vary so smoothly with wavelength because of the effect of molecular absorption. In fact, significant absorption bands define the so-called infrared windows (yJHKLM), as shown in the near IR in Figure 6. At longer wavelengths, the broad absoprtion band behavior continues, as shown in Figure 7. In this figure, $transmission = f(b_\lambda l)$ where l is path length (units of airmass):

$b_\lambda l$	f
-3	1
-2	0.97
-1	0.83
0	0.5
1	0.111
2	0.000

The L band is at 3.5μ, M band at 5μ.

Note that even within the IR "windows", there can still be significant telluric absorption features, e.g. from CO_2, H_2O, and CH_4. When doing IR spectroscopy, one needs to be aware of these and possibly attempt to correct for them, taking care not to confuse them with real stellar features!



Infrared band absorption of atmospheric gases

Figure 7: Atmospheric transmission in the IR (Allen, Astrophysical Quantities).

Know the spectral energy distribution of the main sources of telluric absorption. Understand how H_2O absorption leads to "windows" in the near-infrared.

6.2.2 Airmass and zenith distance dependence

Clearly, if the light has to pass through a larger path in the Earth's atmosphere, more light will be scattered/absorbed; hence one expects the least amount of absorption directly overhead (zenith), increasing as one looks down towards the horizon.

Definition of airmass: path length that light takes through atmosphere relative to length at zenith: $X \equiv 1$ vertically (at $z = 0$). Given the zenith distance, z, which can be computed from:

$$\sec z = (\sin\phi \sin\delta + \cos\phi \cos\delta \cos h)^{-1}$$

where ϕ is the latitude, δ is the declination, and h is the hour angle ($=$ local sidereal time$-\alpha$, where α is the right ascension), we have

$$X \sim \sec z$$

which is exactly true in the case of a plane parallel atmosphere. Since the earth's atmosphere is not a plane, the plane parallel approximation breaks down for larger airmasses. For $X > 2$, a more precise formula is needed, the following gives a higher order approximation:

$$X = \sec z - 0.0018167(\sec z - 1) - 0.002875(\sec z - 1)^2 - 0.0008083(\sec z - 1)^3$$

How much light is lost going through the atmosphere?

Consider a thin sheet of atmosphere, with incident flux F, and outcoming flux $F + dF$. Let the thin sheet have opacity $\kappa = N\sigma$, where N is the number density of absorbers/scatterers, and σ is the cross-section/absorber-scatterer.

$$dF = -\kappa F dx$$

$$F = F_{top} e^{-\int \kappa dx} \equiv F_{top} e^{-\tau}$$

where τ is the *optical depth* of atmosphere, which parameterizes how much light is lost.

If the optical depth through the atmosphere is just proportional to the physical path length (true if same atmospheric structure is sampled in different directions), then

$$\tau(X) \sim \tau_0 X$$

where τ_0 is the optical depth at the zenith.

$$F = F_{top} e^{-\tau_0 X}$$

Expressing things in magnitudes, we have:

$$m = m_0 + 1.086\tau_0 X$$

We can define the *extinction coefficient* k_λ:

$$m_0 = m + k_\lambda X$$

$$k_\lambda \equiv -1.086\tau_0$$

so the amount of light lost in magnitudes can be specified by a set of extinction coefficients. Note by this definition, the extinction coefficient will be negative; others may use the opposite sign convention (e.g. defining $m_0 = m - k_\lambda X$). Of course, use of the scaling of τ or k with airmass assumes *photometric weather*!!

Since the amount of absorption from the Earth's atmosphere varies with time, to apply this characterization requires measurement of the extinction coefficient(s) on each given night in which one wants to apply them. The basic principle is straightforward: if one observes a star at a range of airmasses as it moves across the sky, one can determine the extinction coefficient in each filter by fitting (e.g., in a least squares sense) for slope of the relation between airmass and the measured instrumental mangitude, and this gives the extinction coefficient.

In practice, the determination of extinction is often combined with the determination of the zeropoint and the transformation coefficients as discussed earlier. The basic equations are:

$$M_i = m_i + k_i X + t_i(M_i - M_j) + z_i$$

where the subscripts refer to measurements in different filters, and where I have ignored second order coefficients. The advantage of combining the equations is that you can use the information about the known magnitudes of the standard stars for the extinction term, so you can combine observations of different standards at different airmasses to derive the extinction coefficient and do not need to observe the *same* star at multiple airmasses.

Know what airmass is. Know the functional dependendence of atmospheric transmission (flux and magnitudes) and what typical values of the extinction coefficient are.

6.3 Atmospheric refraction

The direction of light as it passes through the atmosphere is also changed because of refraction since the index of refraction changes through the atmosphere. The amount of change is characterized by Snell's law:

$$\mu_1 \sin\theta_1 = \mu_2 \sin\theta_2$$

Let z_0 be the true zenith distance, z be the observed zenith distance, z_n be the observed zenith distance at layer n in the atmosphere, μ be the index of refraction at the surface, and μ_n be the index of refraction at layer n. At the top of the atmosphere:

$$\frac{\sin z_0}{\sin z_N} = \frac{\mu_N}{1}$$

At each infinitessimal layer:

$$\frac{\sin z_n}{\sin z_{n-1}} = \frac{\mu_{n-1}}{\mu_n}$$

as so on for each layer down to the lowest layer:

$$\frac{\sin z_1}{\sin z} = \frac{\mu}{\mu_1}$$

Multiply these to get:

$$\sin z_0 = \mu \sin z$$

from which we can see that the refraction depends only the index of refraction near the earth's surface.

We define astronomical refraction, r, to be the angular amount that the object is displaced by the refraction of the Earth's atmosphere:

$$\sin(z + r) = \mu \sin z$$

In cases where r is small (pretty much always):

$$\sin z + r \cos z = \mu \sin z$$

$$r = (\mu - 1) \tan z$$
$$\equiv R \tan z$$

where we have defined R, known as the "constant of refraction".

A typical value of the index of refraction is $\mu \sim 1.00029$, which gives R = 60 arcsec (red light).

The direction of refraction is that a star apparently moves towards the zenith. Consequently in most cases, star moves in both RA and DEC:

$$r_\alpha = r \sin q$$

$$r_\delta = r \cos q$$

where q is the *parallactic* angle, the angle between N and the zenith:

$$\sin q = \cos \phi \frac{\sin h}{\sin z}$$

Note that the expression for r is only accurate for small zenith distances ($z <\sim 45$). At larger z, can't use plane parallel approximation. Observers have empirically found:

$$r = A \tan z + B \tan^3 z$$

$$A = (\mu - 1) + B$$

$$B \sim -0.07"$$

but these vary with time, so for precise measurements, you'd have to determine A and B on your specific night of observations.

Why is it important to understand refraction? Clearly, it's relevant for pointing a telescope, but this is generally always automatically handled in the telescope pointing software. If you're just taking images, then the stars are just a tiny bit moved relative to each other, but who really cares? One key issue today is the use of multiobject spectrographs, where slits or fibers need to be placed on objects to accuracies of a fraction of an arcsec. For small fields, refraction isn't too much of an issue, but for large fields, it can be .... note SDSS plates!

The other *extremely* important effect of refraction arises because the index of refraction varies with wavelength, so the astronomical refraction also depends on wavelength:

λ	R
3000	63.4
4000	61.4
5000	60.6
6000	60.2
7000	59.9
10000	59.6
40000	59.3

This gives rise to the phenomenon of *atmospheric dispersion*, or *differential refraction*. Because of the variation of index of refraction with wavelenth, every object actually appears as a little spectrum with the blue end towards the zenith. The spread in object position is proportional to $\tan z$.

This effect is critical to understand for spectroscopy when using a slit or a fiber, since the location of an object in the sky depends on the wavelength. If you point to the location at one wavelength, you can miss an another wavelength and the relative amount of flux you collect will be a function of wavelength, something you may want to take into account if you're interested in the relative flux (continuum shape) over a broad wavelength range. Note the consequent importance of the relation between the orientation slit orientation and the parallactic angle: a slit aligned with the parallactic angle will not lose light as a function of wavelength, but otherwise it will. However, for a slit at the parallactic angle, be careful about matching up flux at different wavelengths for extended objects!

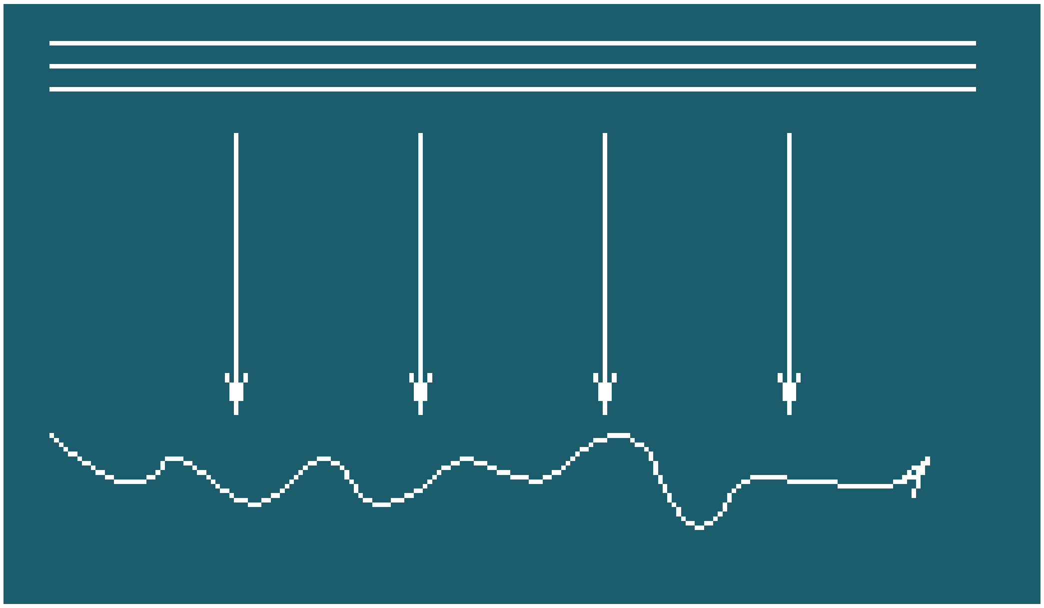


Figure 8: The effect of the Earth's atmosphere on a plane wavefront.

Understand the concept of atmospheric refraction and know its dependence on airmass and approximate amplitude. Understand differential refraction and its effect on slit/fiber spectroscopy, and how it can be mitigated in slit spectroscopy by aligning the slit with the parallactic angle.

6.4 Seeing: theory and practice

References: Coulson, ARAA 23,19; Beckers, ARAA 31, 13; Schroeder 16.II.

Generally, a perfect astronomical optical system will make a perfect (diffraction-limited) image for an incoming plane wavefront of light. The Earth's atmosphere is turbulent and variations in the index of refraction cause the plane wavefront from distant objects to be distorted, as shown in Figure 8.

These cause several astronomical effects:

- scintillation, which is amplitude variations, which typically vary over scales of cm: generally very small for larger apertures, and

- seeing: positional changes and image quality changes. The effect of seeing depends on aperture size: for small apertures, one sees a diffraction pattern moving around, while for large apertures, one sees a set of diffraction patterns (speckles) moving around on scale of ~ 1 arcsec; see also here. These observations imply:
  - local wavefront curvatures flat on scales of small apertures, and
  - instantaneous slopes vary by $\sim$ an arcsec.

The time variation scales are several milliseconds and up.

The effect of seeing can be derived from theories of atmospheric turbulence, worked out originally by Kolmogorov, Tatarski, Fried. Here, I'll quote some pertinent results, without derivation.

A turbulent field can be described statistically by a *structure function*:

$$D_N(x) == < |N(r+x) - N(r)|^2 >$$

where x is separation of points, N is any variable (e.g. tempereature, index of refraction, etc), r is position.

Kolmogorov turbulence gives:

$$D_n(x) = C_n^2 x^{2/3}$$

where C_n is the refractive index structure constant. From this, one can derive the phase structure function at the telescope aperture:

$$D_\phi(x) = 6.88 \left(\frac{x}{r_o}\right)^{5/3}$$

where the coherence length r_0 (also known as the Fried parameter) is:

$$r_0 = 0.185 \left(\lambda^{6/5}\right) \left(\cos^{3/5} z\right) \left[\int (C_n^2 dh)\right]^{-3/5}$$

where z is zenith angle, λ is wavelength. Using optics theory, one can convert D_ϕ into an image shape.

Physically, r_0 is (roughly) inversely proportional to the image size from seeing:

$$d \sim \lambda / r_0$$

as compared with the image size from diffraction-limited images:

$$d \sim \lambda / D.$$

Seeing dominates when $r_0 < D$; a larger r_0 means better seeing.

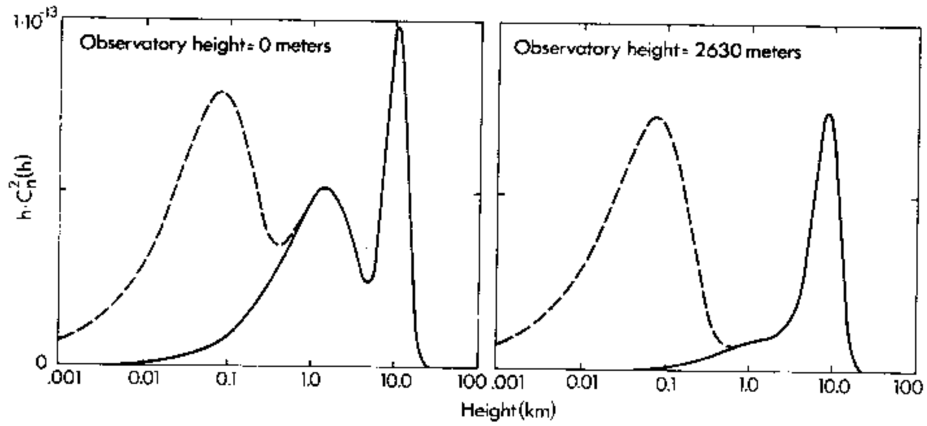


Figure 2 Average C_n^2 profile with local height h_L (in km). (*Left*) Profile for a sea level site. (*Right*) Profile for a 2630 meter high mountain site. The solid curve follows the expression given by Valley (1980) for height h above sea level: $C_n^2 = [2.05 \times 10^{-23} \cdot h^{10} \cdot \exp(-h) + 0.93 \times 10^{-16} \cdot \exp(-h/1.5)]\ m^{-2/3}$. It ignores near ground, local seeing. It is scaled to give 0.5 arcsec seeing at $=0.55\ \mu$m at sea level. The dashed line corresponds to $C_n^2 = (2.17 \times 10^{-15} + 5 \times 10^{-17} \cdot h_L{}^{-2/3}) \cdot \exp(-h_L/0.08)$ which also results in 0.5 arcsec seeing by itself. It approximates this local nighttime seeing. For the sea level site the resulting seeing is 0.76 arcsec; for the mountain site 0.63 arcsec. For daytime condition the local seeing will be worse. The $h \cdot C_n^2$ vs log h presentation was chosen to better visualize the contributions of the different heights to r_o.

Figure 9: From Beckers, ARAA 1993.

