Chapter 4: Optical and Infrared Imaging and Astrometry

4.1 Imagers

Imagers capture the two-dimensional pattern of light at the telescope focal plane. They consist of a detector array along with the necessary optics, electronics, and cryogenic apparatus to put the light onto the array at an appropriate angular scale and wavelength range, to collect the resulting signal, and to hold the detector at an optimum temperature while the signal is being collected. Imaging is basic to a variety of investigations, but is also the foundation for the use of other instrument types that need to have their target sources located accurately. In this chapter we discuss the basic design requirements for imagers in the optical and the infrared and guidelines for obtaining good data and reducing it well. We finish with a section on astrometry, a particularly demanding and specialized use of images.

4.2. Optical imager design

A simple optical imager consists of a CCD in a liquid nitrogen dewar with a window through which the telescope focuses light onto the CCD. Broad spectral bands are isolated with filters, mounted in a wheel or slide to allow different ones to be placed conveniently over the window. Although this imager is conceptually simple, good performance requires attention to detail. For example, if the filters are too close to the CCD and telescope focus, any...
imperfections or dust on them will produce large-amplitude artifacts in the image.

Other components can improve the quality of the data. A shutter can provide accurate and uniform exposures. For instance, a curtain can initiate the exposure by opening across the CCD, and a second curtain can close off light

Figure 4.2. Wynne-type field correcting optics.

Figure 4.3. Layout of a near-infrared camera. After entering through the dewar window, the light is focused by the field lens plus collimator on to the pupil where the cold stop is placed. The remaining optics relay the focal plane to the detector array.
in the same direction to end the exposure, providing the same exposure time across the entire CCD. More ambitious instruments use optical systems to provide good images over a large field that can be filled by a mosaic of detectors. There are a broad variety of possible corrector designs. Figure 4.2 shows a generic example, based on the concept proposed by Wynne (Section 2.4.1) that is used widely. The first three lenses work together to correct the telescope aberrations over the field; lens 4 flattens the field to match the detector. An atmospheric dispersion corrector (ADC) counters the chromatic effects due to atmospheric refraction. Without the ADC, images taken at small elevation angles will look like tiny spectra, with the blue end pointing to the zenith because the atmospheric refraction bends blue light more than red light. In this design, the ADC is moved as the telescope elevation changes to provide a good correction.

4.3. Infrared imagers

Figure 4.3 shows the optical layout of a near infrared camera (after Hodapp et al. 2003). To avoid being flooded with thermal background from the telescope and other surroundings, the entire camera is cooled in a vacuum enclosure. A number of fold mirrors are placed in the optical train to make it fit into as small a cryostat as possible. In addition, to minimize the view of the telescope, the optical train forms a pupil around which is placed a tight, cold stop. The “camera” optics behind the pupil reimage the beam onto the array at the desired pixel scale. These design considerations result in an instrument configuration changed substantially from that for the CCD camera, although the two instruments take very similar-appearing data.

There are some serendipitous benefits from this design. For example, the filters can be placed at a pupil. This placement is optically equivalent to placing the filter over the primary mirror, so small flaws in a filter result in uniform loss of light but do not introduce artifacts into the images. In contrast, neither the simple optical CCD camera nor the Wynne corrector version provide a pupil where filters can be placed. Therefore, the accuracy of their photometry depends on a high degree of filter uniformity.

However, all those optics impose their own issues:
1. \( \cos^2 \theta \) effects refer to a general drop in signal with increasing off-axis angle (\( \theta \)). In general, optical systems are most efficient on axis. For example, anti-reflection coatings are generally designed for normal incidence and their effectiveness is reduced for off-axis rays. The efficiency of a detector array is likely to fall off with increasing angles of incidence. Vignetting may reduce the response at the edges of the field.
2. Ghost images - light reflected from refractive optics can get back to the array and provide point-like or extended images, depending on the geometry. Even simple optical imagers can produce ghosts due to reflections from the dewar window and/or filter.

4.4 Nyquist Sampling

A basic requirement in imager design is to set the equivalent angular size of the array pixels to extract the maximum possible information. Pixels that project to angles that are too small will under-utilize the detector array and reduce the size of the field of view unnecessarily. Pixels that project to angles that are too large will lose information on small angular scales. From Section 2.3.1, we know how to control the projected pixel sizes; now we will determine what size we should aim for. Fortunately, there is a rule that says how small is small enough that essentially no details are lost in an image.

We can determine the required pixel size by determining the MTF of a detector array and seeing under what conditions the spatial frequency spectrum from the telescope is preserved. We only have to compute the MTF of a pixel, since the full suite of pixels will have the same impact on the result. The function

\[ \prod (x) = 1 \text{ for } |x| \leq \frac{1}{2} \]

\[ = 0 \text{ otherwise} \]

is the response function for a pixel of width 1. We need to modify it for a pixel of width \( w \):

\[ \prod \left( \frac{x}{w} \right) = 1 \text{ for } \left| \frac{x}{w} \right| \leq \frac{1}{2} \]

Using equations 2.17 and 2.24, the corresponding Fourier transform is

\[ F(w) = \frac{1}{|1/w|} \frac{\sin(\pi f_s/(1/w))}{\pi (1/(1/w))} = |w| \frac{\sin(\pi w f_s)}{\pi w f_s} \]

where \( f_s \) is the spatial frequency.

