Notes for Instrumentation for Astronomy
UC Santa Cruz, Winter 2023

Aditya Sengupta

July 11, 2023

Contents

Instrumentation for Astronomy **Winter 2023**

Lecture 1: Introduction

Lecturer: Connie Rockosi *10 January* *Aditya Sengupta*

Note: *LaTeX format adapted from template for lecture notes from CS 267, Applications of Parallel Computing, UC Berkeley EECS department.*

1.1 Overall goals

This class will help us become both instrument designers and informed observational astronomers. We'll get the tools to identify science opportunities with new instruments, and to help make them happen as informed members of science teams.

We'll be talking a lot about signal-to-noise; a lot of what makes great instruments so great is that they've made a big jump in S/N for some kind of data well matched to a new science opportunity. Another aspect is delivering the desired resolution, FOV, etc. for science on a new satellite or telescope. We'll also learn why we can't always have the instrument we want.

The first part of this class will focus on optical and near-IR regimes. We'll learn about detectors, optics, PSFs, gratings/diffraction, spectrographs, and fibers. The second part will focus more on measuring flux, photometric systems, calibration and so on, with some time spent on sub-mm and radio astronomy and high-energy detectors.

1.2 Observables

We're going to start by building vocabulary.

What we do in order to collect data is project the sky onto our detector. A spectrograph is similar: it images the entrance aperture onto the detector. So we have to talk about the optics that capture that, the detector that records it, and numerical issues in reading it later.

The specific luminosity L_ν is the energy per second per frequency, and the total energy per unit time is $\mathrm{d}E = L_\nu \mathrm{d}t\mathrm{d}\nu$. We observe Flux (flux density) which is the fraction of the luminosity intercepted by our collector. Flux has units of ergs/sec/cm^2/Hz, or Janskys, where $1Jy = 10^{-26}W/m^2/Hz$. The flux you measure depends on the distance from the source.

We don't usually measure in ergs, but instead in magnitudes:

$$m = -2.5 \log \frac{\int F_\nu S_\nu \, \mathrm{d}(\log \nu)}{\int S_\nu \, \mathrm{d}(\log \nu)} - zp. \tag{1.1}$$

An extended source of arbitrary shape and size gives us a received energy per second of I_ν, where we divide by a solid angle here. We usually deal with surface brightness in magnitudes per square arcsecond. Flux is specific intensity integrated over solid angle of the source: $F_\nu = \int I_\nu \cos\theta \mathrm{d}\Omega$. For a point source, we measure flux, and for an extended source, we measure total flux over the entire area, or surface brightness.

Surface brightness is independent of distance from the source (until cosmology changes geometry), but flux is dependent.

We describe observations by wavelength range because there's a huge amount of variation in the nature of the background. The sky brightness gets stronger at longer wavelengths, up to the mid-IR – it's 10,000 times brighter in the K band than the B band. After that, it starts dropping off again. The contribution of the sky magnitude to an I-band observation of a 1 arcsec source is 18; we've been able to achieve 24.5, or about 400 times fainter than the sky.

1.3 Detectors

Detectors detect photons, and the photons we're interested in can span 12 orders of magnitude. The highest-energy things are gamma rays and x-rays. At the other end, we have radio and consumer electronics.

There's a few different types of photon detectors:

- photoconductors, where photons interact with electrons and boost their energy over a bandgap threshold to create current. CCDs, near-IR detectors, and several others fall under this
- photoemissive detectors, which use the photoelectric effect to eject an electron out of a metal photocathode. These are things like photomultiplier tubes. These are good at collecting high-energy photons, but they're less efficient.
- Thermal detectors, like bolometers, which absorb low-energy photons and create heat. They generate a small change in current that, over a material with high $\frac{\mathrm{d}R}{\mathrm{d}T}$, creates an appreciable potential difference.
- There's also antennas, which detect coherent radiation.

1.4 CCDs

Most of our detectors are silicon, because we have a very-well developed industry for making things out of silicon! In silicon, the electron wavefunctions overlap, and due to Pauli exclusion, the levels split. The levels become states in energy bands. In a semiconductor, there's a conduction band and a valence band, with a band gap energy between them. If you excite an electron in the valence band, it can cross over into the conduction band, where it can move freely in the lattice in response to an electrical potential gradient. Materials with small/less-than-zero band gaps are metals, where you have a continuous band with many unfilled states and it's easy to put electrons there with thermal energy – this isn't particularly useful for us to be able to detect if anything's happened. If the band gap is too large, you get an insulator, and you can't make it do anything without a ton of energy. If the band gap is just right, you get a semiconductor!

The Si bandgap is 1.11 eV, which means if you give an electron that much energy, it'll make the jump. A 5000Å photon has 2.25 eV, so it's got enough energy to get an electron to jump and we can detect it. This is why there's a limit at 1 μm in silicon detectors, because photons at that wavelength have 1.2 eV. If you want to go longer, into the mid-IR and beyond, you have to engineer semiconductors like HgCdTe (a mix of CdTe, 1.55 eV, and HgTe, 0 eV as it's a metal) with smaller bandgaps. We can also dope semiconductors to change the bandgap. Defects in the Si crystal lattice can also change the bandgap. These defects can result in phenomena like charge "traps" or blocked columns.