Seeing is more important than diffraction at shorter wavelengths (and for larger apertures) since r_0 scales roughly with wavelength. Diffraction more important at longer wavelengths (and for smaller apertures); the effects of diffraction and seeing cross over in the IR for most astronomical-sized telescopes (∼5 microns for 4m); the crossover falls at a shorter wavelength for smaller telescope or better seeing.

The meat of r_0 is in $\int(C_n^2 dh)$; as you might expect, this varies from site to site and also in time. At most sites, there seems to be three regimes of "surface layer" (wind-surface interactions and manmade seeing), "planetary boundary layer" (influenced by diurnal heating), and "free atmosphere" (10 km is tropopause: high wind shears), as seen in Figure 9.

A typical astronomical site has $r_0 \sim 10$ cm at 5000Å.

We also want to consider the coherence of the same turbulence pattern over the sky: this coherence angle is called the *isoplanatic angle*, and the region over which the turbulence pattern is the same is called the *isoplanatic patch*. This is relevant to adaptive optics, where we will try to correct for the differences across the telescope aperture; if we do a single correction, how large a field of view will be corrected?

$$\theta \sim 0.314 r_0/H$$

where H is the average distance of the seeing layer:

$$H = \sec z[\int(C_n^2 h^{5/3} dh)/\int(C_n^2 dh)]^{3/5}$$

For $r_0 = 10$ cm, $H \sim 5000$ m , $\theta \sim 1.3$ arcsec.

In the infrared $r_0 \sim 70$ cm, $H \sim 5000$ m, $\theta \sim 9$ arcsec, i.e. for free atmosphere. For boundary layer, however, isoplanatic patch is considerably larger (part of motivation for ground-layer AO).

Note however, that the "isoplanatic patch for image motion" (not wavefront) is $\sim 0.3D/H$. For $D = 4$ m, $H \sim 5000$ m, $\theta_{kin} = 50$ arcsec. This is relevant for low-order atmospheric correction, i.e., tip-tilt, where one is doing *partial* correction of the effect of the atmosphere.

As a final practical discussion of seeing, note that atmospheric turbulence is not directly correlated with the presence of clouds. In fact, the seeing is often better with thin cirrus than when it is clear!

Understand the basic concept of scintillation and seeing. Know the terminology of the coherence length (r_o) and the isoplanatic patch, and know what typical values are.

6.4.1 Other sources of seeing

Although the Earth's atmosphere provides a limit on the quality of images that can be obtained, at many observatories, there are other factors that can dominate the image quality budget. These have been recognized over the past several decades to be significant effects.

Dome seeing arises from turbulence pattern around the dome and the interface between inside the dome and outside the dome. Even small temperature differences can lead to significant image degradataion

Mirror seeing arises from turbulence right above the surface of the mirror, which can arise if the mirror temperature differs from that of the air above it.

Wind shake of the telescope can also contribute to image quality.

Poor telescope tracking can also contribute.

Finally, the design, quality, and alignment of the telescope optics can contribute to image quality. In general, however, telescope design is done such that the image degradation from telescope design is significantly smaller than that arising from seeing.

Understand the multiple man-made sources of seeing.

6.4.2 What does seeing cause the image to look like?

The "quality" of an image can be described in many different ways. The overall shape of the distribution of light from a point source is specified by the *point spread function*. Diffraction gives a basic limit to the quality of the PSF, but seeing, aberrations, or image motion add to structure/broadening of the PSF. For a good ground-based telescope, seeing is generally the dominant component of the PSF. The PSF is intrinsically a 2D function. In the case where image quality is azimuthally symmetric, then this can be represented by a 1D function or some parameterization of a 1D function.

Probably the most common way of describing the seeing is by specifying the full-width-half-maximum (FWHM) of the image, which may be estimated either by direct inspection or by fitting a function (usually a Gaussian); note the correspondence of FWHM to σ of a gaussian: $FWHM = 2.354\sigma$. Note that when you observe a PSF on a detector, you're really getting the PSF integrated over pixels: some people call the pixel-integrated PSF the effective PSF. Remember that the FWHM doesn't fully specify a PSF, and one should always consider how applicable the quantity is.

Another way of describing the quality of an image is to specify its *modulation transfer function* (MTF). The MTF and PSF are a Fourier transform pair, so the MTF gives the power in an image on various spatial scales. Turbulence theory makes a prediction for the MTF from seeing:

$$MTF(\nu) = \exp[-3.44(\lambda\nu/r_o)^{5/3}]$$

where ν is the spatial (not energy!) frequency. Note that a gaussian goes as $\exp[-\nu^2]$, so this is close to a gaussian. This is relevant because the seeing is often described by fitting a Gaussian to a stellar profile. The shape of seeing-limited images is roughly Gaussian in core but has more extended wings, as shown in Figure 10.

Note that stellar images don't have edges! How much of this profile you can "see" depends on brightness of star and background noise: stellar profile is independent of stellar brightness.

A potentially better empirical fitting function is a Moffat function as shown in Figure 11.

$$I = p_1(1 + (x - p_2)^2/p_4^2 + (y - p_3)^2/p_5^2)^{-p_6}$$

POINT-SPREAD FUNCTION 701

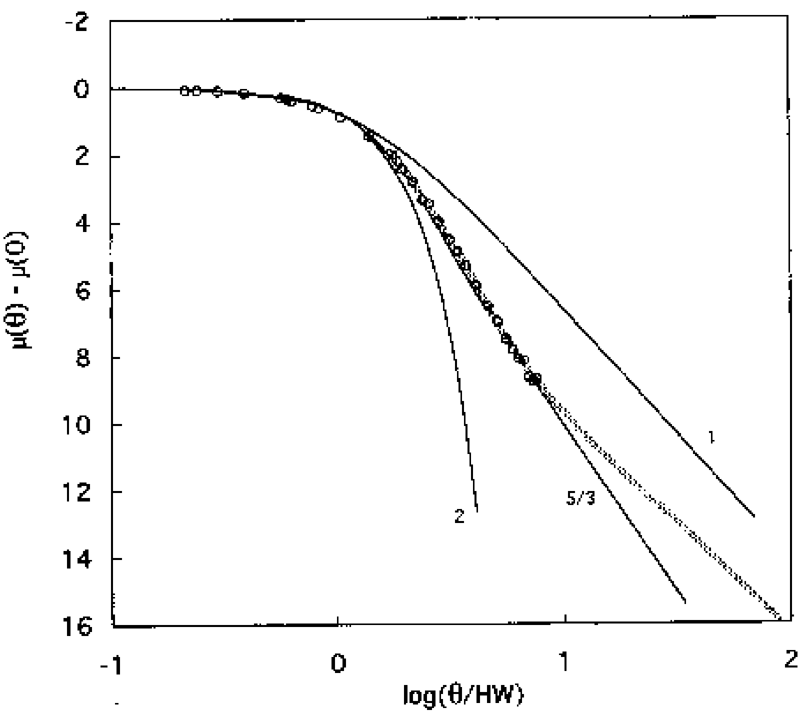


FIG. 2—The observed PSF (open circles, stippled line) is compared to theoretical models for MTFs of different indices *n*.

Figure 10: From Racine,

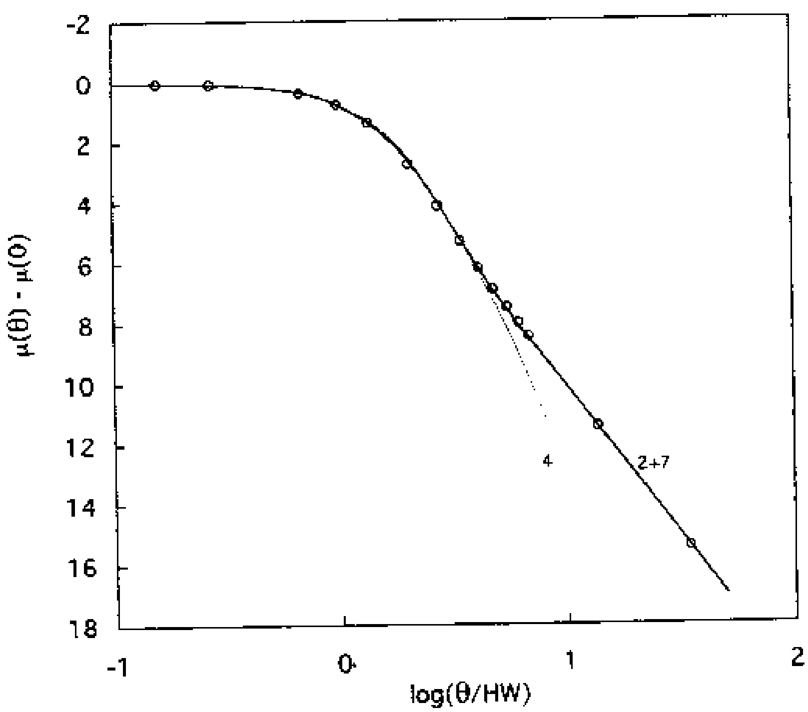


FIG. 3—The Kolmogorov PSF (open circles) fitted by a Moffat function with β=4 (thin line) and by the sum of two Moffat functions (β_1=7 and β_2=2). All functions have the same HW as the PSF.

Figure 11: From Racine,

Another way of characterizing the PSF is by giving the *encircled energy* as a function of radius, or at some specified radius. The encircled energy is just the cumulative integral of the PSF. Encircled energy requirements are often used for specifying optical tolerances.

A final way of characterizing the image quality, more commonly used in adaptive optics applications, is the Strehl ratio. The Strehl ratio is the ratio between the peak amplitude of the PSF and the peak amplitude expected in the presence of diffraction only. With normal atmospheric seeing, the Strehl ratio is *very* low. However, the Strehl ratio is often used when discussing the performance of adaptive optics systems (more later).

Know the terminology: FWHM, PSF, MTF, EE, and Strehl ratio. Understand how the PSF extends to large distances and does not depend on the brightness of the star, although your ability to recognize it might.

7 Astronomical optics and telescopes

Because astronomical sources are faint, we need to collect light. We use telescopes/cameras to make images of astronomical sources. Example: a 20th magnitude star gives $\sim 0.01 photons/s/cm^2$ at 5000 A through a 1000 A filter! However, using a 4m telescope gives 1200 photons/s.

Telescopes/optics are the bread and butter tool of the observational astronomer, so it is worthwhile to be familiar with how they work.

7.1 Single surface optics and definitions

We will define an optical system as a system which collects light; usually, the system will also make images. This requires the bending of light rays, which is accomplished using lenses (refraction) and/or mirrors (reflection), using *curved surfaces*.

The operation of refractive optical systems is given by Snell's law of refraction:

$$n \sin i = n' \sin i'$$

where n are the indices of refraction, i are the angles of incidence, relative to the normal to the surface. For reflection:

$$i' = -i$$

An optical element takes a source at s and makes an image at s'. The source can be *real* or *virtual*. A real image exists at some point in space; a virtual image is formed where light rays apparently emanate from or converge to, but at a location where no light actually appears. For example, in a Cassegrain telescope, the image formed by the primary is virtual, because the secondary intercepts the light and redirects it before light gets to the focus of the primary.

Considering an azimuthally symmetric optic, we can define the optical axis to go through the center of the optic. The image made by the optic will not necessarily be a perfect image: rays at different height at the surface, y, might not cross at the same point. This is the subject of aberrations, which we will get into in a while. For a "smooth" surface, the amount of aberration will depend on how much the different rays differ in y, which depends on the shape of the surface. We define *paraxial* and *marginal* rays, as rays near the center of the aperture and those on the edge of the aperture. We define the *chief ray* as the ray that passes through the center of the aperture. To define nominal (unaberrated) quantities, we consider the *paraxial* regime, i.e. a small region near the optical axis, surrounding the chief ray. In this regime, all angles are small, aberrations vanish, and a surface can be wholly specified by its radius of curvature R.

The *field angle* gives the angle formed between the chief ray from an object and the z-axis. Note that paraxial does not necessarily mean a field angle of zero; one can have an object at a field angle and still consider the paraxial approximation.

Note also that for the time being, we are ignoring *diffraction*. But we'll get back to that too. We are considering *geometric* optics, which is what you get from diffraction as wavelength tends to 0. For nonzero wavelength, geometric optics applies at scales $x >> \lambda$.

We can derive the basic relation between object and image location as a function of a surface where the index of refraction changes (Schroeder, chapter 2).

$$\frac{n'}{s'} - \frac{n}{s} = \frac{(n' - n)}{R}$$

The points at s and s' are called *conjugate*; the behavior is independent of which direction the light is going . If either s or s' is at infinity (true for astronomical sources for s), the other distance is defined as the *focal length*, f, of the optical element. For $s = \inf$, $f = s'$.

We can define the quantity on the right side of the equation, which depends only the the surface parameters (not the image or object locations), as the *power*, P, of the surface:

$$P \equiv \frac{(n' - n)}{R} = \frac{n'}{f'} = \frac{n}{f}$$

We can make a similar derivation for the case of reflection:

$$\frac{1}{s'} + \frac{1}{s} = \frac{2}{R}$$

This shows that the focal length for a mirror is given by $R/2$.

Note that one can treat reflection by considering refraction with $n' = -n$, and get the same result:

$$\frac{n'}{s'} + \frac{n'}{s} = \frac{(n' + n')}{R}$$

Given the focal length, we define the *focal ratio* to be the focal length divided by the aperture diameter. The focal ratio is also called the *F-number* and is denoted by the abbreviation $f/$. Note $f/10$ means a focal ratio of ten; f is not a variable in this! The focal ratio gives the beam "width"; systems with a small focal ratio have a short focal length compared with the diameter and hence the incoming beam to the image is wide. Systems with small focal ratios are called "fast" systems; systems with large focal ratios are called "slow" systems.

The *magnification* of a system gives the ratio of the image height to the object height:

$$\frac{h'}{h} = \frac{(s' - R)}{(s - R)} = \frac{ns'}{n's}$$

The magnification is negative for this case, because object is flipped. The magnification also negative for reflection: $n' = -n$. Magnification is an important quantity for multi-element systems.

We define the *scale* as the motion of image for given incident angle of parallel beam from infinity. From a consideration of the chief rays for objects on-axis and at field angle α, we get:

$$\tan\alpha \approx \alpha = \frac{x}{f}$$

or

$$\text{scale} \equiv \frac{\alpha}{x} = \frac{1}{f}$$

In other words, the scale, in units of angular motion per physical motion in the focal plane, is given by $1/f$. For a fixed aperture diameter, systems with a small focal ratio (smaller focal length) have a larger scale, i.e. more light in a patch of fixed physical size: hence, these are "faster" systems.

Exercise: the APO 3.5m telescope is a f/10 system. A typical CCD might have 15 micron pixels. What angle in the sky would one pixel subtend? Once you get this, comment on whether you think this is a good pixel scale and why or why not?

Know the terminology: real/virtual images, paraxial/marginal/chief rays, field angle, focal ratio, magnification. Understand what is meant by the paraxial approximation. Know the basic lens/mirror equation. Know how to calculate the scale of an optical system.

7.2 Multi-surface systems

To combine surfaces, one just takes the image from the first surface as the source for the second surface, etc., for each surface. We can generally describe the basic parameters of multi-surface systems by equivalent single-surface parameters, e.g. you can define an effective focal length of a multi-surface system as the focal length of some equivalent single-surface system. The effective focal length is the focal length of the first element multiplied by the magnification of each subsequent element. The two systems (single and multi) are equivalent in the paraxial approximation ONLY.