The MTF is \( \sqrt{F F^*} \), normalized to 1 at \( f_s=0 \). Since

\[ \frac{\sin(x)}{x} \rightarrow 1 \quad \text{as } x \rightarrow 0 \]
it follows that

\[ MTF = \left| \frac{\sin(\pi f_s)}{\pi f_s} \right| \]  \hspace{1cm} (4.1)

Some samples are shown in Figure 4.4. We can identify \( f_s = 1 \) with the natural
cutoff frequency of a telescope, \( D/\lambda \). Since \( f_s = 1/P_s \) where \( P_s \) is the spatial
period, the MTF at \( D/\lambda \) for sampling with two pixels across the spatial period is

\[ MTF = \left| \frac{\sin(\pi/2)}{\pi/2} \right| = 0.64 \]

That is, the attenuation of high spatial frequencies is modest. With one pixel
per period the loss at high spatial frequencies is already severe. That is, we
have shown that two pixels across the image diameter (FWHM = \( \lambda/D \)) is a
good goal for the pixel size that does not lose any spatial information.

This result is often described as Nyquist sampling, after the Nyquist or
Sampling Theorem. This theorem states that a bandwidth-limited signal with
maximum frequency \( f \) and period \( P = 1/f \) can be completely reconstructed

Figure 4.4. MTFs for sampling at 0.5, 1, and 2 pixels per spatial period. The
spatial frequencies are in units of \( D/\lambda \).
from time samples at a time interval of P/2 (strictly at slightly smaller than P/2). The situation with finite pixels is analogous but not completely identical with the assumptions in proving this mathematical result (e.g., the pixels have significant size compared with their spacing whereas the sampling theorem applies to instantaneous samples). Furthermore, a number of people derived the result independently of Nyquist, but the “Nyquist” terminology applied to imaging is firmly entrenched.

Images where the pixels are larger than half the image FWHM are described as being undersampled, whereas those with pixels smaller than this limit are oversampled. Besides the irretrievable loss of information at high spatial frequencies, there is a related issue with undersampling, called aliasing. With undersampled pixels, the source distribution that fits a detector array output image is not uniquely determined. Consider the case with ½ pixel per FWHM of the image in Figure 4.6. The null at D/2λ occurs because a signal containing only this spatial frequency has one full period across the pixel width, so it cancels its own signal. An example of a signal contributing to the secondary peak at higher frequencies would be one 1.5 times higher than the null frequency; it would have 1.5 periods across the pixel and so it would only partially cancel itself. The high spatial frequency signals can get confused with source structures at lower spatial frequency and there may be no unique distribution of sources that can be associated with an image. It may be easier to visualize this problem in image space; if the pixels seriously undersample the input image, then a number of arrangements of diffraction-limited images distributed over pixels could produce identical images. Fortunately, for mild degrees of undersampling, if we take multiple exposures offset in pointing by a fraction of a pixel size we can mitigate aliasing. However, Figure 4.4 shows that this strategy does not recover all the information; once the pixels are as large as λ/D, the highest spatial frequencies are lost completely from the images.

A problematic case occurs with seeing-limited images. Our derivation of the required sampling assumed that there is a unique spatial frequency cutoff associated with the telescope aperture. Seeing-limited images behave in a more complex manner. Nonetheless, a useful rule of thumb is that the images should have at least two pixels across their FWHMs.

There are situations where oversampling is beneficial. One reason is that real arrays fall slightly short of the ideal (see Section 4.5 below) and some of their flaws can be overcome by finer sampling. Another case is groundbased imaging where one might try to extract information at spatial scales finer than the simple “seeing limit.”
4.5. Imager Data Reduction

4.5.1 Benefits and Issues

Imagers have huge advantages over single detector instruments for nearly all astronomical observations. They permit very accurate position determination and enable astrometry, as discussed in Section 4.7. For photometry they:

- Allow centering on the source and setting other parameters of extraction of photometry after the fact
- Enable use of small apertures in crowded fields and to reject backgrounds for improved sensitivity
- Allow differential photometry relative to other sources in the field for accurate measurements under non-optimum conditions
- Enable removal of foreground stars that might interfere with the signal from an extended source
- Enable use of custom extraction apertures with shapes optimized to the shape of the source
- Allow construction of multiple-color images for comparison of the behavior of a source in all the colors precisely

However, to gain these advantages, there are number of steps that are required:

- Various issues with detector arrays need to be anticipated in data-taking
- The data reduction strategy needs to allow for elimination of problems such as transient signals
- Calibration must take into account the differing properties of the detectors in the array

Some of the optical phenomena that affect array data are:

- Interpixel gaps and intrapixel response: Since array pixels are discrete, the sensitivity may have minima between pixels (Figure 4.5, image from JWST HgCdTe array). These "gaps" can have big effects if the pixels do not sample the PSF well. It is also possible that the response varies over

Figure 4.5. Interpixel response gaps, from John Hutchings
Imagers

the face of a pixel. Figure 4.6 shows the dependence of the signal on the centering of a source on a NICMOS Camera 3 pixel - for this example, 1 pixel $\sim 1.5 \lambda/D$, where $D$ is the telescope aperture. These effects will not be detected through normal flat fielding image processing.

- Crosstalk: Small amounts of charge can diffuse from one pixel in an array to the next, blurring the image. This issue is much more prominent if the pixels undersample the intrinsic image.
- Fringing/channel spectra: Arrays are based on thin parallel plate components. When the absorption in the detectors is low, interference within the material causes fringing. Figure 4.7 shows fringing in a GMOS-S CCD at $\sim 0.95 \mu m$. The nature of the fringing is a sensitive function of the spectral content of the illumination. Therefore, fringing can best be corrected by generating response (or flat field) images with the identical illumination conditions as those for the science data.