A MOS capacitor is a basic CCD unit (a pixel). There's a metal gate which is given positive voltage, and there's a depletion region with no free charge carriers and no current. When there's a photon impact, it's translated to energy given to electrons which can jump across to the metal gate.

(Something about MKIDs.)

A feature of semiconductors that work this way is *dark current*. The number of electrons in the conduction band is set by the thermal distribution at $\exp(-E_{gap}/kT)$. At $T = 0$K, all electrons are below the Fermi energy, and as T goes up, some start to occupy higher-energy states and eventually jump to the conduction band. At 300K, there are $1.38 \times 10^{13}\text{cm}^{-3}$ electrons in the conduction band for Si, but this goes down to about 10^{-18} at 77K. So cooling your detector helps! IR detectors with smaller band gaps have even bigger sensitivity to thermal electrons, so those go even colder, at liquid helium temperatures.

Modern CCDs have arrays of pixels that can be read out at amplifiers. To read out a CCD, each column transfers one pixel at a time into the serial register (a row of pixels turned 90 degrees), which in turn transfers one pixel at a time to the readout amplifier. We read out by 'dragging' each pixel to the edge. It's impractical to have one amplifier per pixel, so we work in arrays.

The transfer to the serial register and then to the readout amplifier happens with a sequence of gates, to which we apply voltages to move charges along the row and then the column. In a CCD, the serial register is just the same kind of thing as a pixel; this lets them be low-noise and cosmetically nice. In other detectors, like for the IR, we have to combine silicon with more complicated architectures.

To read this out to a value per pixel, we use a technique called "double-correlated sampling". We transfer the pixel charge to a capacitor with a tiny capacitance. We read the voltage across a known resistor to get a reference, transfer the pixel charge to see the new voltage across a known resistor due to the discharge of the capacitor, and close the switch to reset to the reference voltage. In this, we have two sources of noise:

1. Poisson fluctuations in the current through the output amplifier (read noise). This is irreducible no matter what sampling technique is used.
2. Thermal (Johnson) noise in the value across the resistor when we reset to the reference voltage. This is remedied by double-correlated sampling.

To lower read noise, we can average over current fluctuations in the output. This means we can incur a penalty in the time we take to read each pixel, so that we can average for some time, and get lower read noise, but there's an irreducible floor we can't go below.

Instrumentation for Astronomy **Winter 2023**

Lecture 2: SNR Hack Day

Lecturer: Connie Rockosi *12 January* *Aditya Sengupta*

We'll assume

- a photon counting detector
- sources of noise add in quadrature

In the optical and IR, photon arrival obeys Poisson statistics, where at a photon detection rate μ, the probability of observing n photons in time t is $P(n) = \frac{(\mu t)^n}{n!} e^{-\mu t}$.

We measure some number of photons from our source. We care about the underlying physical properties of the source, primarily luminosity. We observe flux, which is related to luminosity by distance to the source, collecting area, and other factors that set the fraction of photons from the source we can detect. The rate μ is the fundamental physical quantity we'd like to estimate.

Our object creates a sub-image on the detector, and we care about the number of photons falling within that sub-image. We integrate for time t and gather photons from the source and the sky. We count up all the electrons in the aperture, we estimate the rate from the sky by using some other area on the detector that doesn't have the source, and we rescale and subtract to get the number of photons from the source in time t.

The signal-to-noise equation is

$$SNR = \frac{S_\gamma t}{\sqrt{S_\gamma t + B_\gamma t A_\Omega + d_{e-} N_{pix} t + R N_{e-}^2 N_{pix,eff}}} \tag{2.1}$$

This doesn't include systematics, but we're considering irreducible uncertainties here. Here,

- S_γ is the photons from the source per second
- t is the time spent collecting signal
- B_γ is the photons from the background per second/solid angle/unit wavelength. We multiply it by the solid angle on the sky, A_Ω.
- N_{pix} is the number of pixels on the detector that we're summing to measure the signal, and $N_{pix,eff}$ is that number after binning.
- d_{e-} is the dark current, the number of charge carriers per second due to thermal energy.
- RN_{e-} is read noise: Gaussian uncertainty in the measurement of the signal in any pixel.

To improve SNR, we can try and get

- a larger photon rate: a larger telescope, reduced transmission and reflection losses

- a lower background rate
- a smaller area
- longer integration times

Instrumentation for Astronomy **Winter 2023**

Lecture 3: Geometric optics and aberrations

Lecturer: Connie Rockosi *17 January* *Aditya Sengupta*

We have two views of light: as a ray, or as a wave. As waves, we observe expanding spherical wavefronts (surfaces of constant optical path length from source/constant light travel time), and as rays, we observe lines perpendicular to wavefront, along which light travel time is constant.