7.2.1 a lens (has two surfaces)

Consider a lens in air ($n \sim 1$). The first surface give

$$\frac{n}{s_1'} - \frac{1}{s_1} = \frac{(n-1)}{R_1} = P_1$$

The second surface gives:

$$\frac{1}{s'_2} - \frac{n}{s_2} = \frac{(1-n)}{R_2} = P_2$$

but we have $s_2 = s'_1 - d$ (remember we have to use the plane of the second surface to measure distances for the second surface).

After some algebra, we find the effective focal length (from center of lens):

$$P = \frac{1}{f'} = P_1 + P_2 - \frac{d}{n} P_1 P_2$$

$$P = \frac{(n-1)}{R_1} + \frac{(1-n)}{R_2} - \frac{d}{n}\frac{(n-1)(1-n)}{R_1 R_2}$$

From this, we derive the *thin lens* formula:

$$P = \frac{1}{f'} = \frac{(n-1)}{R_1} + \frac{(1-n)}{R_2} = (n-1)(\frac{1}{R_1} - \frac{1}{R_2})$$

$$\frac{1}{f'} = \frac{1}{f_1} + \frac{1}{f_2}$$

7.2.2 plane-parallel plate

Zero power, but moves image laterally: $\Delta = d[1-(1/n)]$. Application to filters: variation of focus.

7.2.3 Two-mirror telescopes:

In astronomy, most telescopes are two-mirror telescopes of Newtonian, Cassegrain, or Gregorian design. All 3 types have a concave primary. The Newtonian has a flat secondary, the Cassegrain a convex secondary, and the Gregorian a concave secondary. The Cassegrain is the most common for research astronomy; it is more compact than a Gregorian and allows for magnification by the secondary. Basic parameters are outlined in Figure 12.

Each of these telescope types defines a *family* of telescopes with different first-order performances. From the usage/instrumentation point of view, important quantities are:

- the diameter of the primary, which defines the light collecting power
- the scale of the telescope, which is related to the focal length of the primary and the magnification of the secondary:

  $$f_{eff} = f_1 m$$

  (alternatively, the focal ratio of the telescope, which gives the effective focal length with the diameter)

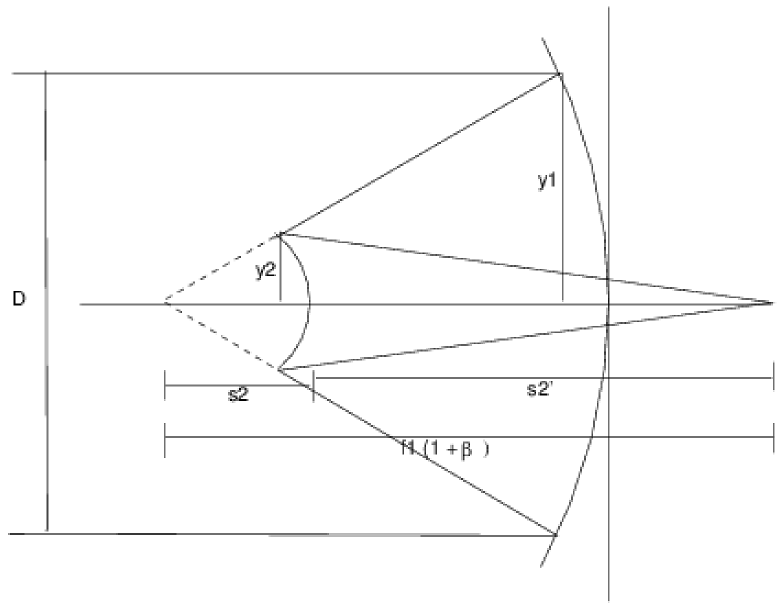


Figure 12: Schematic of a Cassegrain telescope.

- the back focal distance, which is the distance of the focal plane behind the telescope

From the design point of view, we need to specify:

- the radii of curvature of the mirrors
- the separation between the mirrors

The relation between the usage and design parameters can be derived from simple geometry. First, accept some basic definitions:

- ratio of focal lengths, ρ:
$$\rho = R_2/R_1 = f_2/f_1$$
- magnification of the secondary, m (beware that s_2' is negative for a Cassegrain!):
$$m = -s_2'/s_2$$
- *back focal distance*, the distance from the primary vertex to the focal plane (often expressed in units of the primary focal length, or primary diameter):
$$f_1\beta = D\eta$$

- primary focal ratio, F_1:

$$F_1 = f_1/D$$

- ratio of marginal ray heights, k (directly related to separation of mirrors):

$$k = y_2/y_1$$

Using some geometry, we can derive some basic relations between these quantities, in particular:

$$\rho = \frac{mk}{(m-1)}$$

and

$$(1+\beta) = k(m+1)$$

Usually, f_1 is limited by technology/cost. Then choose m to match desired scale. k is related to separation of mirrors, and is a compromise between making telescope shorter and blocking out more light vs. longer and blocking less light; in either case, have to keep focal plane behind primary!

One final thing to note is how we focus a Cassegrain telescope. Most instruments are placed at a fixed location behind the primary. Ideally, this will be at the back focal distance, and everything should be set as designed. However, sometimes the instrument may not be exactly at the correct back focal distance, or it might move slightly because of thermal expansion/contraction. In this case, focussing is usually then done by moving the secondary mirror.

The amount of image motion for a given secondary motion is given by:

$$\frac{d\beta}{dk} = \frac{d}{dk}k(m+1) - 1$$

Working through the relations above, this gives:

$$\frac{d\beta}{dk} = m^2 + 1$$

so the amount of focal plane motion ($f_1 d\beta$) for a given amount of secondary motion ($f_1 dk$) depends on the magnification of the system.

If you move the secondary you change k. Since ρ is fixed by the mirror shapes, it's also clear that you change the magnification as you move the secondary; this is expected since you are changing the system focal length, $f = mf_1$. So it's possible that a given instrument could have a slightly varying scale if its position is not perfectly fixed relative to the primary. Alternatively, if you need to independently focus and set the scale (e.g., SDSS!), then you need to be able to move two things!

Note that even if the instrument is at exactly the back focal distance, movement of the secondary is required to account for mechanical changing of spacing between the primary and secondary as a result of thermal expansion/contraction.

7.2.4 Definitions for multi-surface system: stops and pupils

- aperture stop: determines the amount of light reaching an image (usually the primary mirror)
- field stop: determines the angular size of the field. This is usually the detector, but for a large enough detector, it could be the secondary.
- pupil: location where rays from all field angles fill the same aperture.
- entrance pupil: image of aperture stop as seen from source object (usually the primary).
- exit pupil: image of aperture stop formed by all subsequent optical elements.

In a two-mirror telescope, the location of the exit pupil is where the image of the primary is formed by the secondary. This can be calculated using $s = d$ as the object distance (where d is the separation of the mirrors), then with the reflection equation, we can solve for s' which gives the location of the exit pupil relative to the secondary mirror. If one defines the quantity δ, such that $f_1\delta$ is the distance between the exit pupil and the focal plane, then (algebra not shown):

$$\delta = \frac{m^2 k}{m + k - 1} = \frac{m^2(1+\beta)}{m^2 + \beta}$$

This pupil is generally not accessible, so if one needs access to a pupil, additional optics are used.

The exit pupil is an important concept. When we discuss aberrations, it is the total wavefront error at the exit pupil which gives the system aberration. Pupils are important for aberration compensation. They can also be used to put light at a location that is independent of pointing errors.

Understand in principle how you can calculate multi-surface systems as a sequence of individual surface. Know the different types of two-mirror telescopes (Cassegrain, Gregorian, Newtonian). Know the terminology: aperture stop, field stop, pupil.

7.3 Telescope technology

While we are on the topic of telescope designs, it seems reasonable to include a bit of discussion on the technology of large telescopes

7.3.1 Large mirror types

One real-world issue for large telescopes is the technology of how to build a large mirror which will not be so heavy that it will sag under its own weight. Additionally, since it has been recognized that good image quality requires that the mirrors be at the same temperature as the outside air, the mirror technology must be such that the mirror has a short thermal time constant, or, in other words, it must be able to change temperature to match the outside air fairly quickly. If necessary, one can consider thermally controlling the mirror, e.g., with heating or air conditioning.

In the large mirror regime, there are currently three leading technologies. The first is the construction of a single large mirror (monolithic) made from borosilicate glass, but having large hollowed out regions to keep the weight down. This *borosilicate honeycomb* design has been pioneered by Roger Angel at the Mirror Lab of the University of Arizona. This type of mirror has been successfully cast in a 3.5m size (used in the ARC 3.5m (APO), WIYN 3.5m (KPNO), and the Starfire Optical Range Telescope near Albuquerque), and in a 6.5m format for the MMT conversion (Mt. Hopkins, AZ) and the Magellan (Las Campanas Observatory, Chile) telescopes; they have also been made in an 8m format (x2) for the Large Binocular Telescope (Mt. Graham, AZ). The second design is also monolithic but has a mirror which is significantly thinner than the borosilicate mirror. These *thin mirrors* are being built primarily by two companies, Corning (USA) and Schott (Germany). They use materials with good thermal properties, ULE (Corning) and Zerodur (Schott). Thin mirrors are being used in ESO's 3.5m New Technology Telescope (La Silla, Chile), Japan's 8m Subaru telescope (Mauna Kea, Hawaii), the two 8m Gemini telescopes (Mauna Kea and Cerro Pachon, Chile), and ESO's Very Large Telescopes (4 8m's on Cerro Paranal). Finally, the third design make use of *segmented mirrors*, in which a large mirror is made by combining many small mirrors. This design is currently operational in the 10m Keck telescope (Mauna Kea), the 11m Hobby-Eberly Telescope, the 11m SALT telescope, and the 10m Gran Telescopio de las Canarias. Future 30m class telescopes: TMT, GMT, and E-ELT.

See http://astro.nineplanets.org/bigeyes.html for a nice tabular summary, and http://cosmicpursuits.com/wp-content/uploads/2015/06/Comparison_optical_telescope_primary_mirrors.svg_.png for a graphic.

The borosilicate mirrors have the advantage that they are stiffer than the other designs, so the mirror support is less complicated. For thin mirrors, the support system must be activated to allow for changing shape as a function of telescope pointing. For segmented mirrors, each segment must be controlled to make sure the entire surface is smooth. The thick mirror is also less susceptible to wind shake, which can adversely affect image quality. The thin and segmented mirrors have the advantage of better thermal properties since they contain less total material.

The choice of a primary mirror technology can be complicated. In designing a large telescope, one generally first decides on an optical prescription which is chosen considering the main scientific goals for the project (e.g., large field, IR, good image quality, etc.). The primary mirror choice is made considering the choice of site (e.g, are there large temperature changes, lots of wind, etc.), availability, issues of engineering complexity, and, especially, cost (and politics). The choice of a mount and control system to use is basically a cost and operations issue.

7.3.2 Mirror coatings

Aluminum, silver, gold (JWST) most commonly used. See, e.g. http://www.optiforms.com/metallic.htm for relative reflectances as a function of wavelength, also here. Curves like these can be incorporated into an exposure time calculator to account for the efficiency of the telescope as a function of wavelength. Note the effect of multiple mirrors: if you have three mirrors with 90% reflectance, you will have a total loss of almost 30% of the light!

Issues with mirror cleaning and recoating; coatings get dirty and also degrade over time. The degradation depends on the exposure, hence observatories often have constraints on humidity and dust levels, for example.

7.3.3 Telescope mounts

We've talked about the optics that go into telescopes. However, it's clear that these optics need to be supported in some structure and kept in alignment with each other. The support structures needed are really an engineering issue (and a challenging one for large telescopes), and we won't disucss it here. In addition to supporting the optics, the structure also needs to be capable of tracking astronomical objects as they move across the sky because of the rotation of the earth.

There are two main different sorts of telescope mounts found in observatories: the *equatorial* mount and the *altitude-azimuth (alt-az)* mount. The equatorial mount is by far the most common for older telescopes, but the alt-az design is being used more frequently for newer, especially larger, telescopes. In the equatorial design, the telescope move along axes which are parallel and perpendicular to the polar axis, which is the direction parallel to the earth's rotation axis. In such a mount, tracking the earth's rotation only requires motion along one axis, the one perpendicular to the polar axis, and the tracking motion is at a uniform rate. In the alt-az mount, the telescope moves along axes which are perpendicular and parallel to the local vertical axis. With this mount, however, tracking of celestial objects requires motions of variable speed along both axes. An additional complication of an alt-az mount is the fact that, for a detector which is fixed to the back of the telescope, the image field rotates as the telescope tracks an object. Note, however, that the telescope pupil does not rotate with the object.

An equatorial mount is much easier to control for pointing and tracking. However, from an engineering point of view, it is much more demanding to construct, especially for large telescopes which have significant weight. The engineering complications generally result in a significantly larger cost (for large telescopes) than for an alt-az design. An alt-az telescope, however, has a significantly more complex control system, and must have an image rotator for the instruments. Given the advances in digital motor control and computing, the control system usually no longer poses a very significant challenge.

Regardless of mount type, the mount is never built absolutely perfectly, i.e. with axes exactly perpendicular, exactly aligned as they should be, totally round surfaces, optics aligned with mechanics, etc. As a result, a telescope does not generally point perfectly. However, many effects of an imperfect telescope are quite repeatable, so they can be corrected for. This corrrection is done by something called a pointing model, which records the difference in true position from prediction position over the sky, and, once derived, the pointing model can be implemented to significantly improve pointing. A good telescope points to within a few arcseconds after implementation of a good pointing model.

Related to pointing is tracking performance. The issue here is how long the telescope can stay pointed at a given target. You can consider this question as how well the telescope can point over the area of the sky through which your object will drift. Since your required pointing stability should be significantly less than one arcsec, so that tracking does not degrade the image quality significantly, almost no telescopes have sufficiently good pointing to track to within an arcsecond for an arbitrarily long time. Most telescopes can track sucessfully for several minutes, but will give significant image degradation for exposures longer than this. Consequently, most telescopes/instruments are equipped with *guide cameras*, which are used to continually correct the pointing by observing an object somewhere in the field of view of the telescope (possibly the object you are interested in, but usually not, since that's where your detector is looking). These days, most guiders are *autoguiders*, meaning that they automatically find the position of the guide object, compute the pointing offsets needed to keep this object in one position, and send these offsets as commands to the telescope. The observer generally just has to choose a guide object for the autoguider to use, though they also may have to adjust the guide camera sensitivity or gain to insure that the guide star has a strong signal. These days, many autoguiders can automatically find guide stars in the field or from some on-line catalog (e.g., the HST Guide Star Catalog, which catalogs stars down to V 14). However, if one is taking long exposures and knows that they'll need to use guide stars, make sure to find out whether such a facility is available ; if not, it may still be possible to find guide stars in advance of your observing run, e.g., from the sky survey. If so, you should seriously consider doing so, as it can take a frustratingly long amount of time to search for a guide star at the telescope in real time. Since telescope time is heavily oversubscribed at most facilities, you really want to maximize your efficiency, and doing so is a large part of what will make you a "expert" observer.