There are also a number of electrical issues (see Figer 2002 for illustrative examples):

![Figure 4.6. Signal vs. centering, (from Hook and Fruchter 2000)](image1)

![Figure 4.7. Fringing in a CCD at 0.95 microns](image2)
Imagers

- **Hot and dead pixels.** Some pixels have anomalously high dark current (termed “hot”) and some do not respond at all (“dead”). In either case, they return no useful data; the missing information can be supplied by interpolation, or better by using an observing strategy that moves the image on the array so the effects of these pixels can be filled in with valid data.

- **Cosmic ray hits, other transients.** When a cosmic ray passes through a pixel, it generates a cloud of free charge carriers that produce a transient signal. Cosmic ray hits can be identified in two basic ways. The first is to use some unique aspect of such events. For example, if one is sampling up the ramp on an integrating detector output, a cosmic ray will cause a discontinuous jump that can be identified and excised from the data. If an image is significantly oversampled, it may also be possible to identify cosmic ray hits as producing an “image” that is impossibly small in size. The second approach is to obtain a large number of exposures and reject the cosmic rays by comparing them. For example, one can take the median for each position on the sky, rather than the mean. That way, an outlier or two (due to a cosmic ray hit or other stochastic event) will automatically be excluded from the data. Another approach is called sigma clipping. One calculates the mean and standard deviation for the values for repeated measurements of a position, then rejects those falling more than some number of standard deviations from the mean and recalculates. Both approaches are examples of robust statistics; that is, they are approaches in which the results are not strongly perturbed by a few anomalous readings, but which also impose no bias on the final value.

- **Nonlinearity and soft saturation:** Modest nonlinearity can be corrected with a suitable fit to the detector behavior as a function of signal strength. Soft saturation occurs when the output continues to grow with increasing signal, but the nonlinearity is strong enough to make it difficult to recover accurate measurements.

- **Latent images:** many electronic arrays retain an image at the 0.1 - 1% level on the next readout after a bright source has been observed. The images are from charge trapped at surface or interface layers in the detector. These images usually decay over about 10 minutes. If a short exposure on a bright source is followed by a long one (~ 10 minutes) on a faint one, then the total latent charge collected can be much larger than 1% since many readouts similar to that on the bright source are combined. In addition, much longer decay times can result if the array has been saturated by an extremely strong signal. In such cases, it may be necessary to generate an ad hoc response image from the images with the impressed latent and use it to make a correction.
• MUX glow: the array readout transistors are sources of light through electroluminescence. This can be picked up by the array, contributing to a lack of flatness in the images and also contributing noise. Standard reduction procedures correct for the MUX glow signals but the noise cannot be removed.
• Electrical crosstalk/ghosting: Various effects can produce secondary images. A common example is inadequate drive power on the output amplifiers of an array with multiple outputs. The result is that when a bright source requires a lot of drive power to push its signal out of the array, the supply to the other amplifiers can be affected.
• Pedestal effects: readouts sometimes have electrical offsets that appear as structure in images. Typically, these effects can be removed by standard methods with no degradation in the final images.

4.5.2 Taking Good Data
Addressing these issues in the final reduced images requires care in taking the data. The first rule of array imaging is repetition
• Multiple images allow systematic identification of outlier signals due to cosmic rays and other transients
• They also allow replacing areas compromised by cosmic rays, latent images, hot or dead pixels, ghosting, etc. with real data.
• By dithering the pointing on the sky between exposures (moving the telescope slightly so the images fall on different parts of the array), the sky signal itself can be used to flatten the image (as discussed below). If the sky dominates the signal, then fringing effects are removed to first order, along with many other potential contributors to non-flatness. These benefits require at least three dither positions and more are better.
• Properly sampled images are another form of repetition - more than one pixel contributes to the signal. Accurate photometry benefits from spreading the light over multiple pixels (which can also be done with dithering).

The second rule of imaging photometry is don't change anything. Artifacts like MUX glow, pedestals, and many others will disappear from the reduced data virtually completely if all the input data - science and calibration frames – are taken in identical ways (for example, constant exposure times and readout cadences, plus constant temperatures, and backgrounds).

4.5.3 Calibration
We now address what is sometimes called the inverse problem – working with the observations to remove artifacts and generate a high fidelity version
of the photon input that yielded the measurements. A more mundane term is to calibrate the data, which we assume is in the form of a two-dimensional field of digital units (DU, or DN for digital numbers) representing the outputs of all of the array pixels. In discussing this process, we assume that the data consist of repeated exposures of the field, moving the source on the array between exposures (dithering). This is the most common way to take imaging data (because it works well). We also assume that there is an offset frame (sometimes called bias frame) with a very short exposure and no signals, plus a dark current frame with a long exposure (ideally the same as for the sky frames) and no signals. We will describe how to generate a response frame from the data.

The first step in reducing the data is to pre-process the images to get rid of artifacts. Cosmic ray hits should be removed at this state, if they are being identified by discontinuities in the integration ramp; however, there is another step late in the processing that can also remove them. Latent images may require generation of a special calibration frame using data including the latent; this frame is used analogously to the response frames described below. One might also have to identify electrical and optical ghost images and other such effects and fix them by hand or with custom routines.

Modern infrared arrays often include non-active reference pixels that are electrically identical to those that detect photons. They can be used to correct the images for slow drifts in the electronics and other such effects, although how useful they are depends on the array and the use it is being put to. Their use may be hard-wired into the reduction pipeline, or one may need to experiment or seek advice on whether they improve the quality of the data. For CCDs, the same benefit is obtained by overscanning the array, that is continuing the readout clocking beyond the area of the real pixels. The patterns of noise and electrical artifacts in the overscan data are also impressed on the data from the real pixels. As appropriate, the reference pixel or overscan information can be used to correct all the images.