Wavefronts expand/rays propagate at a speed of c/n where n is the index of refraction of the medium they're propagating in.

The optical path length is proportional to the time it takes for light to travel between two points, and it's the integrated refractive index:

$$OPL = \int_a^b n(s)\mathrm{d}s \tag{3.1}$$

for s, a position along an optical path. If n is constant, $OPL = ns$.

As wavefronts expand out from the source, their radius of curvature gets bigger. A wavefront with a small radius of curvature is close to its source, and one with a large radius of curvature is far from its source. Wavefronts that are really really far away from their source are essentially flat, so we can think about them as planes.

A telescope essentially transforms an incoming flat wavefront into an image some distance away from the primary mirror. The length over which it does this is called the *focal length* f, and the telescope's ratio f/# is defined as the focal length divided by the primary mirror diameter. Another useful relationship is that the object size φ in angular units is related to the image size in linear units (like mm) by $d = \varphi f$.

At the telescope focal plane, if two points on the sky are separated by an angle θ, their separation in the image is θf. Sources at the center of the field of view have light come in parallel to the optical axis, and those at some angular separation come in "off-axis".

Systems with small focal ratios are called "fast" and those with large focal ratios are called "slow". This refers to the speed at which you can accumulate signal-to-noise. Smaller f concentrates signal in a smaller area of the focal plane while noise terms remain relatively constant.

If we have a ray from a source at a point B and an image is formed at B', we say B and B' are *conjugate*. We'll use a convention where the optical axis is along z, the perpendicular axis in the page is y, and the axis out of the page is x. So a mirror is a function $z(y)$, or in 3D it could be $z(x, y)$.

Recall Snell's law, $n \sin i = n' \sin i'$, and the lens/image equation, $\frac{1}{s} + \frac{1}{s'} = \frac{2}{R} = \frac{1}{F}$. Here, $F = \frac{R}{2}$ is the focal length of a mirror, and this is what defines F. It's the image location for an object at infinity.

Fermat's principle tells us that rays follow paths between conjugate points such that for small changes in path, the optical path length is constant to first order:

$$\delta OPL = \delta \int n\mathrm{d}s = 0. \tag{3.2}$$

We can use this to derive Snell's law and the lens equation.

What Fermat's principle is saying is all paths from B to B' are adjacent to each other and have the same OPL. If a ray from B doesn't go through B', it can't have the same OPL. Therefore rays with the same OPL get us a perfect image of the source at the destination.

We can use this idea to see what happens when we don't have perfect OPL. What are the things that can go wrong with CCDs?

- **Charge transfer inefficiency**: on a 2k by 2k CCD, if there's an efficiency of 0.9999, there's some accumulated charge if you do this 2000 times per row, and this adds up over time. We can read extra pixels to quantify this.
- **Pixel response variations**: flat fields are usually wavelength-dependent. It's hard to calibrate these over a broad wavelength range. What we want is a source illuminating the detector evenly in wavelength and space, but this is hard to do. People use many sky images and superimpose them and median out the objects, to use the sky as the flat field.
- **Cosmic rays**: muons and stray electrons may have enough energy to tunnel across your detector's band gap and show up in your data, and will usually come in at an angle and will therefore create streaks on your image. Straight tracks are muons and squiggles are electrons.
- **Saturation**: if you fill too much charge into a given pixel, it'll spill over into its neighbour. This is not the same as maxing out the analog-to-digital converter!
- **Traps and bad columns**: some pixels don't transfer their charge to their neighbour correctly.
- **Fringing**: this is what happens when a photon doesn't get absorbed and keeps bouncing and reflecting around the detector. The *quantum efficiency* is the probability that a photon will be converted to an electron, and it's a function of this and of reflection coating.
- **Brighter-fatter effect**: Thick CCDs, which have become more common over the last decade, exhibit this problem. As electrons hit the detector, their accumulation at the other end creates charge that changes the electrical potential of the CCD overall, and over the distance of a thick CCD, this can change the probability that an electron makes it through at the right position.
- **"Tree rings"**: as silicon crystals grow, their resistivity can change, which can cause variations in the effective pixel size.

Instrumentation for Astronomy **Winter 2023**

Lecture 4: Optical aberrations and telescope design

Lecturer: Connie Rockosi *24 January* *Aditya Sengupta*

In a paraboloid mirror, angular aberrations increase with the off-axis angle of the source. There are two sources of this: *astigmatism* and *coma*. Faster (smaller) f/ratios have worse aberrations (but better SNR) and coma is the larger aberration. Coma increases linearly with the off-axis angle of the source, and the slope is inversely proportional to the f-ratio. Astigmatism follows similar trends but is more quadratic, and its curve remains lower than that of coma.