Note guiding in spectrographs is often done off of the slit with a slit-viewing camera.

7.4 Aberrations

7.4.1 Surface requirements for unaberrated images

Next we consider non-paraxial rays. We first consider what surface is required to make an unaberrated image.

We can derive the surface using Fermat's principle. Fermat's principle states that light travels in the path such that infinitessimally small variations in the path doesn't change the travel time to first order: d(time)/d(length) is a minimum. For a single surface, this reduces to the statement that light travels the path which takes the least time. An alternate way of stating Fermat's principle is that the *optical path length* is unchanged to first order for a small change in path. The OPL is given by:

$$OPL = \int cdt = \int \frac{c}{v} vdt = \int nds$$

Fermat's principle has a physical interpretation when one considers the wave nature of light. It is clear that around a stationary point of the optical path light, the maximum amount of light can be accumulated over different paths with a minimum of destructive interference. By the wave theory, light travels over all possible paths, but the light coming over the "wrong" paths destructively interferes, and only the light coming over the "right" path constructively interferes.

Fermat's principle can be used to derive the basic laws of reflection and refraction (Snell's law).

Now consider a perfect imaging system that takes all rays from an object and makes them all converge to an object. Since Fermat's principle says the only paths taken will be those for which the OPL is minimally changed for small changes in path, the only way a perfect image will be formed is when all optical path lengths along a surface between an image and object point are the same - otherwise the light doesn't get to this point!

Instead of using Fermat's principle, we could solve for the parameters of a perfect surface using analytic geometry, but this would require an inspired guess for the correct functional form of the surface.

We find that the perfect surface depends on the situation: whether the light comes from a source at finite or infinite distance, and whether the mirror is concave or convex. We consider the various cases now, quoting the results without actually doing the geometry. In all cases, consider the z-axis to be the optical axis, with the y-axis running perpendicular. We want to know the shape of the surface, $y(z)$, that gives a perfect image.

Concave mirror with one conjugate at infinity

Sample application: primary mirror of telescope looking at stars.

Fermat's principle gives:

$$y^2 = 2Rz$$

where $R = 2f$, the radius of curvature at the mirror vertex. This equation is that of a parabola. Note, however, that a parabola makes a perfect image only for on axis images (field angle=0).

Concave mirror with both conjugates at finite distance

Sample application: Gregorian secondary looking at image formed by primary.

For a concave mirror with both conjugates finite, we get an ellipse. Again, this is perfect only for field angle = 0.

$$(z - a)^2/a^2 + y^2/b^2 = 1$$

$$y^2 - 2zb^2/a + z^2b^2/a^2 = 0$$

where

$$a = (s + s')/2.$$

$$b = \sqrt{(ss')}$$

$$R = ss'/(s + s') = 2b^2/a$$

Convex mirror with both conjugates at finite distance

Sample application: Cassegrain secondary looking at image formed by primary.

For a convex mirror with both conjugates finite, we get a hyperbola:

$$(z - a)^2/a^2 - y^2/b^2 = 1$$

$$y^2 + 2zb^2/a - z^2b^2/a^2 = 0$$

where

$$a = (s + s')/2$$

$$b^2 = -ss'$$

(s is negative)

$$R = -2b^2/a$$

Convex mirror with one conjugate at infinity

For a convex mirror with one conjugate at infinity, we get a parabola.

2D to 3D

Note that in all cases we've considered a one-dimension surface. We can generalize to 2D surfaces by rotating around the z-axis; for the equations, simply replace y^2 with $(x^2 + y^2)$.

Conic sections

As you may recall from analytic geometry, all of these figures are *conic sections*, and it is possible to describe all of these figures with a single equation:

$$\rho^2 - 2Rz + (1 + K)z^2 = 0$$

where

$$\rho^2 = x^2 + y^2$$

and R is the radius of curvature at the mirror vertex, K is called the conic constant ($K = -e^2$, where e is the eccentricity for an ellipse, $e(b, a)$).

$K > 0$ gives a prolate ellipsoid
$K = 0$ gives a sphere
$-1 < K < 0$ gives a oblate ellipsoid
$K = -1$ gives a paraboloid
$K < -1$ gives a hyperboloid

Know what optical shapes produce perfect images for different situations. Know the terminology: conic constant.

7.4.2 Aberrations: general description and low-order aberrations

Now consider what happens for surfaces that are not perfect, e.g. for the cases considered above for field angle≠0 (since only a sphere is symmetric for all field angles), or for field angle 0 for a conic surface which doesn't give a perfect image?

You get *aberrations*; the light from all locations in aperture does **not** land at **any** common point.

One can consider aberrations in either of two ways:

1. aberrations arise from all rays not landing at a common point,
2. aberrations arise because wavefront deviates from a spherical wavefront.

These two descriptions are equivalent. For the former, one can talk about the *transverse* aberrations, which give the distance by which the rays miss the paraxial focus, or the *angular* aberration, which is the angle by which the rays deviate from the perfect ray which will hit paraxial focus. For the latter, one discusses the wavefront error, i.e., the

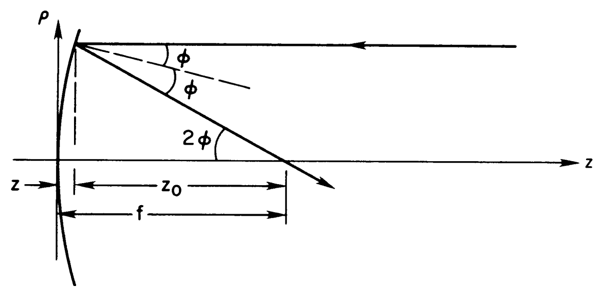


Fig. 4.1. Geometry of ray from distant object reflected from concave mirr

Figure 13: Spherical aberration diagram

deviation of the wavefront from a spherical wavefront as a function of location in the exit pupil.

In general, the angular and transverse aberrations can be determined from the optical path difference between a given ray and that of a spherical wavefront. The relations are given by:

$$\text{angular aberration} = \frac{d(2\Delta z)}{d\rho}$$

$$\text{transverse aberration} = s' \frac{d(2\Delta z)}{d\rho}$$

If the aberrations are not symmetric in the pupil, then we could define angular and transverse x and y aberrations separately by taking derivatives with respect to x or y instead of ρ.

Spherical aberration

First, consider the axisymmetric case of looking at an object on axis (field angle equal zero) with an optical element that is a conic section. We can consider where rays land as $f(\rho)$, and derive the effective focal length, $f_e(\rho)$, for an arbitrary conic section:

in Figure 13.

$$z_0 = \rho / \tan(2\phi) = \rho(1 - (\tan\phi)^2)/(2\tan\phi)$$

$$\tan\phi = dz/d\rho$$

from conic equation:

$$\rho^2 - 2Rz + (1 + K)z^2 = 0$$

$$z = \frac{R}{(1+K)}\left[1 - \left(1 - \frac{\rho^2}{R^2}(1+K)\right)^{1/2}\right]$$

$$z \approx \frac{\rho^2}{2R} + (1+K)\frac{\rho^4}{8R^3} + (1+K)^2\frac{\rho^6}{16R^5} + \ldots$$

$$dz/d\rho = \rho/(R-(1-K)z)$$

$$z_0 = \frac{\rho}{2}\left[\frac{R-(1+K)z}{\rho} - \frac{\rho}{R-(1+K)z}\right]$$

$$\begin{aligned} f_e &= z + z_0 \\ &= \frac{R}{2} + \frac{(1-K)z}{2} - \frac{\rho^2}{2(R-(1+K)z} \\ &= \frac{R}{2} - (1+K)\frac{\rho^2}{4R} - (1+K)(3+K)\frac{\rho^4}{16R^3} - ... \end{aligned}$$

$$\Delta f = f_e - \frac{R}{2}$$

Note that f_e is independent of z only for $K = -1$, a parabola. Also note that Δf is symmetric with respect to ρ.

We define spherical aberration as the aberration resulting from $K \neq -1$. Rays from different radial positions in the entrance aperture focus at different locations. It is an aberration which is present on axis as seen in Figure 14.

Spherical aberration is symmetric in the pupil. There is no location in space where all rays focus at a point. Note that the behavior (image size) as a function of focal position is not symmetric. One can define several criteria for where the "best focus" might be, leading to the terminology paraxial focus, marginal focus, diffraction focus, and the circle of least confusion.

The asymmetric nature of spherical aberration as a function of focal position distinguishes it from other aberrations and is a useful diagnostic for whether a system has this aberration. This is shown in in Figure 15.

We define *transverse spherical aberration* (TSA) as the image size at paraxial focus. This is not the location of the minimum image size.

in Figure 16.

$$\frac{TSA}{\Delta f} = \frac{\rho}{(f - z(\rho))}$$

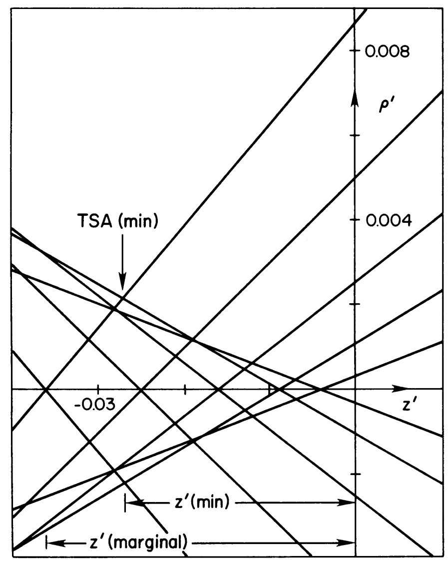


Fig. 4.5. Ray distribution near paraxial focus for image with spherical aberration. Paraxial focus is at (0, 0). See Eq. (4.2.8) for definition of parameters.

Figure 14: Spherical aberration

$$TSA = -(1+K)\frac{\rho^3}{2R^2} - 3(1+K)(3+K)\frac{\rho^5}{8R^4} + ...$$

The difference in angle between the "perfect" ray from the parabola and the actual ray is called the *angular aberration*, in this case *angular spherical aberration*, or ASA. in Figure 17.

$$ASA = 2(\phi_p - \phi) \approx \frac{d}{d\rho}(2\Delta z) \approx -(1+K)\frac{\rho^3}{R^3}$$

where $2\Delta z$ gives the optical path difference between the two rays.

This is simply related to the transverse aberration:

$$TSA = \frac{R}{2} ASA$$

We can also consider aberration as the difference between our wavefront and a spherical wavefront, which in this case is the wavefront given by a parabolic surface. in Figure 18.

$$\Delta z = z_{parabola} - z(K) = -\frac{\rho^4}{8R^3}(1+K) + \dots$$

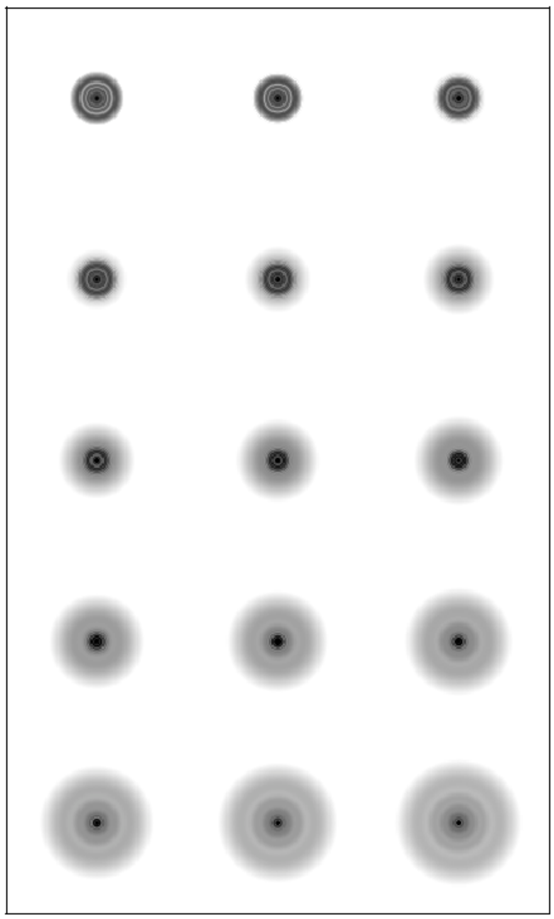


Figure 15: Spherical aberration image sequence as function of focal position

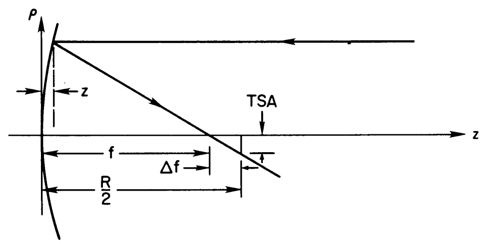


Fig. 4.2. Transverse spherical aberration (TSA) at paraxial focus. See Eqs. (4.1.6) and (4.2.1).

Figure 16: Spherical aberration diagram

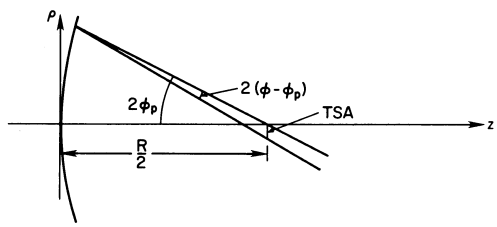


Fig. 4.4. Relation between TSA and angular difference between ray paths after reflection. See Eq. (4.2.3).

Figure 17: Spherical aberration diagram

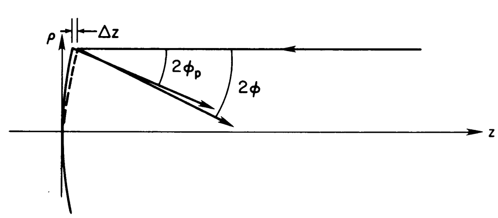


Fig. 4.3. Path difference between ray reflected from paraboloid (solid curve) and conic (dashed curve). Size of Δz, given in Eq. (4.2.2), is greatly exaggerated in the diagram.

Figure 18: Spherical aberration diagram

This result can be generalized to any sort of aberration: the angular and transverse aberrations can be determined from the optical path difference between a given ray and that of a spherical wavefront. The relations are given by:

$$\text{angular aberration} = \frac{d(2\Delta z)}{d\rho}$$

$$\text{transverse aberration} = s' \frac{d(2\Delta z)}{d\rho}$$

If the aberrations are not symmetric in the pupil, then we could define angular and transverse x and y aberrations separation by taking derivatives with respect to x or y instead of ρ.

General aberration description

We can describe deviations from a spherical wavefront generally. Since all we care about are optical path *differences*, we write an expression for the optical path difference between an arbitrary ray and the chief ray, and in doing this, we can also include the possibility of an off-axis image, and get

$$OPD = OPL - OPL(chiefray)$$

$$OPD = A_0 y + A_1 y^2 + A_1' x^2 + A_2 y^3 + A_2' x^2 y + A_3 \rho^4$$

where we've kept terms only to fourth order and chosen our coordinate system such that the object lies in the y-z plane. The coefficients, A, depend on lots of things, such as (θ, K, n, R, s, s').

Note that rays along the y-axis are called *tangential* rays, while rays along the x-axis are called *sagittal* rays.

