Chapter 2

Data Acquisition and Reduction

The vast majority of the data for this thesis are in the form of continuum imaging. That is true for all wavelength regimes, optical, infrared, and radio. The remaining small fraction of data consists of mid-infrared photometry (with no spatial resolution).

2.1 Obtaining Optical Data

In the optical, a charge-coupled device (CCD) was used in combination with either broadband or narrowband filters. For most comets, in order to get a sufficient signal-to-noise ratio in a short amount of time, the ~1000 Å filters were necessary. We typically used the Cousins R and I filters. Bessel (1990) discusses the spectral responses of these filters; Bessel (1979) and Zombeck (1990, p.100) discuss the photometric zero points.

For the bright comets, Hyakutake and Hale-Bopp, narrowband (~50 Å) "comet filters" could be employed. The narrow widths can isolate portions of the spectrum that are relatively gas-free, to just sample the scattered solar continuum. There are currently two sets of narrowband filters in existence, one from the International Halley Watch (A'Hearn 1991, Osborn *et al.* 1990), the other recently developed specifically for Hale-Bopp, with improvements in the wavelength ranges to remove contamination to the continuum filters by unwanted gaseous emission (Farnham *et al.* 1999). Fortunately, it turns out that most of the strong gas emission occurs in the bluer end of the optical spectrum, making R and I bands fairly free of gas emission lines.

The basic procedure for obtaining calibration data for the CCD is as follows. Images of the blank twilight sky (or, if not possible, of a blank space inside the telescope's dome) were used to remove pixel-to-pixel variations in the CCD response, i.e., to "flatten" it with a "flat field." Sets of zero-exposure frames were taken, at least twice during a night, to measure the bias count level of the CCD. All CCDs used in this study had a low enough dark current to make it unnecessary to perform that calibration procedure. To measure the photometry and account for the extinction of the atmosphere, standard stars were observed during the night at various zenith distances.

I note that some of the optical imaging has come via the *Hubble Space Telescope*. The Space Telescope Science Institute of course has a detailed set of calibration and reduction procedures that they incorporate into the *HST* data, so the scientist frequently obtains science-quality images with very little further processing necessary. The only processing I personally have done to *HST* data that I use in this dissertation is to remove cosmic ray-affected bad pixels.

2.2 Obtaining Infrared Data

Of the infrared data I have used for this dissertation, all sets measure the thermal emission from the comets, and reside in what is loosely called the "mid-infrared" wavelength regime, from about 5 to 25μm. Thus my use of the word "infrared" or "IR" should be taken to refer to this wavelength range. Strictly speaking the "infrared" part of the electromagnetic spectrum includes 1 to 4 μm flux that in comets is usually dominated by scattered sunlight. For my purposes it is important to be only measuring the thermal emission, not the scattered, in the infrared.

Recent advances in infrared detector technology have made it possible to create array detectors, thus bringing high-spatial resolution imaging to these wavelengths. This is a critical aspect to this dissertation, as will be seen, since it allows us to separate the comatic and nuclear contributions to the flux.

At this wavelength range, room temperature objects near the detector (e.g., the telescope, the sky itself) provide the vast majority of the counts; the astronomical source is usually only a small 0.001% or 0.01% excess on top of all that terrestrial flux. Thus "chopping" and "nodding" are employed to remove all of that. The former involves the secondary mirror of the telescope oscillating back and forth, usually 2 to 5 times per second, so that the detector sees alternately the field of view with the comet and a field of view some distance away – I often used a "throw" (offset) of 30 to 60 arcseconds. The difference of the two fields leaves the comet, although the subtraction is not perfect because the sky's apparent brightness is not necessarily the same in the two frames. To correct this one nods the telescope off the source by some distance – again, I used 30 to 60 arcseconds – and does the same procedure as before with chopping and subtracting. If the nod is not too far then the difference of the two difference frames will remove all of the focus problems and sky variations and retain just the comet. In summary, one obtains four frames, first one on the source, then one off the source after chopping the secondary, then another one off the source after nodding the telescope, and finally yet another one off the source after chopping the secondary with the telescope still at the nodded position. The workable image is: (first minus second) minus (third minus fourth), that is, the result of a double difference. A caveat here is that for the bright comets the nod and chop frames cannot be so close to the comet's photocenter that one accidentally incorporates coma in the three off-source positions, since then some of the coma signal would be subtracted off! A schematic of this chopping-and-nodding idea is shown in Fig. 2.1.