Assuming these steps have been successful, to produce good images the data must still be processed to take into account and correct for the differing properties of the detectors in the array, that is pixel-to-pixel variations in: 1.) amplifier offset; 2.) dark current; and 3.) response. The first two calibration frames (offset and dark current) should already be available. A response frame (sometimes called a flat field) can be generated as the median average of the dithered frames on sky – as a result of the dithering, sources will disappear because they do not appear at the same place on any two frames.

Image data reduction then consists of:
1.) Subtract the offset frame from the data, dark, and response frames to obtain data’, dark’, and response’
2.) Scale dark’ to exposure time of data and response and subtract from them to get data’” and response’”
3.) Divide data’” by response’”

That is:

\[
\begin{align*}
\text{data’} &= \text{data} - \text{offset} \\
\text{dark’} &= \text{dark} - \text{offset} \\
\text{response’} &= \text{response} - \text{offset} \\
\text{data’”} &= \text{data’} - \alpha \text{dark’} \\
\text{response’”} &= \text{response’} - \beta \text{dark’} \\
\text{final frame} &= \text{data’”/response’”}
\end{align*}
\]

Here \( \alpha \) and \( \beta \) are the scaling factors to adjust the dark frame to be equivalent to the dark that would be obtained with the same exposure as the frame being adjusted. If the data frame has a uniform exposure, then the product of these reduction steps will be a uniform image at a level corresponding to the ratio of the exposure on the data frame to the exposure on the response frame (exposure = level of illumination multiplied by the exposure time). Sources will appear on top of this uniform background.

There will be one such reduced frame for each position of the telescope (at least for each dither position and if multiple exposures were taken at a given position, the telescope may have drifted in pointing so each exposure may be usefully treated as a separate reduced frame). These frames need to be shifted to a common pointing and combined. In general, the images are unlikely to have been taken with exactly an integral number of pixel widths difference in pointing. The simplest solution is just to assign the pixel values to the nearest point on the master grid of the image. An improvement is to use bilinear interpolation among the four nearest positions to assign the values appropriately for the surrounding pixel positions. Various other interpolation functions such as sinc, spline, or polynomial can produce better performance under specific conditions. Finally, after shifting all the frames, another median average will eliminate bad pixels and cosmic rays, while gaining signal to noise on the source image. This final step automatically fills in the positions of bad pixels with valid data from other pointings of the telescope, so the impact of the bad pixels is only a local reduction of the total integration time in the final image.

We have gone through one possible calibration sequence. In general, the image reduction software will include standard or recommended procedures.
to generate the necessary calibration frames from your data, and to shift and add all your science frames into one high-quality image. In addition, for simplicity, we have specified the minimum number of calibration frame types. Additional types that may be useful include: 1.) dome flats obtained with the telescope pointing at an illuminated screen; they provide an alternative response frame set with high signal to noise but with the disadvantages that the spectral character and the illumination do not match the sky well; 2.) twilight flats obtained in the evening or morning when the sky is bright enough to provide high signal to noise in short exposures; they overcome both of the issues with dome flats to a large extent; 3.) shutter shading exposures, flats obtained with a range of integration times, especially short ones, so the effect of a shutter on the exposure over the field can be determined; and 4.) orientation exposures obtained with the telescope drive off, so the precise orientation of the field on the sky can be determined from the direction of the star trails.

For well-sampled images, the approaches described above work well. For undersampled images, or where one is being exceptionally careful to preserve the image characteristics, a different class of combination works better, exemplified by the “drizzle” algorithm (Fruchter & Hook 2002). A fine grid is imagined to be projected onto the sky, representing an ideal array of pixels for the image. The data for the observed pixels are projected onto this grid as in Figure 4.8. This projection not only reflects the true pixel sizes, but also includes any distortions from the optics, rotations of the instrument with regard to the ideal grid, and any other issues that result in the real data not projecting perfectly onto the ideal array. The signals are divided according to the overlapping areas of the actual pixels and the ideal ones represented by the fine grid. For illustration in Figure 4.8, the observed pixel area was set to
exactly four times the area of a “pixel” in the ideal grid, but it is distorted in shape and rotated. The algorithm would assign 25% of the signal received by this imager pixel to the ideal one 2 down and 3 from the left (since it overlaps entirely), and would distribute the remaining 75% of the signal to the other ideal pixels in proportion to the overlapping areas with the imager pixel. Thus, if one of the ideal pixels overlaps with 15% of the projected real pixel, 15% of the signal from that real pixel is assigned to the ideal one. Once this process has been completed over the entire image, the signals over the ideal grid of pixels represent the source field as it might have been imaged starting with the ideal array. A similar approach can be used to remove optical distortions in co-adding multiple images (Gordon et al. 2005).

4.5.4 How to carry the measurements around

Now that the images are reduced, they need to be recorded. In the late 1970's astronomers developed the Flexible Image Transport System, FITS, as an archive and interchange format for astronomical data files. In the past decade FITS has also become the standard format for on-line data that can be directly read and written by data analysis software. FITS is much more than just an image format (such as JPG or GIF) and is primarily designed to store scientific data sets consisting of multidimensional arrays and 2-dimensional tables containing rows and columns of data.

A FITS file consists of one or more Header + Data Units(HDUs), where the first HDU is called the "Primary Array". The primary array contains an N-dimensional array of pixels. This array can be a 1-D spectrum, a 2-D image or a 3-D data cube. Any number of additional HDUs, called "extensions", may follow the primary array. They are followed by an optional "Data Unit".