Analytically, people generally restrict themselves to talking about *third-order* aberrations, which are fourth-order (in powers of $x, y, \rho, \mathrm{or}\theta$) in the optical path difference, because of the derivative we take to get transverse or angular aberrations. In the third-order limit, one finds that $A_2 = A_2'$, and $A_1 = -A_1'$. Working out the geometry, we find for a mirror that:

$$A_0 = 0$$

$$A_1 = \frac{n\theta^2}{R}$$

$$A_2 = -\frac{n\theta}{R^2}\left(\frac{m+1}{m-1}\right)$$

$$A_3 = \frac{n}{4R^3}\left[K + \left(\frac{m+1}{m-1}\right)^2\right]$$

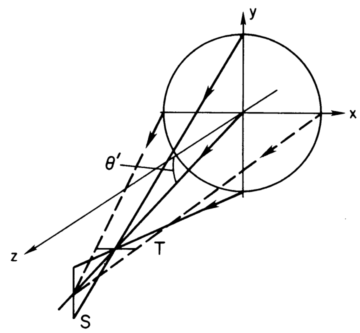


Fig. 5.2. Location and orientation of astigmatic line images. T and S denote tangential and sagittal images, respectively.

Figure 19: Rays in the presence of astigmatism.

From the general expression, we can derive the angular or the transverse aberrations in either the x or y direction. Considering the aberrations in the two separate directions, we find:

$$AA_y = 2A_1 y + A_2(x^2 + 3y^2) + 4A_3 y\rho^2$$
$$AA_x = 2A'_1 x + 2A_2 xy + 4A_3 x\rho^2$$

The first term is proportional to $\theta^2 y$ and is called *astigmatism*. The second term is proportional to $\theta(x^2 + 3y^2)$ and is called *coma*. The final term, proportional to $y\rho^2$ is *spherical aberration*, which we've already discussed (note for spherical, $AA_x = AA_y$ and in fact the AA in any direction is equal, hence the aberration is circularly symmetric).

Astigmatism

For astigmatism, rays from opposite sides of the pupil focus in different locations relative to the paraxial rays. At the paraxial focus, we end up with a circular image. As you move away from this image location, you move towards the tangential focus in one direction and the sagittal focus in the other direction. At either of these locations, the astigmatic image looks like a elongated ellipse. Astigmatism goes as θ^2, and consequently looks the same for opposite field angles. Astigmatism is characterized in the image plane by the *transverse* or *angular* astigmatism (TAS or AAS), which refer to the height of the marginal rays at the paraxial focus. Astigmatism is symmetric around zero field angle.

Figure 19 shows the nature of astigmatism.

Figure 20 shows the behavior of astigmatism as one passes through paraxial focus.

Coma

For coma, rays from opposite sides of the pupil focus at the same focal distance. However, the tangential rays focus at a different location than the sagittal rays, and

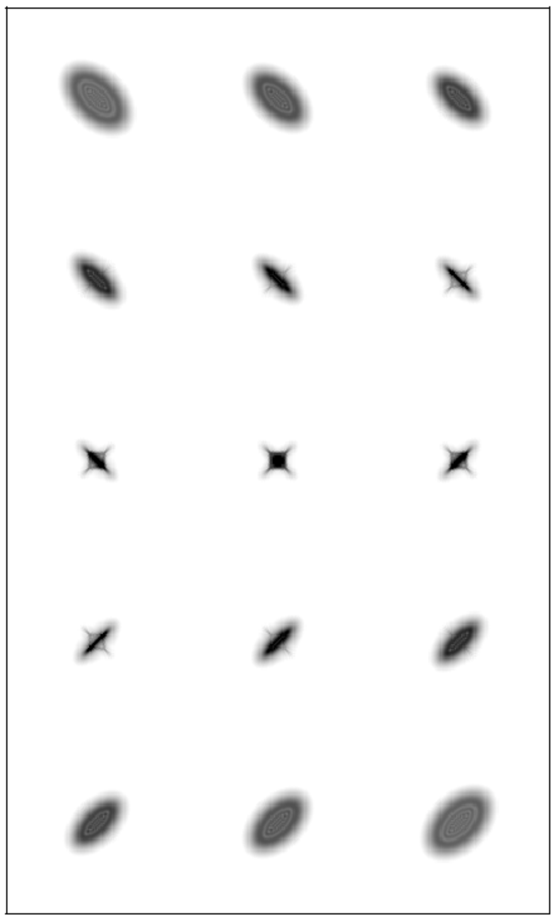


Figure 20: Astigmatism image sequence as function of focal position

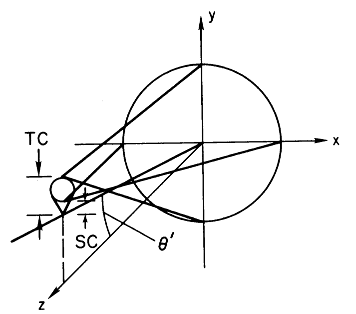


Fig. 5.5. Sketch of comatic image profile. TC and SC are tangential and sagittal coma, respectively.

Figure 21: Rays in the presence of coma.

neither of these focus at the paraxial focus. The net effect is to make an image that vaguely looks like a comet, hence the name coma. Coma goes as θ, so the direction of the comet flips sign for opposite field angles. Coma is characterized by either the *tangential* or *sagittal transverse/angular coma* (TTC, TSC, ATC, ASC) which describe the height/angle of either the tangential or sagittal marginal rays at the paraxial focus: $TTC = 3TSC$.

Figure 21 shows the nature of coma.

Figure 22 shows the behavior of coma as one passes through paraxial focus.

In fact, there are two more third-order aberrations: *distortion* and *field curvature*. Neither affects image quality, only location (unless you are forced to use a flat image plane!). Field curvature gives a curved focal plane: if imaging onto a flat detector, this will lead to focus deviations as one goes off-axis. Distortion affects the location of images in the focal plance, and goes as θ^3. The amount of field curvature and distortion can be derived from the aberration coefficients and the mirror parameters.

We can also determine the relevant coefficients for a surface with a displaced stop (Schroeder p 77), or for a surface with a decentered pupil (Schroeder p89-90); it's just more geometry and algebra. With all these realtions, we can determine the optical path differences for an entire system: for a multi-surface system, we just add the OPD's as we go from surface to surface. The final aberrations can be determined from the system OPD.

Understand the basic concepts of aberration. Know what the five third-order aberrations are (spherical, coma, astigmatism, field curvature, distortion) and have a basic idea about how they affect image quality and/or location.

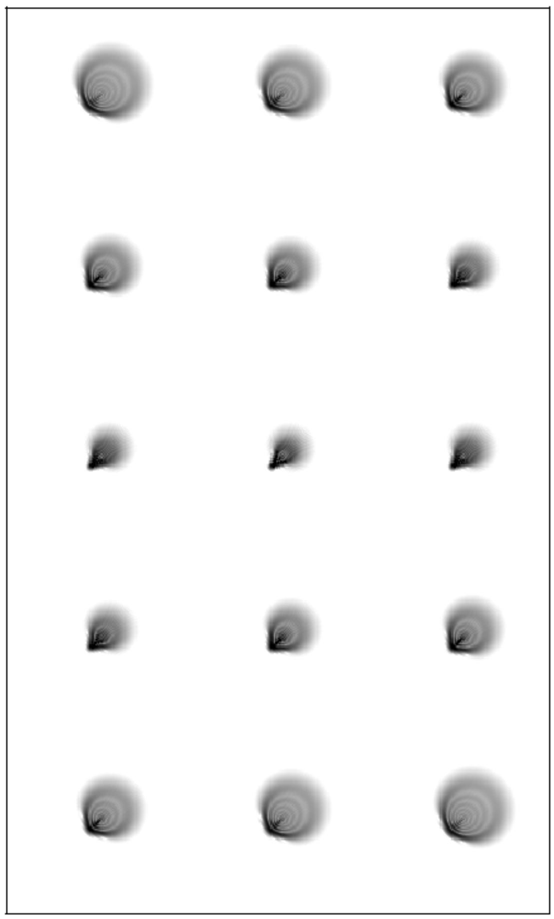


Figure 22: Coma image sequence as function of focal position

7.4.3 Aberration compensation and different telescope types

Using the techniques above, we can write expressions for the system aberrations as a function of the surface figures (and field angles). If we give ourselves the freedom to choose surface figures, we can eliminate one (or more) aberrations.

For example, given a conic constant of the primary mirror, we can use the aberration relations to determine K_2 such that spherical aberration is zero; this will give us perfect images on-axis. We find that:

$$K_2 = -\left(\frac{(m+1)}{(m-1)}\right)^2 + \frac{m^3}{k(m-1)^3}(K_1+1)$$

satisfies this criterion. If we set the primary to be a parabola ($K_1 = -1$), this gives the conic constant of the secondary we must use to avoid spherical aberration. This type of telescope is called a *classical* telescope. Using the aberration relations, we can determine the amount of astigmatism and coma for such telescopes, and we find that coma gives significantly larger aberrations than astigmatism, until one gets to very large field angles.

If we allow ourselves the freedom to choose both K_1 and K_2, we can eliminate both spherical aberration and coma. Designs of this sort are called *aplanatic*. The relevant expression, in terms of the magnification and back focal distance (we could use the relations discussed earlier to present these in terms of other paraxial parameters), is:

$$K_1 = -1 - \frac{2(1+\beta)}{m^2(m-\beta)}$$

We can only eliminate two aberrations with two mirrors, so even this telescope will be left with astigmatism.

There are two different classes of two-mirror telescopes that allow for freedom in the shape of both mirrors: Cassegrain telescopes and Gregorian telescopes (Newtonians have a flat secondary). For the classical telescope with a parabolic primary, the Cassegrain secondary is hyperbolic, whereas for a Gregorian it is ellipsoidal (because of the appropriate conic sections derived above for convex and concave mirrors with finite conjugates). For the aplanatic design, the Cassegrain telescope has two hyperbolic mirrors, while the Gregorian telescope has two ellipsoidal mirrors. An aplanatic Cassegrain telescope is called a *Ritchey-Chretien* telescope.

The following table gives some characteristics of "typical" telescopes. Aberrations are given at a field angle of 18 arc-min in units of arc-seconds. Coma is given in terms of tangential coma.

Characteristics of Two-Mirror Telescopes

Parameter	CC	CG	RC	AG
m	4.00	-4.00	4.00	-4.00
k	0.25	-0.417	0.25	-0.417
1 - k	0.75	1.417	0.75	1.417
mk	1.000	1.667	1.000	1.667
ATC	2.03	2.03	0.00	0.00
AAS	0.92	0.92	1.03	0.80
ADI	0.079	0.061	0.075	0.056
$\kappa_m R_1$	7.25	-4.75	7.625	-5.175
$\kappa_p R_1$	4.00	-8.00	4.00	-8.00

The image quality is clearly better for the aplanatic designs than for the classical designs, as expected because coma dominates off-axis in the classical design. In the aplanatic design, the Gregorian is slightly better. However, when considerations other than just optical quality are considered, the Cassegrain usually is favored: for the same primary mirror, the Cassegrain is considerably shorter and thus it is less costly to build an enclosure and telescope structure. To keep the physical length the same, the Gregorian would have to have a faster primary mirror, which are more difficult (i.e. costly) to fabricate, and which will result in a greater sensitivity to alignment errors. Both types of telescopes have a *curved* focal plane.

Understand how multi-surface systems can be used to reduce or remove aberrations. Know the terminology: aplanatic telescope, Ritchey-Chretien telescope.

7.5 Sources of aberrations

So far, we have been discussing aberrations which arise from the optical design of a system when we have a limited number of elements. However, it is important to realize that aberrations can arise from other sources as well. These other sources can give additional third-order aberrations, as well as higher order aberrations. Some possible sources include:

- design: as we have seen, it may not be possible to remove all aberrations with a limited number of surfaces
- misfigured or imperfectly figured optics : rarely is an element made exactly to specification!
- misalignments. If the mirrors in a multiple-element system are not perfectly aligned, aberrations will result. These can be derived (third-order) from the aberration expressions for decentered elements. For two mirror systems, one finds that decenter-

ing or tilting the secondary introduces a *constant* amount of coma over the field. Coma dominates astigmatism for a misaligned telescope.

- mechanical/support problems. When the mirrors are mounted in mirror cells the weight of the mirror is distributed over some support structures. Because the mirrors are not infinitely stiff, some distortion of the mirror shape will occur. Generally, such distortion will probably change as a function of which way the telescope is pointing. Separate from this, becuase the telescope structure itself is not perfectly stiff, one expects some flexure which gives a different secondary (mis)alignment as a function of where one is pointing. Finally, one might expect the spacing between the primary and secondary to vary with temperature, if the telescope structure is made of materials which have non-zero coefficients of expansion.

- chromatic aberration. Generally, we've only been discussing mirrors since this is what is used in telescopes. However, astronomers often put additional optics (e.g., cameras or spectrographs) behind telescopes which may use refractive elements rather than mirrors. There are aberration relations for refractive elements just as we've discussed, but these have additional dependences on the indices of refraction of the optical elements. For most refractive elements, the index of refraction varies with wavelength, so one will get wavelength-dependent aberrations, called chromatic aberrations. These can be minimized by good choices of materials or by using combinations of different materials for different elements; however, it is an additional source of aberration.

- seeing. The earth's atmosphere introduces optical path differences between the rays across the aperture of the telescope. This is generally the **dominant** source of image degradation from a ground-based telescope. Consequently, one builds telescopes in good sites, and as far as design and other sources of image degradation are concerned, one is generally only interested in getting these errors small when compared with the smallest expected seeing errors.

7.6 Ray tracing

For a fully general calculation of image quality, one does not wish to be limited to third-order aberrations, nor does one often wish to work out all of the relations for the complex set of aberrations which result from all of the sources of aberration mentioned above. Real world situations also have to deal with *vignetting* in optical systems, in which certain rays may be blocked by something and never reach the image plane (e.g., in a two-mirror telescope, the central rays are blocked by the secondary).

Because of these and other considerations, analysis of optical systems is usually done using *ray tracing*, in which the parameters of an optical system are entered into a computer,

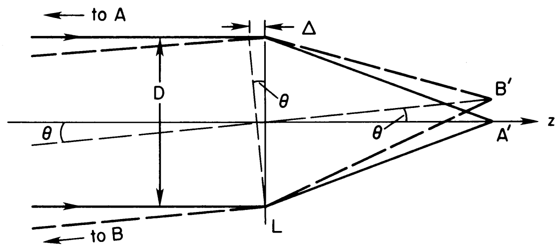


Fig. 3.11. Schematic of perfect optical system from which approximate diffraction limit is derived. See Eq. (3.6.1) and Section 3.VI.

Figure 23: The diffraction limit of a telescope

and the computer calculates the expected images on the basis of geometric optics. Many programs exist with many features: one can produce *spot diagrams* which show the location of rays from across the aperture at an image plane (or any other location), plots of transverse aberrations, plots of optical path differences, etc., etc.