In practice one obtains several "first" and several "second" frames, combining them via the average or the median, to get a more accurate "first" and "second" frame. Then the nod occurs, and the same thing happens for the "third" and "fourth" frames. This is done since nodding takes several seconds but chopping is relatively quick, at a rate of a few hertz. To clarify my nomenclature, an "image" of a comet is built up from averaging or medianing several "frames" together from the 4 positions, and then taking the double difference. Commonly we used 5 to 10 frames at each of the four positions before creating an image.

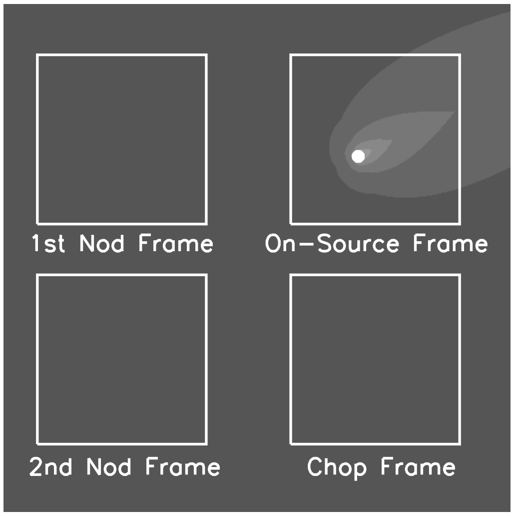

Figure 2.1: Schematic of Mid-Infrared Observing. Here is the basic idea for the ideal method of observing in the mid-IR. One uses four frames and their double difference to actually get an image of the comet. Note that the three off-source frames do not cover any of the comet's coma.

To account for atmospheric absorption and obtain a photometric calibration one observes standard stars. At this wavelength range the behavior of the atmosphere is not necessarily as straightforward as in the optical, so to be safe it is wise to pay attention to the humidity and see if the magnitudes of the standard stars as a function of airmass are not following a straight line.

The flattening of the array can be done by a variety of methods. One method is to observe a star multiple times at various locations on the array, calculate the relative photometry, and then interpolate for the rest of the array. One drawback is just that – the uncertainty in interpolation. Moreover you *a priori* have to know that the pixel-to-pixel variations in the array are smooth enough to be well sampled by this shotgun technique. Another subtlety is that one must be sure to observe the star over a large enough region on the array to include all of the observations of the comet; i.e., it is difficult to extrapolate the flat field, so any images of the comet near the edge of the array have much larger errors associated with their flux.

An alternate method is to stare at a blank sky and then at the inside of the telescope dome, and take the difference of the two images. The sky is a fairly uniform emitter but when looking at the "blank sky" one is really seeing the contribution from the hot telescope as well (not just atmospheric emission) and indeed that can dominate the signal. The telescope's dome on the other hand is brighter than the telescope and swamps the detector; that is, in a sense one sees more flux in the mid-infrared with the dome shutter closed than when it is open! Subtracting the "blank sky" image from the dome image effectively takes away the telescope's contribution, and the observer is left with a flat field for the IR array. Of course one does this multiple times, say ten times, to build up good statistics.

2.3 Obtaining Radio Data

In this wavelength regime again I have only looked at a small fraction of the full part of the electromagnetic spectrum classified as the "radio" part. My radio data covers the X band, i.e, a wavelength of 3.55 cm, and has a bandwidth of 100 MHz. This wavelength was chosen mainly for two reasons: (a) I desired to detect as little of the coma as possible and the longer the wavelength the fewer dust grains there are, and (b) the sensitivity of the centimeter-wave receivers is near its maximum.

Only comets Hyakutake and Hale-Bopp were observed at this wavelength, both at NRAO's interferometer, Very Large Array (VLA). At least 26 of the available 27 telescopes were used at all times. The observations were all done almost totally automatically. A VLA user typically writes an observing program (with a syntax applicable to the telescope control computer) and submits it; the observatory does the rest and the user later picks up the data via Internet or magnetic recording material. For the Hyakutake observations, a colleague was dispatched to oversee the experiment; during Hale-Bopp's apparition, everything was done remotely.