We don't get perfect images because all rays do not focus at the same point. This is wavelength-dependent, because there are lenses refracting the light. On a spot diagram, we can see where the rays end up near the focus. In good imaging systems, these will remain well within the footprint of the fiber optic collecting the light. We define the best focus as the best compromise image over the focal plane, averaging over all field angles.

To define our mirror surface in 2D, $z(y)$, we use Fermat's principle: we require that every point on the surface have the same optical path length. Consider two rays: one is on the optical axis and one is at a height y. For the first ray, the input and reflected rays are on the optical axis, so the OPL from B′ to the mirror and back to B′ is $2f$. At height y, the ray travels $f - z(y)$, hits the mirror, and travels back a distance l. This gives us

$$2f = l + (f - z(y)). \tag{4.1}$$

If we write l in terms of z using a right triangle, we can solve this for the curve; we get $y^2 = 4fz$. This is the equation for a parabola.

This holds if we have an object and image at the same point. If this isn't the case, say we have a distance s from the object to the mirror and a distance s' back. Fermat's principle tells us that $s' + s = l' + l$, where we can write l, l' as

$$l^2 = y_0^2 + (s - z(y))^2 \tag{4.2}$$
$$(l')^2 = y_0^2 + (s' - z(y))^2. \tag{4.3}$$

Solving this, we get

$$y^2 - 2z\frac{b^2}{a} + z^2\frac{b^2}{a^2} = 0 \tag{4.4}$$

where $2a = s + s', b = ss'$. This is the equation of an ellipse centered at $(0, a)$ with axes a, b. The foci are at B, B', so you get a perfect image of B at B'.

In a Cassegrain telescope, the primary mirror is a paraboloid, so we get a perfect image of a source at infinity. The secondary mirror is a hyperboloid, with one focus at the image created by the primary and the other at

the focal plane of the telescope. A Cassegrain makes perfect images for an on-axis source, but not for any other object. A Gregorian telescope does the same thing but with an ellipsoidal secondary.

We can generalize the description of the surface to any conic. The equation is $y^2 - 2Rz + (1-e^2)z^2 = 0$. Here, R is the spherical radius of curvature at the steepest point, which by convention is the location of the optical axis. We sometimes talk about the conic constant $K = -e^2$.

We can also do this in reverse: with a specified surface $z(y)$, we can find the distance f along the optical axis at which an incoming ray at height y reflected/refracted by the surface crosses the optical axis. If f is independent of y, we get a perfect image. If we do this analysis for Δf (the change in f relative to the focus) for an object at infinity, we get

$$\Delta f = -\frac{(1+K)y^2}{4R} - \frac{(1+K)(3+K)y^4}{16R^3}. \tag{4.5}$$

This is zero only for $K = -1$, i.e. a parabola. For all other values of K, there is no point on the optical axis at which the conic surface makes a perfect image of the object.

For a sphere, $K = 0$, rays from larger y focus closer to the surface. This is *spherical aberration*. A famous example of spherical aberration is the Hubble Space Telescope, before the first servicing mission.

We can generalize this to 3D, where in place of y we use $r^2 = x^2 + y^2$. After a lot of math, we get the optical path length for an off-axis source:

$$\begin{aligned} OPL = {} & (-ns + n's') - y(n'\sin\theta' - n\sin\theta) \\ & + \frac{x^2}{2}\left[\frac{n'}{s'} - \frac{n}{s} - \frac{n'\cos\theta' - n\cos\theta}{R}\right] + \frac{y^2}{2}\left[\frac{n'\cos^2\theta'}{s'} - \frac{n\cos^2\theta}{s} - \frac{n'\cos\theta' - n\cos\theta}{R}\right] \\ & - \frac{x^2 y}{2}\left[\frac{n\sin\theta}{s}\left(\frac{1}{s} - \frac{\cos\theta}{R}\right) - \frac{n'\sin\theta'}{s'}\left(\frac{1}{s'} - \frac{\cos\theta'}{R}\right)\right] \\ & - \frac{y^3}{2}\left[\frac{n\sin\theta}{s}\left(\frac{\cos^2\theta}{s} - \frac{\cos\theta}{R}\right) - \frac{n'\sin\theta'}{s'}\left(\frac{\cos^2\theta'}{s'} - \frac{\cos\theta'}{R}\right)\right] \\ & + \frac{r^4}{8}[\text{something even longer}]. \end{aligned} \tag{4.6}$$

Let's break this down. The $(-ns + n's')$ term is the OPL of a ray through the center at $r = 0$. We're interested in the difference between the OPL of this ray and others. If $n = n'$, the $n'\sin\theta' - n\sin\theta$ term is always 0, because of Snell's law. We're left with

$$\text{OPL}(x, y) - \text{OPL}(x, y = 0) = A_1 y^2 + A_1' x^2 + A_2 y^3 + A_2' x^2 y + A_3 r^4. \tag{4.7}$$

This is itself a simplification; we've kept terms to third order and used the usual approximations for sin and cos. We can solve an s' that makes only one of A_1 and A_1' zero, or we can set $A_1 = -A_1'$, to balance out the aberration in x and y. This choice of s' gives us

$$A_1 = -A_1' = \tag{4.8}$$

Instrumentation for Astronomy **Winter 2023**

Lecture 5: Point Spread Functions and the Nyquist Limit

Lecturer: Connie Rockosi *31 January* *Aditya Sengupta*

PSFs are complicated! We care about this because Nyquist says we need to sample the PSF at some minimum number of pixels to recover what you need, but we want apertures that are as small as possible to minimize read noise, dark current, and so on. What the Nyquist limit is depends on the shape of the PSF; if it's not a smooth, well-behaved function, that'll change how we think about the minimum allowable aperture size.