Every "Header Unit" is formatted in ASCII and consists of any number of 80-character records that have the general form:

KEYNAME = value / comment string

The keyword names may be up to 8 characters long and can only contain uppercase letters, the digits 0-9, the hyphen, and the underscore character. The value of the keyword may be an integer, a floating point number, a character string, or a Boolean value (the letter T or F). There are many rules governing the exact format of keyword records so it usually best to rely on standard interface software like CFITSIO, IRAF or the IDL astro library to construct or
parse the keyword records rather than directly reading or writing the raw FITS file.

4.6. Astrometry - Coordinates

Well-reduced images invite us to think of the positions of astronomical objects, a direction that leads us to the topic of astrometry – the procedures used to set up systems of positions and to measure the locations of individual objects accurately within these systems.

4.6.1 Coordinate Systems

An astrometric coordinate system can be envisioned as a grid projected up into the sky upon which the positions of celestial objects are measured. Such a grid has a fundamental great circle and a secondary one (a great circle is the intersection of a plane running through the center of a sphere with the surface of that sphere). There are four such grids in common use:

![Figure 4.9. Horizon Coordinates](image-url)
1. **Horizon**: the fundamental circle runs around the horizon and the secondary one runs from it over the zenith. This system is the basis of the altitude-azimuth, alt-az, coordinates used to point most large telescopes. As shown in Figure 4.9, the great circle passing through the zenith and north and south celestial poles defines the zero point of azimuth where it intersects the horizon circle to the north. Any object on the celestial sphere lies on a great circle perpendicular to the horizon circle, and the azimuth (A) for this object is the angular distance measured eastwards from the zero point to the first intersection of its great circle with the horizon one. The altitude (a) of the object is measured along this circle from the horizon circle, + if it is above the horizon and – if it is below.

2. **Equatorial**: Although the horizon system is convenient for telescopes, it has the disadvantage that the coordinates of any object depend on the place and time of the observation. Another system is needed in which the position of an object remains at fixed coordinates. Equatorial coordinates fulfill this role (Figure 4.10); the fundamental circle is the celestial equator and declination is measured along a secondary great circle that runs through the object and is perpendicular to the celestial equator, + for north and – for south. The zero point is the position of the Vernal Equinox, and right ascension is measured from there eastward to where the declination circle for the object intersects the celestial equator. The Vernal Equinox is defined as the point where the celestial equator and ecliptic (the apparent path of the sun across the sky) intersect in March (i.e., the placement of the sun the moment in March when it is directly overhead as viewed from the equator). Thus, the zero is roughly at midnight (within the vagaries of civil time) at the Autumnal Equinox. The

![Figure 4.10. Equatorial Coordinates](image-url)
units of right ascension ($\alpha$) are hours, minutes, and seconds while declination ($\delta$) is measured in degrees, and minutes and seconds of arc. The celestial meridian, or local meridian, is the great circle along which lie the north and south celestial poles and the zenith (the point directly overhead). The meridian of a source is the great circle along which lie the north and south celestial poles and the source in question. The hour angle of the source is the angular distance from the celestial meridian to the meridian of the source, measured to the west and in hours, minutes, and seconds. It is equivalently the time until the source transits the celestial meridian (negative hour angle) or the time since it transited (positive).

3. **Ecliptic**: This coordinate system is the natural one to use when dealing with members of the solar system. The fundamental great circle is the ecliptic – the apparent path of the sun across the sky. The zero point is the Vernal Equinox and the ecliptic longitude ($\lambda$) is measured from there eastward. The ecliptic latitude ($\beta$) is measured along a great circle perpendicular to the ecliptic and passing through the north and south ecliptic poles; positive if north of the ecliptic and negative if south.

4. **Galactic**: For problems centered on places in the Milky Way, the Galactic system is preferred. Its fundamental great circle is the plane of the Galaxy, the Galactic equator, and the zero point since 1958 is close to the Galactic Center (it was originally intended to be the Galactic Center, but we have learned since where this region is more accurately and it is about 5 arcmin away from the coordinate system zero). The pre-1958 system is designated by I and the newer one by II (Roman numerals). To place an object in Galactic coordinates, we first determine a second great circle passing through it that is perpendicular to the Galactic Equator and passes through the north and south Galactic Poles. The Galactic longitude ($l$) of the source is measured from the zero point eastward to where this circle intersects the equator. The Galactic latitude ($b$) is measured along the second great circle, + for north and – for south.

### 4.6.2 Sidereal Time

To make use of these coordinate systems, we need to synchronize watches – that is, to place the celestial objects in the sky as a function of some time system. A way to do so is to determine when an object is on the local meridian.
Our clocks are set to run (approximately) on solar time (sun time). But for astronomical observations, we need to use sidereal time (star time). Consider the Earth at position E1 in Figure 4.11. The star shown is on the meridian at midnight by the clock. But three months later, when the Earth reaches position E2, the same star is on the meridian at 6 p.m. by the clock. We define one rotation of Earth as one sidereal day, measured as the time between two successive meridian passages of the same star. Because of the Earth's orbital motion, this is a little shorter than a solar day. (In one year, the Earth rotates 365 times relative to the Sun, but 366 times relative to the stars. So the sidereal day is about 4 minutes shorter than the solar day.)

The local sidereal time (LST) is the sidereal time at a particular location. It is zero hours when the Vernal Equinox is on the observer’s local meridian, and by definition of the hour angle, the LST is thus the hour angle of the Vernal Equinox – that is, if the Vernal Equinox is on the local meridian, in two hours it will be two hours (30 degrees on the celestial equator) west of the meridian and the hour angle will be +2 hours. By the definition of right ascension, the LST is also the right ascension of any source that is on the local meridian. Equivalently, the hour angle of a source is the LST minus its right ascension.