For flux calibration one observes a calibration source – in this case, a quasar near the comet – at the beginning and end of each observing day, or "track," in the parlance of the radio astronomer. Since these were interferometric observations, it is necessary to monitor the phase stability of the telescopes; this is done by observing a bright (~ 1 Jy) source near ($\sim 10^\circ$) the target roughly every 45 minutes or so.

The reduction uses the Astronomical Image Processing System (AIPS) software package specifically designed for this interferometric data. The procedure is outlined in *The AIPS Cookbook* (National Radio Astronomy Observatory 1997). The basic idea is to flag the bad visibility data points, compare with the flux calibrator, then do the inverse fourier transform to obtain an image. Deconvolution can then be employed using the CLEAN algorithm (Högbom 1974), although in this case there was not much difference since in one case the comet was not detected, and in the other case the comet was a point source.

2.4 The Ideal Dataset

It is worthwhile to clearly spell out exactly how the ideal observing campaign would proceed for the study of the nucleus. Of course reality often prevents one from performing this, but here are my observational goals during an experiment.

Two observing runs would be scheduled simultaneously, one at an optical telescope and one at an IR telescope. Obviously colleagues' assistance is vital. Each run would last at least four nights. This length of time and the simultaneity allows us to follow the rotational variations in the comet's brightness in both wavelength regimes. At both telescopes we would obtain continuum images at two or three wavelengths, cycling through them continuously. We would use another filter every so often to have better spectrophotometric wavelength coverage. The images would contain coma, and we would see the coma out to several PSF FWHMs away from the photocenter. Of course the data would be photometric since we are after absolute brightnesses.

We choose the targets that are observed during our telescope time by two methods. First, we find which short-period comets are within roughly 1 AU of Earth; of course we try to choose a time for the observing run when we would maximize the number of possible targets. It was our experience that the typical comet that is farther than about 1 AU from Earth is exceedingly difficult to observe, so much so that one cannot usually even find the comet on the instrument monitor. Hale-Bopp of course was an exception to this.

The second criterion for choosing targets is more up to random chance. Occasionally a long-period comet that was discovered after the telescope's proposal deadline will be visible in the infrared sky at the time of the scheduled run. This is usually the only way to observe long-period comets: by fortuitous accident. Hale-Bopp again was a notable exception. If a long-period comet is available, and all else is equal, that new comet will take observational precedence during the run over the short-period objects.

2.5 Processing the Data: Coma Removal

A cometary image usually includes flux from the coma. To understand the nucleus requires accounting for this contribution and deleting it. For this thesis, this was a severe problem for comets Hyakutake and Hale-Bopp, and a less severe but still appreciable problem for the other comets. One way to deal with the coma is to model its shape in the skirt and extrapolate back to the photocenter to calculate its contribution in those few central pixels, since that is where the nucleus is. We dubbed this method the "coma-fitting method." Dr. C. M. Lisse and I codeveloped the computer program that uses it, although we are not the first: Lamy and Toth (1995), Lamy *et al.* (1998a), and Jorda *et al.* (1999) have done similar experiments, although they have concentrated on *HST* optical and low spatial resolution *ISO* IR data.

To use the coma-fitting method, the PSF is required. It is desirable to have as high a signal-to-noise PSF as possible, so usually a bright flux standard star is used. Not only should the total integrated signal-to-noise be high, but in each pixel near the center as well. It is best also if the PSF's wings are apparent. Naturally of course the PSF should be well-sampled spatially, since that will make it easier to find the location of the point-source nucleus within in the image. Unfortunately high spatial resolution and high signal-to-noise per pixel are competing desires, but usually one has no choice about the spatial resolution, since it just depends on the instrument and telescope that is being used. It is also desirable to image the star close to the time at which the comet image was obtained, so that effects that change the seeing – like thermal flexure of the instrument, temperature changes of the telescope, and evolving sky conditions – are not significant.