The Nyquist limit says if you want to recover signals with frequency f, you need to sample at a rate of at least $2f$. We can see this by going back to Fourier transforms. Higher frequencies are faster wave oscillations, so you need fine sampling in time to capture them. Remember that a square wave and sinc are Fourier conjugate, which is useful because a square wave is sort of like a pixel. Also remember that the Fourier transform of a Gaussian is another Gaussian, where the spread of the frequency Gaussian is inversely related to that of the time Gaussian.

If we have a continuous function $f(t)$, we can sample it by a sampling function $s(t;\tau) = \sum_{n\in\mathbb{Z}} \delta(t - n\tau)$, and we have access to the sampled function $f(n\tau) = f(t)s(n\tau;\tau)$.

If we take the Fourier transform of the sampled version of $Q(t)$, we get a series of copies of the Fourier transform of $Q(t)$ spaced by the sampling frequency $F = \frac{1}{\tau}$. The larger the sampling frequency, the more spaced out the copies are in frequency space.

Let the Fourier transform of $Q(t)$ be *band-limited*, i.e. it has a maximum frequency. Lots of functions are not band-limited. Band-limited functions tend to be pretty smooth, without a lot of sharp corners, because we're saying they can be well approximated/exactly described by a sum of sine waves.

If we want to recover a function from its FT, and we know it's band-limited, we can take an inverse FT and we'll know we're getting everything even if we cut off everything past $f_{\max}$. But to do this, we need to sample at least twice as fast as $f_{\max}$. If we don't, we get aliasing.

What this gets us in the context of PSFs is: if we undersample the PSF (by having pixels that are too large) we won't be able to fit the center, width, etc.

A standard rule of thumb is that you need 2 pixels across the FWHM of the PSF. For a Gaussian, the FWHM is 2.35σ, so we want 1 pixel = 1.2σ. Most PSFs are not perfect Gaussians and are not as band-limited as we'd like, so 2 pixels isn't really good enough. For example, precision radial-velocity spectrographs sample much higher than 2 pixels across their line-spread functions.

Bickerton and Lupton: a way to get subpixel information! Set $f_{\max} = 0.5/\text{px}$. Do an inverse FT to pixel space, and cut off everything above the Nyquist limit, which gives us the same as convolution with a sinc. This convolution can reproduce the PSF everywhere, not just at the centers of pixels. The error from assuming a Gaussian PSF is band-limited is small, but it's nonzero.

The basic trick for recovering sub-pixel information is to have a very stable system, where you can recover it over a long period of time. With HST, whose PSF was undersampled, it's possible to be wrong about the center of the PSF by about a tenth of a pixel, so long-term characterization is necessary.

The effective PSF is the instrumental PSF convolved with the actual pixel response. This is necessary because pixels are less sensitive at the corners. Even though a pixel does not have uniform response, all of them are very alike, so we can observe the same star at many different locations in a pixel grid, and map the effective PSF using the changes since we can be sure those aren't intrinsic.

One more example of space-based data that's very undersampled is Kepler K2 photometry. We can see the impact of the Kepler spacecraft thrusters correcting pointing several times a day in the sawtooth pattern of the average pixel center. To fix this, they take out large-timescale variations by subtracting off a fit spline.

Ground-based imaging usually doesn't have such a stable PSF, unless you have AO running.

PSF photometry involves fitting a model that weights pixels with more signal in them more highly. The PSF is the weighting function. We can prove that a least-squares fit of the model parameters including the total flux is the minimum variance estimate of the actual flux, so we get the maximum S/N. To do this, we need a good PSF model; if we don't have that, the flux measurements may be biased.

Instrumentation for Astronomy **Winter 2023**

Lecture 6: Gratings and spectroscopy

Lecturer: Connie Rockosi *7 February* *Aditya Sengupta*

Diffraction patterns: if light comes in at an angle α, we get constructive interference when $\sin\theta = m\frac{\lambda}{D}$ where D is the slit spacing. If we have a bunch of slits repeating, we get an intensity pattern consisting of regularly spaced peaks in a sinc-squared envelope. For larger N (number of slits), in principle the maxima are more narrow. Peaks have width $\frac{\lambda}{ND}$ and spacing $\frac{\lambda}{D}$. Peak locations are wavelength-dependent, which is good for spectroscopy so we can see peaks at different locations, but this isn't the case at the central peak: those will all overlap, which is bad because that's where most of the intensity is.