4.6.3 Coordinate Transformations
Although each of the coordinate systems has its use, they do pose the problem of transforming from one to another. Here are the formulae you are most likely to need for that purpose. First, to transform from azimuth, $A$, and
altitude, \( a \), to hour angle, \( h \), and declination, \( \delta \), for an observer at latitude on the earth of \( \phi \):

\[
\begin{align*}
\cos \delta \sin h &= \cos a \sin A \\
\sin \delta &= \sin \phi \sin a - \cos \phi \cos a \cos A \\
\cos \delta \cos h &= \cos \phi \sin a + \sin \phi \cos a \cos A
\end{align*}
\]

The inverse goes from hour angle and declination to azimuth and altitude:

\[
\begin{align*}
\cos a \sin A &= -\cos \delta \sin h \\
\sin a &= \sin \phi \sin \delta + \cos \phi \cos \delta \cos h \\
\cos a \cos A &= \cos \phi \sin \delta - \sin \phi \cos \delta \cos h
\end{align*}
\]

From equatorial to ecliptic coordinates, where the obliquity (inclination of the equator of the earth against the ecliptic) is \( \varepsilon=23^\circ26'21.448'' \), the transformation is

\[
\begin{align*}
\cos \beta \cos \lambda &= \cos \delta \cos \alpha \\
\cos \beta \sin \lambda &= \cos \delta \sin \alpha \cos \varepsilon + \sin \delta \sin \varepsilon \\
\sin \beta &= -\cos \delta \sin \alpha \sin \varepsilon + \sin \delta \cos \varepsilon
\end{align*}
\]

where \( \lambda \) and \( \beta \) are the ecliptic longitude and latitude, respectively. The inverse transformation from ecliptic to equatorial is:

\[
\begin{align*}
\cos \delta \cos \alpha &= \cos \beta \cos \lambda \\
\cos \delta \sin \alpha &= \cos \beta \sin \lambda \cos \varepsilon + \sin \delta \sin \varepsilon \\
\sin \delta &= \cos \beta \sin \lambda \sin \varepsilon + \sin \beta \cos \varepsilon
\end{align*}
\]

The previous two sets of transformations are relatively simple mathematically because all the systems are centered on the earth. The conversion to Galactic coordinates does not have this attribute and is more complex. There are a number of web-based coordinate transformation calculators that can be used, e.g.,

http://nedwww.ipac.caltech.edu/forms/calculator.html or
http://heasarc.gsfc.nasa.gov/cgi-bin/Tools/convcoord/convcoord.pl

or one can find details in Lang (2006) or Kattunen et al. (2007).
4.6.4 Defining coordinate systems

The most accurate celestial positions are obtained through very long baseline interferometry (VLBI) in the radio, accurate to a milliarcsec or better. Therefore, in 1997 the IAU adopted the International Coordinate Reference System (ICRS), based on VLBI coordinates for 212 radio sources. Because these objects are extragalactic (and indeed very distant) they should have no significant proper motions and the definition should remain in place indefinitely. We will discuss VLBI position determination in Chapter 9. The ICRS is transferred in the optical to 118218 stars, all with accurate measurements of positions and proper motions based on the Hipparcos satellite data.

4.6.5 World Coordinate System

As astronomy becomes more and more panchromatic, it has become a necessity to have an efficient method to place an image of a field accurately on the sky and in the appropriate equatorial coordinates, so it can be matched with identifications at other wavelengths. To implement this capability, suitable information is now placed in the FITS header of many types of astronomical data. A standard format for the keywords in the FITS header is defined as the world coordinate system (WCS). A common example is to link each pixel in an astronomical image to a specific direction on the sky (such as right ascension and declination). The basic idea is that each axis of the image has a coordinate type, a reference point given by a pixel value, a coordinate value, and an increment. A rotation parameter may also exist for each axis. The FITS WCS standard defined 25 different projections that are specified by the CTYPE keyword. For a complete description of the FITS/WCS projections and definitions see Greisen and Calabretta (2002), Calabretta and Greisen (2002), and Greisen et al. (2006).

There are a number of software packages that aid the astronomer in accessing the astrometric information using the WCS of the image or to write the WCS of an image to the header. A few of the most commonly used packages are WCS tools, WCS LIB, IRAF, and packages in the astronomy IDL library.

If an adequate WCS does not exist for an image the basic steps are to

1. Read in the FITS image and its header
2. Find all the stars in the image
3. Find all stars in a reference catalog in a region of the sky where the image header says the telescope is pointing.

4. Match the reference stars to the image stars.

5. Using one of the above WCS software packages perform a fit between the matched star's pixel and reference positions. Write the resulting WCS information to the header.

Unless very accurate pointing data can be associated with the data, obtaining this information often requires conducting an automated search to match objects detected in the image with a catalog of objects on the sky. If this search is to be fast, it cannot proceed by brute force. One strategy is to sort the objects in the image in order of decreasing brightness and then to match them with a list similarly sorted of catalog objects in the same region of sky. Once a match has been achieved, it is usually necessary to correct the image data for distortions and other effects that might make the coordinates less accurate away from the specific region of the match.

4.6.6 Changes in Celestial Coordinates

Unfortunately, we are not done. The linking of the equatorial coordinate system to the celestial equator and poles means that the grid of the system shifts due to a number of motions of the earth. In addition, to use the system accurately there are additional effects to be taken into account. Fortunately, all of the items listed below are well understood and with care can be compensated sufficiently well that they do not interfere with obtaining accurate positions for any objects we wish to observe.