7.7 Physical (diffraction) optics

Up until now, we have avoided considering the wave nature of light which introduces *diffraction* from interference of light coming from different parts of the aperture. Because of diffraction, images of a point source will be slightly blurred. From simple geometric arguments, we can estimate the size of the blur introduced from diffraction: as shown in Figure 23

Diffraction is expected to be important when $\Delta \sim \lambda$, i.e.,

$$\theta \sim \frac{\lambda}{D}$$

Using this, we find that the diffraction blur is smaller than the blur introduced by seeing for $D > 0.2$ meters at 5500 Å, even for the excellent seeing conditions of 0.5 arcsecond images. However, the study of diffraction is relevant because of several reasons: 1) the existence of the Hubble Space Telescope (and other space telescopes), which is diffraction limited (no seeing), 2) the increasing use of infrared observations, where diffraction is more important than in the optical, and 3) the development of adaptive optics, which attempts to remove some of the distortions caused by seeing. Consequently, it's now worthwhile to understand some details about diffraction.

To work out in detail the shape of the images formed from diffraction involves understanding wave propagation. Basically, one integrates over all of the source points in the aperture (or exit pupil for an optical system), determining the contribution of each point at each place in the image plane. The contributions are all summed taking into account phase differences at each image point, which causes reinforcment at some points and cancellation at others. The expression which sums all of the individual source points is called the *diffraction integral*. When the details are worked out, one finds that the intensity in the image plane is related to the intensity and phase at the exit pupil. In fact the wavefront is described at any plane by the *optical transfer function*, which gives the intensity and phase of the wave at all locations in that plane. The OTF at the pupil plane and at the image plane are a Fourier transform pair. Consequently, we can determine the light distribution in the image plane by taking the Fourier transform of the pupil plane; the light distribution, or point spread function, is just the modulus-squared of the OTF at the image plane. Symbolically, we have

$$PSF = \left| \int (OTF(pupil)) \exp ikx \right|^2$$

where

$$OTF(pupil) = P(x, y) \exp ik\phi(x, y)$$

$P(x, y)$ is the *pupil function*, which gives the transmission properties of the pupil, and usually consists of ones and zeros for locations where light is either transmitted or blocked (e.g., for a circular lens, the pupil function is unity within the radius of lens, and zero outside; for a typical telescope the pupil function includes obscuration by the secondary and secondary support structure). ϕ is the phase in the pupil. More relevantly, ϕ can be taken to be the optical path difference in the pupil with some fiducial phase, since only OPDs matter, not the absolute phase. Finally the wavenumber k is just $\frac{2\pi}{\lambda}$.

For the simple case of a plane wave with no phase errors, the diffraction integral can be solved analytically. The result for a circular aperture with a central obscuration, when the fractional radius of the obscuration is given by ϵ, the expression for the PSF is:

$$PSF \propto \left[\frac{2J_1(v)}{v} - \epsilon^2 \frac{2J_1(\epsilon v)}{\epsilon v} \right]^2$$

$$v = \frac{\pi r}{\lambda F}$$

where J_1 is a first order Bessel function, r is the distance in the image plane, λ is the wavelength, and F is the focal ratio ($F = f/D$).

This expression gives the so-called *Airy pattern* which has a central disk surrounded by concentric dark and bright rings. One finds that the radius of the first dark ring is at the physical distance $r = 1.22\lambda F$, or alternatively, the angular distance $\alpha = 1.22\lambda/D$. This gives the size of the *Airy disk*.

For more complex cases, the diffraction integral is solved numerically by doing a Fourier transform. The pupil function is often more complex than a simple circle, because there are often additional items which block light in the pupil, such as the support structures for the secondary mirror.

Figure 24

shows the Airy pattern, both without obscurations, and with a central obscuration and spiders in a setup typical of a telescope.

In addition, there may be phase errors in the exit pupil, because of the existence of any one of the sources of aberration discussed above. For general use, ϕ is often expressed as an series, where the expansion is over a set of orthogonal polynomials for the aperture which is being used. For circular apertures with (or without) a central obscuration (the case most often found in astronomy), the appropriate polynomials are called *Zernike* polynomials. The lowest order terms are just uniform slopes of phase across the pupil, called tilt, and simply correspond to motion in the image plane. The next terms correspond to the expressions for the OPD which we found above for focus, astigmatism, coma, and spherical aberration, generalized to allow any orientation of the phase errors in the pupil. Higher order terms correspond to higher order aberrations.

Figure 25

shows the form of some of the low order Zernike terms: the first corresponds to focus aberration, the next two to astigmatism, the next two to coma, the next two to trefoil aberration, and the last to spherical aberration.

A wonderful example of the application of all of this stuff was in the diagnosis of spherical aberration in the Hubble Space Telescope, which has been corrected in subsequent instruments in the telescope, which introduce spherical aberration of the opposite sign. To perform this correction, however, required and accurate understanding of the amplitude of the aberration. This was derived from analysis of on-orbit images, as shown in Figure 26.

Note that it is possible in some cases to try to recover the phase errors from analysis of images. This is called *phase retrieval*. There are several ways of trying to do this, some of which are complex, so we won't go into them, but it's good to know that it is possible. But an accurate amplitude of spherical aberration was derived from these images. This derived value was later found to correspond almost exactly to the error expected from an error which was made in the testing facility for the HST primary mirror, and the agreement of these two values allowed the construction of new corrective optics to proceed...

Some figures from HST Optical Systems Failure Report. Link to full report, and a NASA summary.

Understand the principles of diffration optics and, in particular, how diffraction scales with wavelength and aperture diameter. Know the terminology: optical transfer function, pupil function, and how phase errors across the pupil can be decomposed into a series of Zernike polynomials.

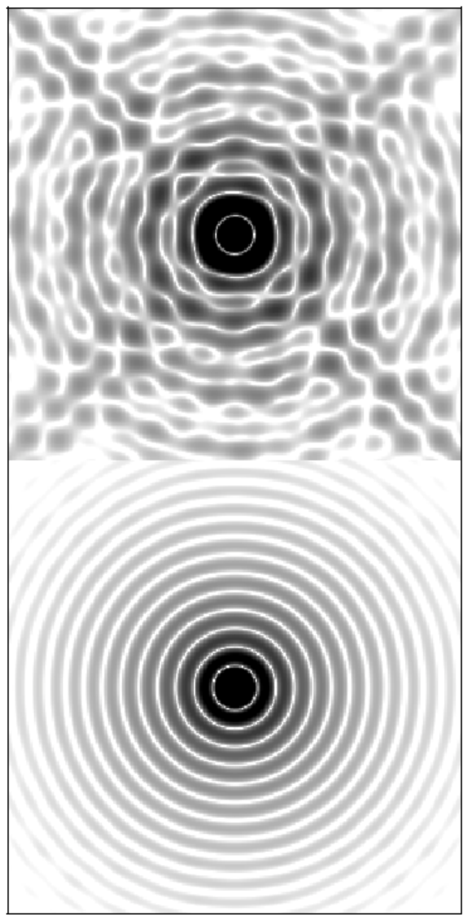


Figure 24: The Airy pattern, with and without obscurations

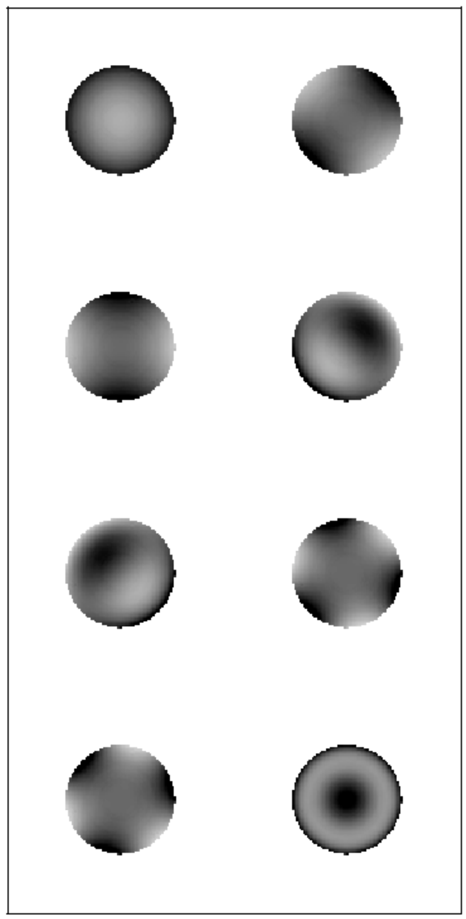


Figure 25: Images of Zernike terms, orders 4-11.

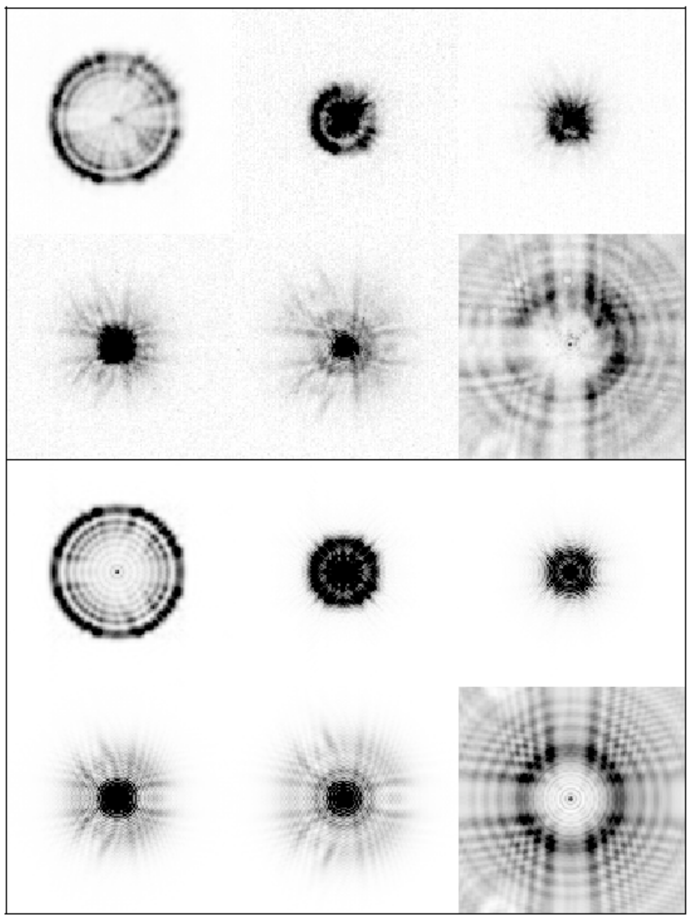


Figure 26: HST images and corresponding models

7.8 Adaptive Optics

The goal of *adaptive optics* is to partially or entirely remove the effects of atmospheric seeing. Note that these day, this is to be distinguished form *active optics*, which works at lower frequency, and whose main goal is to remove aberrations coming from the change in telescope configuration as the telescope moves (e.g., small changes in alignment from flexure or sag of the primary mirror surface as the telescope moves). Active optics generally works as frequencies less than (usually significantly) 1 Hz, whereas adaptive optics must work at 10 to 1000 Hz. At low frequencies, the active optics can be done with actuators on the primary and secondary mirrors themselves. At the high frequencies reqiured for adaptive optics, however, these large mirrors cannot respond fast enough, so one is required to form a pupil on a smaller mirror which can be rapidly adjusted; hence adaptive optics systems are really separate astronomical instruments.

Many adaptive optics systems are functioning and/or under development: see ESO/VLT adaptive optics, CFHT adaptive optics, Keck adaptive optics, Gemini adaptive optics, http://www.cfht.hawaii.edu/Instruments/Imaging/AOB/other-aosystems.html

The basic idea of an adaptive optics system is to rapidly sense the wavefront errors and then to correct for them on timescales faster than those at which the atmosphere changes. Consequently, there are really three parts to an adaptive optics system:

1. a component which senses wavefront errors,
2. a control system which figures out how to correct these errors, and
3. an optical element which receives the signals from the control system and implements wavefront corrections.

There are several methods used for wavefront sensing. Two ones in fairly common use among today's adaptive optics system are Shack-Hartmann sensors and wavefront curvature sensing devices. In a Shack Hartman sensor, as shown in Figure 27,

an array of lenslets is put in a pupil plane and each lenslet images a small part of the pupil. Measuring image shifts between each of the images gives a measure of the local wavefront tilts. Wavefront curvature devices look at the intensity distribution in out-of focus images. Other wavefront sensing techniques include pyramid wavefront sensors and phase diversity techniques. Usually, a star is used as the source, but this is not required for some wavefront sensors (i.e. extended source can be used).

To correct wavefront errors, some sort of deformable mirror is used. These can be generically split into two categories: segmented and continuous faceplate mirrors, where the latter are more common. A deformable mirror is characterized by the number of adjustable elements: the more elements, the more correction can be done. LCD arrays have also been used for wavefront correction.

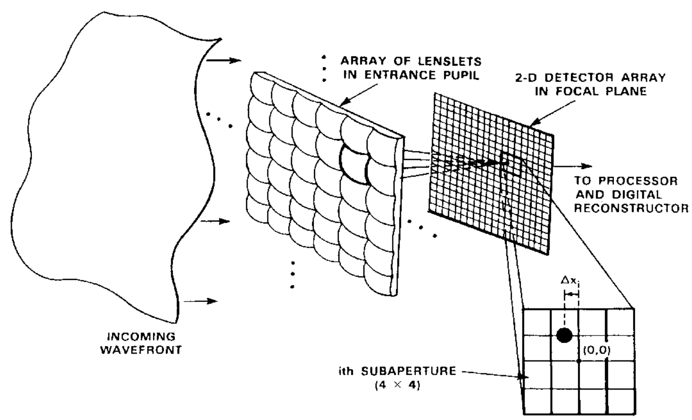


Figure 3 Principle of Hartmann-Shack wavefront sensor. The lenslets array produces an array of star images on the 2-dimensional detector array (generally a CCD or Intensified CCD). Tilt variations in the incoming, distorted wavefront result in position variations ($\Delta x, \Delta y$) of the star images on the detector. These are measured and fed to a digital processor which reconstructs the wavefront distortions. (From Murphy 1992; reprinted with permission of Lincoln Laboratory, MIT, Lexington, MA.)

Figure 27: Shack-Hartmann sensor

In general, it is very difficult to achieve complete correction even for ideal performance, and one needs to consider the effectiveness of different adaptive optics systems. This effectiveness depends on the size of the aperture, the wavelength, the number of resolution elements on the deformable mirror, and the quality of the site. Clearly, more resolution elements are needed for larger apertures. Equivalently, the effectiveness of a system will decrease as the aperture in increased for a fixed number of resolution elements. One can consider the return as a function of Zernike order corrected and aperture size. For large telescopes, you'll only get partial correction unless a very large number of resolution elements on the deformable mirror are available. The following table gives the mean square amplitude, Δ_j, for Kolmogorov turbulence after removal of the first j terms; the rms phase variation is just $\sqrt{\Delta_j}/2\pi$. For small apertures, you can make significant gains with removal of just low order terms, but for large apertures you need very high order terms. Note various criteria for quality of imaging, e.g. $\lambda/4$, etc.