A grating has grooves or facets, which are like slits but are small mirrors, and light is reflected off these grooves instead of passing through slits. Light diffracts as it reflects. The grating equation, $\sin\alpha + \sin\beta = m\frac{\lambda}{d}$, tells us the position of a constructive interference maximum at a wavelength λ, as a function of grating order m.

We have to worry about different orders of diffraction peaks overlapping. We can fix this by finding the *free spectral range*, i.e. the wavelength range such that there is no overlap from an adjacent order, and adding an "order-blocking" filter for this. This range is $\Delta\lambda = \frac{\lambda_1}{m}$.

We blaze the grating by tilting the reflective facets/grooves by the "blaze angle", which gives us the freedom to shift the maximum of the single-slit pattern to a different value of β such that $\frac{\mathrm{d}\beta}{\mathrm{d}\lambda} \neq 0$. This shifts the diffraction pattern in a wavelength-dependent way so we can have peak intensities at different positions.

We isolate the light from a target using a slit, so we don't get spectra from the whole field creating conflicting signals. How big should the slit be? We can imagine doing the reimager calculations we're used to on the slit width (rotating the diagram in our minds 90 degrees, so the slit width is into the page.) Different wavelengths will come in at different angles. The grating angular dispersion comes out to

$$\frac{\mathrm{d}\beta}{\mathrm{d}\lambda} = \frac{m}{d\cos\beta}. \tag{6.1}$$

Keck LRIS diagrams on the slides.

The spectrograph resolving power is $\frac{\lambda}{\Delta\lambda}$ where $\Delta\lambda$ is the resolution. We can find $\Delta\lambda$ using a formula for $\frac{\mathrm{d}\lambda}{\mathrm{d}\beta}$ adjusted to meet the Littrow condition. We need $\Delta\beta$, which we know how to calculate as it's just the slit width. Note that $\Delta\lambda$ is usually quoted as the FWHM (not sigma) of the image of the slit.

$$\frac{\lambda}{\Delta\lambda} = \frac{2\tan\delta}{\theta}\frac{D_{beam}}{D_{tel}} \tag{6.2}$$

There's no lines per millimeter in this formula, which is annoying when planning observations since that's how all spectrographs are specified.

Anamorphic magnification: the anamorphic factor is $r = \frac{\mathrm{d}\beta}{\mathrm{d}\alpha} = \frac{\cos\alpha}{\cos\beta}$.

Instrumentation for Astronomy **Winter 2023**

Lecture 7: IFUs, fibers, pupil stops

Lecturer: Connie Rockosi *14 February* *Aditya Sengupta*

7.1 Integral field units

Integral field spectroscopy involves getting spectra from an entire square field of light via mirror magic or lenslets. For example, KCWI uses slicers. It has a fan of rectangles and it's cured to make an image. Light comes from the telescope and hits each slice at a different angle, so each one gets its own slit that's dispersed by the spectrograph.

The other way to do this is with lenslets. An example of this is OSIRIS. We image light onto a lenslet array, and each lenslet makes a small image. Downstream of this, a mask helps us separate out each small image, and a collimator images this onto a fixed grating. The tradeoff in designing these is it's difficult not to make adjacent spectra overlap. It's possible to mitigate this by tilting the lenslet array, but after a certain point that won't work. So it's hard to do integral field spectroscopy onto small-spaced arrays, because this constrains your achievable resolution.

7.2 Fibers

Fibers in astronomical instruments use total internal reflection to propagate light along the major axis of a solid, flexible tube of glass. The maximum angle at which you can propagate light and have it bounce up and down the fiber is $\alpha = \sqrt{n_1^2 - n_2^2}$ where n_1 is the fiber refractive index and n_2 is the medium refractive index (not air, because then we would have scattering). The maximum angle we can achieve is $\alpha_{\max}$ or the *numerical aperture*. The focal ratio of angles the fiber can accept is $\frac{1}{2\alpha_{\max}}$.

What is this good for? We can use fibers as a geometry organizer, to help break the disconnect between the size of imaging optics/field of view and the size of a spectrum. It can also "scramble" the light from an image so the spectrograph sees a uniform, featureless spot, which can help with increasing wavelength resolution.

Keck LRIS looks at a 6×8 arcminute chunk of the sky, which is large and needs a 500 mm collimator. DEIMOS has a bigger field, so it needs a 1m collimator. Fibers help us mitigate this by rerouting light and lining up all the spectra into a straight line.

There's work in progress to get a fiber spectrograph for Keck: FOBOS is a proposed instrument being led by Kevin Bundy. A fiber spectrograph makes images of the fiber, so if you can collect flux and add it up and put it on a smaller fiber, you can get physically smaller images. We can then co-add fiber images, although this can often be read-noise limited.