In addition, the cometary image itself should be of high signal-to-noise, again per pixel, not just integrated. Modeling the coma's shape is easier if there is decent signal in many pixels away from the photocenter. (I define "decent" and "many" below.) However this only holds up to a point, because at a high cometocentric distance a coma's surface brightness is less likely to be correlated with its behavior close to the nucleus. This distance is different from comet-to-comet, so there is no set rule about how far the coma should be imaged. The dust grains in the coma could be fading, or they could be feeling significant radiation pressure before they reach the edge of the image's field of view, making it much more difficult to model their behavior. Related to this, it is always preferable to obtain images with flux that mostly comes from the comet's continuum. If the flux is heavily contaminated by emission from the gas species in the coma, it again hugely complicates the effort to model the coma's structure since the shape of the gas coma is a much more complicated function.

A rule of thumb that has been employed at the telescope is that one should try to see the coma out to at least a few and probably several FWHMs. This guarantees that there is no flux from the nucleus being spread into the part of the coma that is being modeled, and of course with more coma available it is more modelable. Frequently, however, nature does not follow the rules of thumb and the images that are acquired at the telescope show just a hint of coma. As said above, strong coma was detected in Hale-Bopp and Hyakutake at both optical and mid-infrared wavelengths, while a fairly weak coma existed for Encke and Tempel-Tuttle, even in the optical. Moreover, there is no clearly detectable coma at all in the mid-infrared images of the other comets.

The actual procedure for modeling the coma's shape is straightforward. Assuming the coma is strongly present, first a location for the nucleus within or near the brightest pixel is assumed and the image is "unwrapped" about this point, that is, mapped onto the r-θ plane. This is done using a cubic convolution interpolation method. Then a certain number of azimuths – usually 360 – are chosen and the surface brightness of the coma in each azimuth is fit according to $(A/\rho^n) * \mathrm{PSF}$, i.e., the convolution of a power law with the PSF, and A and n are obtained. This is where it is critical that at each azimuth the coma behaves like a single power law, and not, say, the sum of two power laws. Each azimuth can have a different power law, but each must be characterizable by a single A and n. Presently our computer code finds the value of A and n by trial and error, since it is not so easy to analytically derive the best-fit values when there is a convolution integral involved. The fitted region extends from a cometocentric distance 1 or 2 FWHMs away from the photocenter out until the signal-to-noise is too small to be useful. If there are obvious kinks in the surface brightness profile at the azimuth, the fitting region is shortened to not include that.

There is a subtlety here in the way the surface brightness is fit. The PSF is usually not azimuthally symmetric, so it cannot be unwrapped to get a radial profile. That would make the convolution easy, since it would basically only require a convolution in one dimension, r, the radial dimension, but it is rare that the PSF actually is circular. Instead it is necessary to make a separate model coma image from the trial values of A and n: we assume for the moment that every azimuth in the coma has those values of A and n that are currently being tried, and we make a coma map out of those parameters in the x-y plane, convolve that with the PSF, unwrap this image, and then see how well it fits to what the coma actually looks like.

Strictly speaking, this is not the correct way, since adjacent azimuths contribute to each other upon convolution, and our method does not account for this. To do this rigorously would require fitting hundreds of parameters simultaneously by trial-and-error, a computationally intensive prospect. Hence, this simplification was introduced. It does not create a significant error as long as the fitting is done far enough away from the photocenter so that the surface brightness is not changing rapidly, i.e., at least 1 FWHM away from the photocenter.

Once A and n are found for every azimuth, that is all one needs to recreate an image of the comet's coma. The model coma is subtracted from the image and the residual is compared to the PSF. The only slight complication is the pixelization of the photocenter, since in those pixels one must do an integral of an expression in polar coordinates over a Cartesian area.

The whole process is iterated several times by assuming the nucleus' location in a grid of locations within and near the brightest pixel of the image. Of all these trials, the residual that is most like the PSF and leaves as little flux as possible in the skirt is chosen to be the correct one, and that location is declared to be where the nucleus is. One can then move on to the photometry.

I will make a final note concerning images of comets that only possess a weak dust coma, i.e., a coma that does not extend more than a few pixels away from the photocenter. In this case the same algorithm described above is used except there is no fitting of the exponent n to each azimuth. The lack of data simply just does not justify such an extensive parametrization. Instead I let $n = 1$ for all azimuths and fit a value for A that is applicable for every azimuth. As will be seen in later chapters, this approximation works well for the low signal images.