Z_j	n	m	Expression	Description	Δ_j	$\Delta_j - \Delta_{j-1}$
Z1	0	0	1	constant	1.030 S	
Z2	1	1	$2r\cos\phi$	tilt	0.582 S	0.448 S
Z3	1	1	$2r\sin\phi$	tilt	0.134 S	0.448 S
Z4	2	1	$\sqrt{3(2r^2-1)}$	defocus	0.111 S	0.023 S
Z5	2	2	$\sqrt{6r^2\sin 2\phi}$	astigmatism	0.0880 S	0.023 S
Z6	2	2	$\sqrt{6r^2\cos 2\phi}$	astigmatism	0.0648 S	0.023 S
Z7	3	1	$\sqrt{8(3r^3-2r)\sin\phi}$	coma	0.0587 S	0.0062 5
Z8	3	1	$\sqrt{8(3r^3-2r)\cos\phi}$	coma	0.0525 S	0.0062 S
Z9	3	3	$\sqrt{8r^3\sin 3\phi}$	trifoil	0.0463 S	0.0062 S
Z10	3	3	$\sqrt{8r^3\cos 3\phi}$	trifoil	0.0401 S	0.0062 S
Z11	4	0	$\sqrt{5(6r^4-6r^2+1)}$	spherical	0.0377S	0.0024 S

r = distance from center circle; $\phi =$ azimuth angle; $S = (D/r_0)^{5/3}$.

Another important limitation is that one needs an object on which you can derive the wavefront. Measurements of wavefront are subject to noise just like any other photon detection so bright sources may be required. This is even more evident when one considers that you need a source which is within the same isoplanatic patch as your desired object, and when you recall that the wavefront changes on time scales of milliseconds. These requirements place limitations on the amount of sky over which it is possible to get good correction. It also places limitations on the sorts of detectors which are needed in the wavefront sensors (fast readout and low or zero readout noise!).

band	λ	r_0	τ_0	τ_{det}	V_{lim}	θ_0	Coverage (%)
U	0.365	9.0	.009	.0027	7.4	1.2	1.8 E-5
B	0.44	11.4	.011	.0034	8.2	1.5	6.1 E-5
V	0.55	14.9	.015	.0045	9.0	1.9	2.6 E-4
R	0.70	20.0	.020	.0060	10.0	2.6	0.0013
I	0.90	27.0	.027	.0081	11.0	3.5	0.006
J	1.25	40	.040	.0120	12.2	5.1	0.046
H	1.62	55	.055	.0164	13.3	7.0	0.22
K	2.2	79	.079	.024	14.4	10.1	1.32
L	3.4	133	.133	.040	16.2	17.0	14.5
M	5.0	210	.21	.063	17.7	27.0	71
N	10	500	.50	.150	20.4	64	100

Conditions are: 0.75 arcsec seeing at 0.5 μ; $\tau_{det} \sim 0.3\ \tau_0 = 0.3r/V_{wind}$; $V_{wind} = 10$ mlsec; $H = 5000$; photon detection efficiency (includes transmission and QE) = 20%; spectral bandwidth = 300 nm; $SNR = 100$ per Hartmann-Shack image; detector noise = $5e^-$.

The isoplanatic patch limitation is severe. In many cases, we might expect non-opticmal performance if the reference object is not as close as it should be ideally.

In most cases, both because of lack of higher order correction and because of reference star vs. target wavefront differences, adaptive optics works in the partially correcting regime. This typically gives PSFs with a sharp core, but still with extended wings.

The problem of sky coverage can be avoided if one uses so-called laser guide stars. The idea is to create a star by shining a laser up into the atmosphere. To date, two generic classes of lasers have been used, Rayleigh and sodium beacons. The Rayleigh beacons work by scattering off a layer roughly 30 km above the Earth's surface; the sodium beacons work by scattering off a layer roughly 90 km above the Earth's surface. Laser guide stars still have some limitations. For one, the path through the atmosphere which the laser traverses does not exactly correspond to the path that light from a star traverses, because the latter comes from an essentially infinite distance; this leads to the effect called focal anoisoplanatism. In addition, laser guide stars cannot generally be used to track image motion since the laser passes up and down through the same atmosphere and image motion is cancelled out. To correct for image motion, separate tip-tilt tracking is required.

Note that even with perfect correction, one is still limited by the isoplanatic patch size. As one moves further and further away from the reference object, the correction will gradually degrade, because a different path through the atmosphere is being probed.

To get around this, one can consider the use of multiple laser guides stars (laser guide star constellation) to characterize the atmosphere over a broader column. However, if this is done, one cannot correct all field angles simultaneously at the telescope pupil, because the aberrations are different for different field angles. Instead, one could choose to correct them in a plane conjugate to the location of the dominant source of atmospheric aberration. This is the basis of a *ground layer adaptive optics (GLAO)* system, where a correction is made for aberration in the lower atmosphere.

In principle, even better correction over a wider field of view is possible with *multiple* deformable mirrors, giving rise to the concept of *multi-conjugate* adaptive optics (MCAO) systems. In such systems, each adaptive optic would correct at a different location in the atmosphere.

Systems with single laser guide stars have certainly been tested and appear to work; but remember, only over an isoplanatic patch, and often with partially corrected images. Several implementations of system with multiple guide stars actually exist (at VLT and Keck?) to allow sampling of a larger cylinder/cone through the atmosphere; some of these are designed to correct at particular layers to maximize FOV, e.g. ground layer adaptive optics (GLAO). The bulk of adaptive optics work has been done in the near-IR.

Extreme (high-contrast) AO.

A variant on adaptive optics: lucky imaging.

Science with adaptive optics. Typical AO PSFs. Morphology vs. photometry.

7.8.1 AO Examples

Gemini AO animation video

- http://www.alpao.com/Applications/Adaptive_optics_for_Astronomy.htm
- Galactic center AO (see bottom of page, note scale)
- Neptune.
- Neptune movie.
- on/off sun image.
- Simulated seeing on bench
- Young star video

Understand the principles of how adaptive optics systems work. Understand the challenges of adaptive optics: getting sufficient light for wavefront sensing, isoplanatic patch size, partial correction, and how they can be at least partially addressed, e.g., with laser guide stars and multiconjugate adaptive optics.

8 INSTRUMENTATION

Often, astronomers use additional optics between the telescope and their detector. These, in conjunction with a detector, make up an *instrument*.

8.0.1 Location of optics

Before going into specifics, consider the effect of placing optics at different locations within an optical system, like a telescope.

Optics placed in or near a focal plane will affect images at different field angles differently. Optics in a focal plane will not affect the image quality at any given field angle; however, such optics might be used to control the location of an image of the *pupil* of the telescope.

Optics placed in or near a pupil plane will affect images at all field angles similarly, and will have an effect on the image quality.

Another important general consideration: throughput! All surfaces lose light at some level ....

Understand the implications of putting optics in different locations, e.g., in/near the focal plane vs in/near a pupil.

8.0.2 Refractive optics and chromatic aberration

In many instruments, lenses are used rather than mirrors: they can be cheaper and lead to more compact designs. Recall, however, that when lenses are used, chromatic effects will arise, because the index of refraction of glasses changes with wavelength. While they can often be minimized by the use of use of multiple elements to make achromatic combinations, they are not always negligible. In particular, if an instrument is used at multiple wavelengths, some refocussing may be required.

8.0.3 Field Flatteners

As we've discussed, all standard two-mirror telescopes have curved focal planes. It is possible to make a simple lens to correct the field curvature. We know that a plane-parallel plate will shift an image laterally, depending on the thickness of the plate. If we don't want to affect the image quality, only the location, we want the correcting element to be located near the focal plane.

Consequently, we can put a lens right near the telescope focal plane to flatten the field. For a field which curves towards the secondary mirror, one finds that the correct shape to flatten the field is just a plano-concave lens with the curved side towards the secondary. Often, the field flattener is incorporated into a detector dewar as the dewar window.

8.0.4 Focal plane reimagers

A focal reimager is a reimaging system which demagnifies/magnifies the telescope focal plane.

Motivation: why might you want to magnify or demagnify focal plane?

In a simple form, it consists of two lenses: a collimator and a camera lens. The collimator lens is placed such that the telescope focal plane is put at the focal length of the collimator, so that it converts the telescope beam into a collimated beam (note that the focal ratio of the collimating lens itself will be larger than that of the telescope so that the beam underfills the lens to allow for off-axis light as well). The camera lens then refocuses the light light with the desired focal ratio. The magnification of the system is given by:

$$m = \frac{f_{camera}}{f_{collimator}}$$

Consequently, the scale in the image plane of the focal reimager is just the scale in the telescope focal plane multiplied by the ratio of the focal ratio of the camera to that of the telescope.

Note that with a focal plane reimager, one does not necessarily get a new scale "for free". The focal reimaging system may introduce additional aberrations giving reduced image quality. In addition, one always loses some light at each additional optical surface from reflection and/or scattering, so the more optics in a system, the lower the total throughput.

Note that it is possible to do focal reduction/expansion *without* reimaging, i.e., by putting optics in the converging beam.

Understand the basic design and effect of a focal plane reimager. Be able to draw some rays and to determine if a reimager magnifies or demagnifies.

8.0.5 Pupil reimagers

Often, an additional lens, called a *field lens* is placed in or near the telescope focal plane. This does not affect the focal reduction but is used to reimage the telescope pupil somewhere in the reimager. One reason this may be done is to minimize the size that the collimator lens needs to be to get off-axis images. The size of the field lens itself depends on the desired size of the field that one wishes to reimage.

Another use of reimaging the pupil is when one is building a *coronagraph*, an imaging system designed to observe faint sources nearby to very bright ones. The problem in seeing the faint source is light from the bright one, both from scattered light, from diffraction, and sometimes, from detector effects (e.g., charge bleeding in a CCD). A partial solution is to put an occulting spot in the telescope focal plane which removes most of the light from the bright object. However, the diffraction structure is still a problem. It turns out you can remove this by reimaging the pupil after the occulting spot and putting a mask in around the edges which are the source of the diffraction; this mask is called a Lyot stop. The resulting image in the focal plane of the focal reducer is free of both bright source and diffraction structure.

Note that for really high contrast imaging, you also need to consider other sources of far-field light including light scattered from small-scale features on optical elements, and far-field light from seeing. Minimizing the former required very smooth optics, while minimizing the latter requires high-performance adaptive optics (e.g. "extreme-AO").

Pupil reimagers are also widely used in IR systems to reduce emission via cold pupil stops. The issue here is that the telescope itself contributes infrared emission which acts as additional background in your observations. There is little you can do about emission from the primary, since you need to see light from the primary to see your object! However, you can block out emission from regions of the pupils which are obscured already, for example, by the secondary and/or secondary support structures. To do this you put a mask in the pupil plane. Obviously, however, the mask needs to be colder than the telescope itself or else the mask would contribute the background, so it is usually placed within the dewar that contains the detector and camera optics (which also would otherwise glow!).

Understand why one might want to access a pupil in an instrument. Know the principles of a coronagraph works.

8.0.6 Filters

Filters are used in optical systems (usually imaging systems) to restrict the observed wavelength range. Using multiple filters thus provides color information on the object being studied. Generally, filters are loosely classified as broad band ($>\sim 1000$Åwide), medium band (100$<\sim 1000$ Å), or narrow band ($1 <\sim 100$ Å).

Perhaps a better distinction between different filters is by the way that they filter light. Many broad band filters work by using colored glass, which has pigments which absorb certain wavelengths of light and let others pass. Bandpasses can be constructed by using multiple types of colored glass. These are generally the most inexpensive filters.

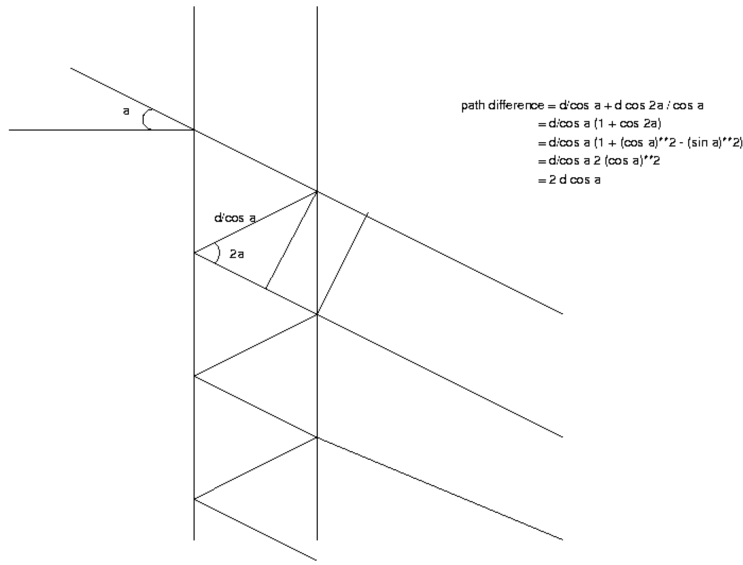


Figure 28: Schematic of an interference filter

A separate filter technique uses the principle of *interference*, giving what are called interference filters. They are made by using two partially reflecting plates separated by a distance d apart. The priciple is fairly simple:

When light from the different paths combines constructively, light is transmitted; when it combines destructively, it is not. Simple geometry gives:

$$m\lambda = 2nd \cos\theta$$

It is clear from this expression that the passband of the filter will depend on the angle of incidence. Consequently narrowband filters will have variable bandpasses across the field if they are located in a collimated beam; this can cause great difficulties in interpretation! If the filter is located in a focal plane or a converging beam, however, the mix of incident angles will broaden the filter bandpass. This can be a serious effect in a fast beam. Bandpasses of interference filters can also be affected by the temperature.

Since interference filters will pass light at integer multiples of the wavelength, the extra orders often must be blocked. This can be done fairly easily with colored glass.

The width of the bandpass of a narrowband filter is determined by the amount of reflection at each surface. Both the wavelength center and the width can be tuned by using multiple cavities and/or multiple reflecting layers, and most filters in use in astronomy are of this more complex type.

The same principles by which interference filters are made are used to make antireflection coatings.

Note filters can introduce aberrations, dust spots, reflections, etc; one needs to consider these issues when deciding on the location of filters in an optical system.

Understand how filters work and the difference between a colored glass filter and an inteference filter. Understand how the optical configuration can modify the bandpass of an interference filter.

8.0.7 Fabry-Perot Interferometer

A *Fabry-Perot* system makes use of a tunable interference filter. The filter is tuned in wavelength by adjusting one of

- the spacing,
- the index of refraction (usually changed by chaning the pressure), or
- the tilt of the interference filter.

A tunable interference filter is called an *etalon*. Often, etalons are made to provide very narrow bandpasses, on the order of 1Å.

A picture taken with a Fabry-Perot system covers multiple wavelengths because the etalon is located in the collimated beam between the two elements of the focal reducer. At each etalon setting, one observes an image which has rings of constant wavelength. By tuning the etalon to give different wavelengths at each location, one build up a "data cube", through which observations at a constant wavelength carve some surface. Consequently, to extract constant wavelength information from the Fabry-Perot takes some reasonably sophisticated reduction techniques. It is further complicated by the fact that to get accurate quantitative information, one requires that the atmospheric conditions be stable over the entire time when the data cube is being taken.

Know what a Fabry-Perot system is.

8.0.8 Spectrographs

A spectrograph is an instrument which separates different wavelengths of light so they can be measured independently. Most spectrographs work by using a *dispersive* element, which directs light of different wavelengths in different directions.

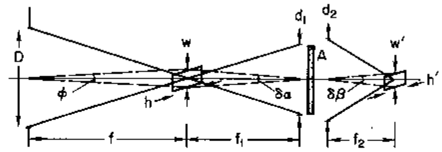


Fig. 12.4. Schematic layout of slit spectrometer with dispersing element *A*. See text, Section 12.II, for definitions of parameters.