All light entering at angles $< \alpha_{\max}$ propagates in the fiber, and in ideal fibers it preserves that angle and preserves focal ratio. In real fibers, this may not hold. Fibers scramble well in the angular direction (an input spot anywhere on the fiber becomes a ring) but not the radial direction. We can have step-index or graded-index fibers, which describes whether or not there is an abrupt transition in index of refraction between core and cladding.

A mode propagating in a cylinder is an orthogonal solution to the Helmholtz wave equations (Bessel functions, standing waves, etc). How many modes propagate is set by the fiber size relative to the wavelength. On a small fiber where you can get a lot of light onto it, you'll only get one mode and you'll get a Gaussian-like circular spread. Larger fibers can create more complicated patterns. If you want to couple light only to the first mode, you need to input a smooth, small, Gaussian beam, so these are usually AO-fed. But this is hard to do, because you need a very stable system.

We can see "modal noise", or fluctuations between the amount of light propagating in different modes, which is observed as fluctuations in the light distribution at the fiber output. This can be reduced just by moving the fibers around mechanically.

Something about attenuation and OH, I didn't quite get it.For reference, we can look at the DESI throughput.

Fibers feature focal ratio degradation: the cone that comes out is larger than the one that goes in. Collimators have to be faster than their telescopes to account for this. This effect largely comes about due to defects or irregularities at the cladding interface, causing extra scattering or reflection.

If a star moves around in a slit, the light profile in the fiber changes, and therefore so does the light profile of the image in the spectrum. It gives us errors in wavelength calibration, velocity, and sky subtraction. We can scramble the fiber on purpose to compensate for the star's motion to first order.

You can scramble both the near field (image at the output of the fiber) and the far field (image of the pupil). It's hard to make double scramblers without throughput loss; KPF is predicting 92%. Modern precision RV spectrographs are so effective because they're so efficient at scrambling.

7.3 Pupil stops, echelle gratings

Lyot stops block out light that does not come from the primary mirror/entrance pupil. Coronagraphs do this but more aggressively: you can mask around the edges and on the central star. Coronagraphs and Lyot stops work in concert to get rid of stray light or contamination. The Lyot stops in these cases can be aggressive.

Echelle gratings help us solve the problem of order overlap in spectrographs by adding another grating. We put semi-overlapping spectra through a cross-dispersing grating that causes dispersion in a perpendicular direction, and this separates out the spectra.

Instrumentation for Astronomy **Winter 2023**

Lecture 8: Photometry

Lecturer: Connie Rockosi *21 February* *Aditya Sengupta*

Although we have photometry references from Landolt fields, it's almost impossible to exactly match the filter and optics used to take those calibration measurements. We try and compensate for this using a photometric transform:

$$B = C_1 B_{\text{measured}} + C_2(B_{\text{measured}} - V_{\text{measured}}) + C_3 \text{airmass} B_{\text{measured}} + \dots \tag{8.1}$$

Magnitude measurements depend heavily on S_ν, the sensitivity of the system as a function of frequency. This depends on several factors from the telescope, like the optics and quantum efficiency. If S_ν is broad enough, very different spectral energy distributions can give the same observed flux.

As a zero reference, Vega is defined to be magnitude 0 in all filters. But Vega gives us different fluxes in different filters, so the zero point is different in all filters. Further, data's since changed so Vega in, e.g., the V band is now +0.03. That said, this method still works fine. We also use AB magnitudes, which are defined by a known calibration source and which have a known constant denominator, 48.6.

Most magnitudes are the same in the V band in any photometric system, but it starts diverging away from this. The AB magnitudes of Vega are J = 0.89, H = 1.37, Ks = 1.84.

It's impossible to avoid transforming between filter systems. What really matters is your total S_ν relative to the S_ν of the system that set up the standards. Usually this is just a linear transformation.

It's important to do this accurately to realize science goals over large samples or whole populations.

How surveys solve this calibration problem: SDSS synchronized their charge movement with the movement of stars along the sky. SDSS also did internal calibration based on repeat surveys.

Another survey is PanSTARRS. They ended up going to about the same depth as SDSS or a bit fainter.

Instrumentation for Astronomy **Winter 2023**

Lecture 9: Survey instruments, gratings

Lecturer: Connie Rockosi *2 March* *Aditya Sengupta*

9.1 Survey instruments

Surveys are important! There's many science examples of this: one is the distribution of galaxies in color/color or color/magnitude. It helps to see this with hundreds of thousands of datapoints. Another is stellar streams in the Milky Way. We can look for overdensities of stars, and using massive amounts of data helps to decouple true overdensities from systematics.

The Zwicky Transient Factory in southern California is a survey telescope with a massive field of view of 47 square degrees. It delivers a 2-arcsecond FWHM (not great) and has a pixel scale of about an arcsecond per pixel (still Nyquist sampled because of the bad seeing). In a single visit to any part of the sky, it gets down to 20th magnitude, so it can get very well-sampled light curves very frequently. Because it's covering the same patch of sky so much, it can pass information onto larger telescopes with better seeing once it's got a transient detection. It's the only survey telescope that doesn't have a special multi-lens corrector system.