**Precession:** Because the earth is not perfectly spherical, the gravitational fields of the moon and sun exert a torque on it. The result is that it precesses like a spinning top, its axis describing cones with a half angle of about 23.5° centered on the north and south ecliptic poles. A precessional cycle takes about 26,000 years. Torques exerted by the other planets add an additional precession term with a period of about 41,000 years. The planets also result in a change in the obliquity (tilt) of the poles of the earth over a range of about 21.5 to 24.5°. Of course, these effects also change the direction of the celestial equator and parameters that depend on it drift. For example, the zero of the equatorial system is set by the intersection of the celestial equator and the ecliptic, which is currently drifting at about 50 arcsec per year.
Therefore, coordinate systems defined by the Vernal Equinox must be specified for a certain date. The specified year is called the Equinox (not epoch as is commonly stated). We currently usually use coordinates for Equinox J2000.0, but one will find coordinates for equinoxes of 1900, 1950, and so forth. Calculators such as

http://nedwww.ipac.caltech.edu/forms/calculator.html or http://heasarc.gsfc.nasa.gov/cgi-bin/Tools/convcoord/convcoord.pl

are convenient for converting from one equinox to another.

**Nutation:** On top of precession, the tidal forces from the sun and moon cause a number of much smaller, short-period motions imposed on the much larger circle traced by the earth’s axis due to precession. The largest is 18.6 years long, but a number of additional periodic terms make the motion relatively complex. These terms are called nutation (after Latin for “nodding”). Like precession, they can be determined and compensated accurately.

**Parallax:** Of course, one of the primary goals of astrometry is to measure parallax and determine stellar distances. For nearby stars, this effect must be accounted for in any accurate position determination. The Hipparcos satellite has measured accurate parallaxes for virtually all nearby stars that are reasonably bright.

**Proper Motion:** Nearby stars also move measurably across the sky – in about 500 cases at a rate of 1 arcsec per year or more. In these cases, the coordinates need to be updated to the current date to have an accurate position for the star. The Hipparcos satellite data have been used with earlier astrometry to provide measurements of proper motions. Where there is a long time baseline or under circumstances permitting very accurate positional measurements, they can be measured by other means also, including to much fainter levels than are reached by Hipparcos.

**Refraction:** The index of refraction of air under standard conditions is about 1.0003 and diminishes with reduced pressure. Therefore, light from outside the atmosphere that enters obliquely is bent slightly. Objects that are really 35 arcmin below the horizon will appear (in visible light) to be right on the horizon (if we could see that clearly). The refractive index is smaller in the near infrared (and at longer wavelengths), reducing this effect.

**Aberration:** Because of the finite speed of light, the apparent position of an object is displaced in proportion to the transverse velocity of the earth moving through space relative to the vector of the beam of light from the object. Since
the direction of the motion of the earth relative to the vector of the light from
the object varies over the year, much of this effect is periodic over a year. It
can have an amplitude as large as 20 arcsec. It should not be confused with
parallax, which becomes larger the closer the object. Aberration occurs with
the same amplitude for all objects in the same direction, since it depends only
on the observer’s instantaneous transverse velocity relative to the direction of
the incoming light.

4.7. Astrometric Instrumentation and Surveys

Astrometry is an important branch of science in its own right (besides
providing coordinates for us all). It is the foundation of our distance scales, a
basic way to identify members of populations of stars, has provided
fundamental evidence for the existence of a supermassive black hole in the
Galactic Center, and is a promising approach to search for planets around
other stars, just to name a few examples.

Stellar positions are measured from images of the sky. The accuracy of the
position measurement of an unresolved source (e.g., a star) can be estimated
roughly as the full width at half maximum divided by the signal to noise. This
guideline clearly breaks down at low signal to noise (if you do not detect a
source you cannot locate it at all!). It also fails at very high signal to noise; the
underlying reason is that the further one pushes the position below the
FWHM, the more demanding are the requirements on understanding the
structure of the image. Fortunately, it is not required that the image be perfect
(aberration-free), nor even that images being compared be identical, just very
well understood. Standard accuracy limits for photographic astrometry based
on combining the results of multiple observations of the same field are about
10 milli-arcsec (mas).

To reach this level of accuracy requires that the telescope be very stable. In
addition, from our discussion of issues like inter-pixel gaps and intra-pixel
response variations (and their analogs for photographic plates), it is very
desirable that the image scale be large enough (f/# of the telescope large
enough) that the image of a star is spread over many pixels. A century ago,
the best solution was long-focus refracting telescopes, but as the engineering
of reflectors improved they became more than competitive. Although
specialized astrometric telescopes are often used for large programs (e.g., the
155 cm telescope with a flat secondary of the U.S. Naval Observatory), with
care good results can also be obtained with ones of more conventional design.

In the 19th Century, visual astrometry was a central topic in astronomy. A
number of specialized telescopes and instruments were developed to allow
Imagers

accurate measurements, such as transit telescopes (sometimes called meridian circles) to mark the passage of a star over the meridian, heliometers (telescopes with split lenses to measure angular distances from their neighbors), and various micrometer adjustable sighting devices integrated with eyepieces.

An ambitious program was initiated in the late 19th Century, the Carte du Ciel, to obtain all-sky astrometry using the newly available high-sensitivity photographic plates and eighteen identical telescopes, each with a 30 cm aperture. In fact, the project proved too ambitious and observations dragged out for more than 50 years, by which time the product was becoming obsolete (254 printed volumes in various formats). This effort was replaced by Hipparcos and Tycho, the latter of which has a similar limiting magnitude (about $11^{	ext{th}}$) and number of stars to the Carte du Ciel. The work invested in the photographic effort has assumed new importance, however, because it provides a long baseline for determining accurate proper motions.

Astrometry based on the Palomar Optical Sky Survey (POSS) has been pursued to provide good guide stars for HST. Guide Star Catalog Version 2.2 provides all-sky measurements to accuracies of 200-250 mas and down to about $19^{	ext{th}}$ magnitude.