Figure 29: Schematic of a basic spectrograph

A conventional spectrograph has a collimator, a dispersive element, a camera to refocus the light, and a detector. There are different sorts of dispersive elements with different characteristics; two common ones are prisms and diffraction gratings, with the latter the most commonly in use in astronomy.

The performance of a spectrograph is characterized by the *dispersion*, which gives the amount that different wavelengths are separated, and the *resolution*, which gives the smallest difference in wavelength that two different monochromatic sources can be separated. The dispersion depends on the characteristic of the dispersing element. Various elements can be characterized by the angular dispersion, $d\theta/d\lambda$, or alternatively, the reciprocal angular dispersion, $d\lambda/d\theta$. In practice, we are often interested in the linear dispersion, $dx/d\lambda = f_2 d\theta/d\lambda$ or the reciprocal linear dispersion, $d\lambda/dx = \frac{1}{f_2} d\lambda/d\theta$ where the latter is often referred to simply as the dispersion in astronomical contexts, and is usually specified in Å/mm or Å/pixel.

If the source being viewed is extended, it is clear that any light which comes from regions parallel to the dispersion direction will overlap in wavelength with other light, leading to a very confused image to interpret. For this reason, spectrographs are usually used with slits or apertures in the focal plane to restrict the incoming light. Note that one dimension of spatial information can be retained, leading to so-called *long-slit* spectroscopy. If there is a single dominant point source in the image plane, or if they are spaced far enough (usually in combination with a low dispersion) that spectra will not overlap, spectroscopy can be done in *slitless* mode. However, note that in slitless mode, one can be significantly impacted by sky emission.

The resolution depends on the width of the slit or on the size of the image in slitless mode, because all a spectrograph does is create an image of the focal plane after dispersing the light. The "width" of a spectral line will be given by the width of the slit or the image, whichever is smaller. In reality, the spectral line width is a convolution of the slit/image profile with diffraction. The spatial resolution of the detector may also be important.

Note that *throughput* may also depend on the slit width, depending on the seeing, so maximizing resolution may come at the expense of throughput.

Given a linear slit or image width, ω (or an angular width, $\phi = \omega/f$, where f is the focal length of the telescope) and height h (or $\phi' = h/f$), we get an image of the slit which has width, ω', and height, h', given by

$$h' = h\frac{f_2}{f_1}$$

$$\omega' = r\omega\frac{f_2}{f_1}$$

where we have allowed that the dispersing element might magnify/demagnify the image in the direction of dispersion by a factor r, which is called the anamorphic magnification.

Using this, we can derive the difference in wavelength between two monochromatic sources which are separable by the system.

$$\delta\lambda = \omega'\frac{d\lambda}{dx}$$

$$\delta\lambda = r\omega\frac{f_2}{f_1}\frac{d\lambda}{dx}$$

The bigger the slit, the lower the resolving power.

The resolution is often characterized in dimensionless form by

$$R \equiv \frac{\lambda}{\delta\lambda} = \frac{\lambda f_1}{r\omega f_2(d\lambda/dx)}$$

Note that there is a maximum resolution allowed by diffraction. This resolution is given aproximately by noting that minimum angles which can be separated is given by approximately λ/d_2, where d_2 is the width of the beam at the camera lens, from which the minimum distance which can be separated is:

$$\omega_{min} = f_2\frac{\lambda}{d_2}$$

The slit width which corresponds to this limit is given by:

$$\omega' = r\omega\frac{f_2}{f_1} = f_2\frac{\lambda}{d_2}$$

or

$$\omega = \frac{f_1}{r}\frac{\lambda}{d_2}$$

and the maximum resolution is

$$R_{max} = \frac{d_2}{f_2(d\lambda/dx)} = d_2\frac{d\theta}{d\lambda}$$

Understand how a typical astronomical spectrograph works. Know the functions of different elements: slit, collimator, dispersing element, camera. Be able to sketch rays showing how the spectrograph works. Understand the concepts of dispersion and resolution, and what about the spectrograph determines what these will be.

8.0.9 Astronomical spectrographs

Slitless spectographs: generally need to work at low dispersion (or narrow spectral range) to avoid spectrum overlap. Issue with background: since light from all field angles is included, this effectively disperses object light, but not background.

Long slit spectrographs: standard spectrograph as discussed above. Avoids spectrum overlap by limiting spectra to a line in the sky.

Slitlets: multiobject data. Break up single long slit into individual slitlets, avoiding overlap by the slitmask design. Note that each slitlit will have it's own wavelength calibration.

Fiber spectrographs: multiobject data. Use fibers to select objects, then line up the other ends of fibers into a pseudo-slit.

Integral field spectrographs. Get spectra information over 2D field. Either use fibers to accomplish, or optical configuration, e.g. with lenslets, oo image slicer.

Understand the different types of astronomical spectrographs.

8.0.10 Dispersing elements

Prisms

Perhaps the simplest conceptual dispersing element is a prism, which disperses light because the index of refraction of many glasses is a function of wavelength. From Snell's law, one finds that:

$$\frac{d\theta}{d\lambda} = \frac{t}{d}\frac{dn}{d\lambda}$$

where t is the base length, and d is the beamwidth. Note that prisms do not have anamorphic magnification ($r = 1$). The limiting resolution of a prism, from above is:

$$R_{max} = \frac{d_2}{f_2(d\lambda/dx)} = d_2\frac{d\theta}{d\lambda}$$

$$R_{max} = t\frac{dn}{d\lambda}$$

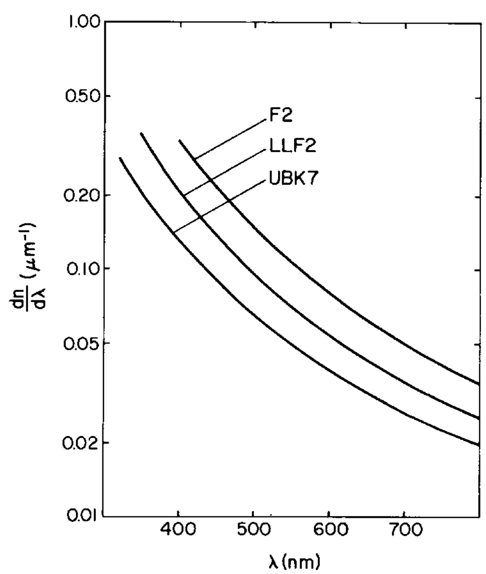


Fig. 13.1. Dispersion curves for three glasses from the Schott glass catalog.

Figure 30: $dn/d\lambda$ for typical glasses used in prisms

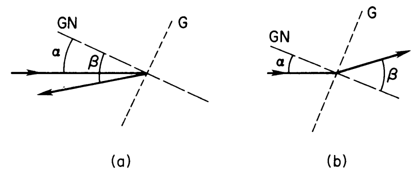


Fig. 13.2. Schematic showing angles of incidence α and diffraction β for (a) reflection grating and (b) transmission grating. See the discussion following Eq. (13.2.1) for the sign convention.

Figure 31: Schematic of a grating

One finds that $dn/d\lambda \propto \lambda^{-3}$ for many glasses.

So dispersion and resolution are a function of wavelength for a prism. In addition, the resolution offered by a prism is relatively low compared with other dispersive elements (e.g. gratings) of the same size. Typically, prisms have $R < 1000$. Consequently, prisms are rarely used as the primary dispersive element in astronomical spectrographs. They are occasionally used as cross-dispersing elements.

Gratings

Diffraction gratings work using the principle of multi-slit interference. A diffraction grating is just an optical element with multiple grooves, or slits (not to be confused with the slit in the spectrograph!). Diffraction gratings may be either transmissive or reflective. Bright regions are formed where light of a given wavelength from the different grooves constructively interferes.

Figure **??** outlines the principle of a grating; light comes in at some incidence angle, and light comes out at a variety of different angles of diffraction. At a given angle of diffraction, light of some wavelengths constructively interferes, while light at another wavelength destructively interferes.

The location of bright images is given by the *grating equation*:

$$m\lambda = \sigma(sin\beta + sin\alpha)$$

for a reflection grating, where σ is the groove spacing, m is the order, and α and β are the angles of incidence and diffraction as measured from the normal to the grating surface.

The dispersion of a grating can then be derived:

$$\frac{d\beta}{d\lambda} = \frac{m}{\sigma \cos\beta}$$

One can see that the dispersion is larger at higher order, and for a finer ruled grating. The equation can be rewritten as

$$\frac{d\beta}{d\lambda} = \frac{\sin\beta + \sin\alpha}{\lambda \cos\beta}$$

Fig. 13.3. Change in beamwidth due to anamorphic magnification of grating. See Eq. (13.2.3).

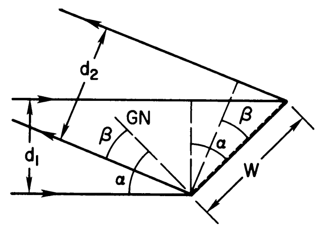


Figure 32: Schematic of anamorphic magnification in a grating

from which it can be seen that high dispersion can also be achieved by operating at large values of α and β. This is the principle of an echelle grating, which has large σ, and operates at high m, α and β, and gives high dispersion and resolution. An advantage of this is that one can get a large fraction of the light over a broad bandpass in a series of adjacent orders.

Typical gratings have groove densities between 300 and 1200 lines/mm. Echelle gratings have groove densities between 30 and 300 lines/mm.

Note that light from different orders can fall at the same location, leading to great confusion! This occurs when

$$m\lambda' = (m+1)\lambda$$

or

$$\lambda' - \lambda = \frac{\lambda}{m}$$

The order overlap can be avoided using either an *order-blocking filter* or by using a cross-disperser. The former is more common for small m, the latter for large m.

Grating order overlap

One can compare grating operating in low order, those operating in high order, and prisms, and one finds that higher resolution is available from gratings, and that echelles offer higher resolution than typical low order gratings.

One can derive the anamorphic magnification for a grating by looking at how β changes as α changes at fixed λ. One finds that:

$$r = \frac{d\beta}{d\alpha} = \frac{\cos\alpha}{\cos\beta} = \frac{d_1}{d_2}$$

where the d's are the beam diameters. Note that higher resolution occurs when $r < 1$, or $\beta < \alpha$.

The limiting resolution can be derived:

$$R_{max} = \frac{d_2}{f_2(d\lambda/dx)} = d_2\frac{d\beta}{d\lambda}$$

Fig. 13.4. Relation between blaze angle δ, grating normal GN, and angles of incidence and diffraction.

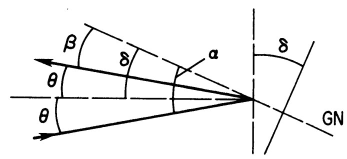


Figure 33: Schematic of how a blazed grating works

$$R_{max} = \frac{d_2 m}{\sigma \cos\beta} = \frac{mW}{\sigma} = mN$$

where W is the width of the grating ($= d_2/\cos\beta$), and N is the total number of lines in the grating.

We can also discuss *grating efficiency*, the fraction of incident light which is directed into a given diffracted order. One finds that for a simple grating, less light is diffracted into higher orders. However, one can construct a grating which can maximize the light put into any desired order by *blazing* the grating, which involves tilting each facet of the grating by some blaze angle. The blaze angle is chosen to maximize the efficiency at some particular wavelength in some particular order; it is set so that the angle of diffraction for this order and wavelength is equal to the angle of reflection from the grating surface. The *blaze function* gives the efficiency as a function of wavelength.

A special case of high efficiency is when the angle of incidence equals the angle of diffraction, i.e. the diffracted light at the desired wavelength comes back to the same direction of in the incoming light. This is called the Littrow configuration; high efficiency spectrographs often try to work close to this configuration.

Typical peak efficiencies of reflective diffraction gratings are of order 50-80%. Recently, a new technology for making diffraction gratings, volume phase holographic (VPH) gratings, as been developed, and these are attractive because they offer the possibility of very high efficiencies ($>$ 90% peak efficiency).

Understand the principle by which gratings work. Understand what different orders means, and how gratings can be blazed to maximize efficiency in a desired order. Understand how the groove density affects dispersion.

Grisms

A grism is a combination of a prism and a diffraction grating. These are combined such that light is dispersed, but light at a chosen central wavelength passed through the grism with direction unchanged. This feature allows grisms to be placed in an imaging system (e.g., in a filter wheel) to provide a spectroscopic (usually low resolution) capability.

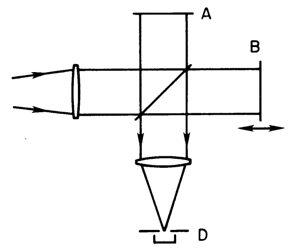


Fig. 13.15. Schematic of Fourier transform spectrometer. A, Fixed mirror; B, movable mirror; D, detector.

Figure 34: Schematic of a fts spectrograph

8.0.11 Operational items: using a spectrograph

Choice of dispersion: wavelength coverage vs. dispersion/resolution, available gratings, etc. Using grating tilt to select wavelength range.

Choice of slit width (science, seeing).

How to put object in slit. Imaging the slit. Slit viewing cameras.

(DEFER FOLLOWING TO SECTION ON DATA REDUCTION???)

Spectrograph calibration (not including basic detector calibration, to be discussed soon).

Wavelength calibration: correspondance between pixel position (in wavelength dimension) and wavelength. Arc lamps, wavelength solutions. Subtleties: extrapolation, line curvature, flexure (using skylines to calibrate).

Flux calibration: relative fluxes at different wavelengths. Spectrophotometric standards. Subtleties: differential refraction

Spectral extraction: object extraction and sky subtraction. Subtleties: S-distortion, differential refraction: spectral traces. Issues: variation of focus along slit and implications for sky line subtraction, scattered light.

Relative fluxes along slit: slit width variations.

Examples of typical spectra: line lamps, flat fields, stellar spectra, galaxy spectra. Night sky emission.

8.0.12 Non-dispersive spectroscopy

It is also possible to use interference effects to measure spectral energy distributions instead of a dispersing element. The Fabry-Perot is an example of such a type of instrument, although it does not record all wavelengths simultaneously.

Another instrument which uses interference to infer spectroscopy information is the Fourier Transform Spectrometer (FTS), which is basically a scanning Michaelson interferometer. The light from the source is split into two parts using a beamsplitter. One part of light is reflected off a fixed flat mirror and the other is reflected off a mirror which can be moved laterally. The two images are combined to form fringes. The fringe pattern changes as the path length of the second beam is changed. The intensity modulation for a given wavelength (λ) or wavenumber ($k = 2\pi/\lambda$) is given by:

$$T(k, \Delta x) = \frac{T_{max}}{2}[a + \cos(2k\Delta x)]$$

and the flux after integrating over all wavelengths is:

$$F(\Delta x) = C \int I(k) T(k, \Delta x) dk = C \int I(k) cos(2k\Delta x) dk$$

where $I(k)$ is the input spectrum. Consequently it is possible to recover the input spectrum by taking the Fourier cosine transform of the recorded intensity. In practice, a discrete Fourier transform is used.

The FTS requires scanning in path spacing. But unlike the Fabry-Perot, it yields information on intensity at all wavelengths simultaneously.