ZTF is a Schmidt telescope, which uses the trick of putting the entrance aperture at the center of curvature of a spherical primary mirror. This means incoming light at all angles see the same spherical mirror, so there are no off-axis aberrations because each ray of light "sees" a spherical primary mirror as if it were on-axis. This allows for a very wide field of view with good image quality.

A wide field on a big telescope is difficult. 8-meter Schmidt correctors don't work. If you specially design optics from the ground up for wide survey instruments, you can get really wide fields of view at reasonable resolutions.

9.2 Reflection gratings

We've seen that there's an optimal condition at which the grating efficiency is maximized, in terms of the blaze angle and some other angles (the Littrow condition). This varies with the polarization of light, and you also get a performance hit for moving away from the optimal blaze angle.

Instrumentation for Astronomy **Winter 2023**

Lecture 10: Radio astronomy

Lecturer: Connie Rockosi *7 March* *Aditya Sengupta*

The National Radio Astronomy Observatory has a lot of great material on radio and sub-mm astronomy and radio interferometry.

ALMA works from a few 100 microns up to 1cm. Radio observatories need to be in EM dark zones because the wavelengths/frequencies at which Wi-Fi (for example) works are close to lines we care about.

We do radio astronomy with large arrays of antennas that we put together to make an interferometer. We can get amazingly good resolution even at long wavelengths because we make these arrays so big.

ALMA is the Atacama Large Millimeter/submillimeter Array. It's an international partnership, represented in the US by NRAO. Between ALMA and the VLA we have excellent resolution on about three or four orders of magnitude in wavelength.

What kind of physics is accessible to sub-mm astronomy? We can look at electromagnetic emission from very cool objects, because they peak at longer wavelengths. Things like bremsstrahlung (free-free) emission from optically thin, ionized gas is visible at these wavelengths. Synchrotron radiation (emission from electrons spiralling around magnetic fields) is also visible at long wavelengths. Here, redshift becomes important: things that are optical features at high redshift can become radio features to us.

A blackbody curve is in specific intensity:

$$I_{\nu,\text{Planck}} = \left(\frac{2h\nu^3}{c^2}\right)\frac{1}{\exp(h\nu/kT)-1}\text{erg/cm}^2\text{/s/Hz/sr} \tag{10.1}$$

When we hear about the brightness temperature, it means the blackbody temperature under the assumption that $h\nu \ll kT$, so $I_{\nu,\text{Planck}} \sim \frac{2\nu^2 k}{c^2}T_B$. Radio astronomers sometimes talk about brightness temperatures for things that aren't blackbodies; think about it as specific intensity in Kelvin. At 30 GHz and $T = 300$K (here), $h\nu = 2\times10^{-24}$ and $kT = 4\times10^{-21}$, so this is reasonable.

The uncertainty in a temperature measurement scales with the system temperature:

$$\Delta T_{RMS} = \frac{T_{sys}}{\sqrt{\Delta\nu\tau}} \tag{10.2}$$

where $\Delta\nu$ is the frequency band and τ is the integration time.

How does detecting signals in the radio work? In a heterodyne receiver, we mix the incoming signal with a mono-frequency signal close to the incoming signal, and get out a lower "beat" frequency with all the information from the incoming science signal intact.

The mixer has $I \sim V^2$, so if you get a science frequency ω_s and have a local oscillator signal ω_{LO}, we have

$$I_{out} \sim (V_s \sin\omega_s t + V_{LO}\sin\omega_{LO}t)^2 \tag{10.3}$$

which a lot of trigonometry can show us is equal to

$$I_{out} \sim V_s V_{LO} \cos(\omega_s - \omega_{LO}) - \underbrace{V_s V_{LO} \cos(\omega_s + \omega_{LO})}_{\text{low pass filter this out}}. \tag{10.4}$$

A signal with frequency $\omega_s - \omega_{LO}$ is easier to analyze.

For 1 arcsecond resolution at 21 cm, we need $D = 40$km, hence the VLA. We can't build something that's physically that big without interferometry. This is made easier by the fact that the atmosphere is basically transparent at radio wavelengths, so it's easy to combine results from different telescopes without any AO.

An interferometer works by having two telescopes look at the same source, and because they're separated on the ground there's a path length difference proportional to that separation. We look for the correlation between those two signals to get one Fourier component of the astrophysical signal, and do that with many pairs of arrays to get many Fourier components.

If your baseline is long, you can get higher resolution. There's also a largest angular scale you're sensitive to; anything that's too big is basically invisible to the interferometer, because your fringes never see the edge of that source. Since we only sparsely sample the Fourier transform, we often end up with artefacts in the image space. Taking this inverse Fourier transform and denoising appropriately is *aperture synthesis*.