More recent astrometric data have been obtained with electronic imagers. One noteworthy example is the Sloan Digital Sky Survey (SDSS) in the optical. The SDSS uses a 2.5-m Ritchey Cretien telescope with a well-corrected 3° diameter field. Toward the edges of this field there are 22 400 X 2048 pixel CCDs optimized for astrometry. These detectors avoid saturation on bright stars (through faster readout and neutral density filters) up to SDSS magnitude $r \sim 8$, and can detect stars down to $r \sim 17$ well. Therefore, they include a huge number of stars from the Tycho-2 catalog (described below) and other astrometric catalogs to establish the overall reference frame, and then extend this frame to their detection limit. The SDSS photometry CCDs saturate at $r \sim 14$, so the astrometric reference can be transferred to them using stars between $14^{	ext{th}}$ and $17^{	ext{th}}$ magnitude, and the photometry CCDs that fill most of the focal plane extend the astrometry to about $r \sim 22$.

We describe the SDSS reduction steps, since they are typical of position determination with digital imaging data. To start, the CCD data are run through a standard reduction computer pipeline, which carries out the steps we described earlier in this chapter to obtain high-quality images. Positions are then measured off these fully reduced images. First, the images are smoothed to minimize noise artifacts; to avoid degrading the resolution, the smoothing length is adjusted according to the image sizes on the data frame.
The pipeline divides each physical pixel into 3 X 3 subpixels, and quartic interpolation is used to estimate the peak position of the image within the subpixels of the peak physical pixel. This result is compared with a first moment position calculation (the sum of the signal times the distance from some fiducial point). A number of additional steps correct these estimates for possible biases. The resulting astrometric accuracy is 50 – 100 mas, the latter at the sensitivity limit of r ~ 22 (Pier et al. 2003).

Another survey that provides accurate positions is 2MASS. The 2MASS survey uses 2 arcsec pixels but with multiple sightings of each source. It has proven possible to use these sightings to obtain accurate astrometry (errors of ~ 80 mas relative to Tycho-2) from the composite images, down to K magnitude of about 14. This accuracy is achieved by modeling the positions of the Tycho-2 stars as detected by the 2MASS cameras to identify and correct a number of error sources, such as wandering of the telescope pointing and drifts in the image distortion. These methods work because 2MASS had a large instantaneous field of view (8.5 arcmin on a side) and was scanned rapidly (~ 1 arcmin per second) so very large areas were covered quickly compared with the time over which the potential error sources changed. Thus, a large number of astrometric calibration stars could be fitted together to obtain an accurate astrometric solution and determine the necessary corrections to the true positions.

Since ~ 1995, Hipparcos has been the ultimate astrometric reference source for the optical range. The instrument concept was very different from the historical imaging approach. It had a Schmidt telescope that included a beam-combining mirror that super-imposed two fields of view about 1° in size and 58° apart on the sky. The detector was an image-dissector-scanner, basically a photomultiplier with the ability to place its sensitive field anywhere over the large sensitive area that filled the usable focal plane of the telescope. A fine grid was placed over the field viewed by this detector, with alternating opaque and transparent bands. The satellite was put into a slow roll causing any star image to conduct a controlled drift across the grid. The region around a known star was isolated with the image dissector scanner, with a field diameter of 38 arcsec. The resulting modulation of the star signal as the telescope scanned it over the grid produced an oscillating signal. A similar signal was produced by a star in the second field of view, differing by 58° in placement on the sky. The phase difference between the two signals could be analyzed for an accurate determination of the apparent angle between the two stars. These relative positions are ambiguous at the level of the period of the grid, 1.2 arcsec, but previous measurements of the stellar positions just to a modest fraction of an arcsec permit this ambiguity to be removed. Positions of more than 100,000 stars were measured (complete to magnitude V = 7.3) to
an accuracy of ~ 1 – 3 mas in this way, with proper motions (based on comparison with earlier astrometry such as the Carte du Ciel) typically accurate to 1 – 2 mas. Comparing these numbers, it is clear that the current-epoch accuracy of the Hipparcos coordinates has degraded significantly due to the uncertainties in proper motion – the epoch of observation is 1991.25, so by 2011 typical errors were about 40 mas.

The satellite carried a second instrument that gathered astrometric data to an accuracy about 25 times lower, but on more than two million stars. It also obtained homogeneous B and V photometry. These measurements are contained in the Tycho-2 catalog (the current best reduction), which is 95% complete to V = 11.5 with positional accuracies of 10 – 100 mas (better for brighter stars) and proper motions typically accurate to 1 – 3 mas/yr.

The next step for space astrometry is the European Space Agency GAIA mission. GAIA goes back to image analysis, but in a grand way. As with Hipparcos, a key element is for the spacecraft to roll slowly, and for images from two widely separated areas on the sky to be brought to a single focal plane. At the focal plane, GAIA will have large number of CCDs, aligned so the stellar images can be tracked across the detector through time-delay-integration (TDI – Section 3.4.2). The readout rate can be relatively slow, just matching the arrival of the charge packets at the CCD output register; as a result, the read noise can be kept low. Because GAIA will use so many detectors and in a high performance mode, it can achieve a huge gain over Hipparcos. It is expected to reach 20th magnitude with positional errors of a few hundred micro-arcsec (µas) and to make measurements to about 4 µas at 12th magnitude, the sensitivity limit of the Tycho catalog (Perryman et al. 2001).

Further Reading:

Howell, Handbook of CCD Astronomy, 2nd edition, 2006 -- Good general coverage of CCDs and their use in professional astronomy


Additional References


Figer, D. F. et al., 2003, SPIE, 4850, 981


