The Sloan Digital Sky Survey Project Book

Index

Survey Strategy

This survey is a very large project and brings technology and techniques which are quite new to astronomy. It is clear that if it is to be successful its planning will have to proceed very carefully, and the many tradeoffs between efficiency and cost on the one hand and the quality of the data on the other must be weighed cautiously. The strategy we are currently planning to follow is outlined in this chapter, along with the motivation for doing what we plan to do. It may still change slightly as we learn more, finish building the hardware, and dream of yet new scientific rewards to be reaped with only small changes in the way we do things.

The Photometric Survey

The purpose of the photometric survey is fourfold: first, to identify and provide positions for a uniformly-selected sample of galaxies, of which the ~ 900,000 galaxies with r' < 18.15 Petrosian magnitudes and the ~ 100,000 luminous red galaxies will constitute the spectroscopic survey; second, to provide precise colors and approximate morphological information for that sample; third, to provide the database needed for identification of quasars by image structure and apparent color; fourth -- wouldn't *you* really like to have a reliable photometric catalog of the brightest 5 x 10^7 galaxies and a comparable number of stars, over some substantial region of sky, with 3% or better colors in five broad filters?

The major factors that need to be addressed in specifying parameters for the photometric survey are angular resolution (pixel size), field size, exposure time, and exposure strategy.

Integration Mode

There are many factors which lead one to consider the time-delay-and- integrate (TDI) or `scanning' mode as the exposure strategy of choice.

First, it results in essentially 100 per cent observing efficiency, since the data are taken and recorded as the exposure progresses. Even with 4-quadrant readout available in the Tek/SITe 2048x2048 CCDs, it requires about 30 seconds to read a chip, and of the order of 7 seconds to prepare one for a new exposure. This has to be compared with the 55 second exposure time provisionally adopted for the survey; readout time would probably dominate the setting time to the next field, but one still has a 67% overhead compared with TDI. The only efficiency loss with TDI is the ramp-up and ramp-down time (one must start a full frame height before taking data and go a full frame height beyond the end of the imaging region for complete data retrieval; even if resuming a previously stopped scan, one must go a *chip* height). This makes TDI rather inefficient for pictures of a single object, but if one is to scan for hours, a few minutes at the beginning and end hardly matter.

Second, the fact that each object traverses each chip vertically reduces the flat-fielding problem to one dimension, and most column trap-type defects which do not completely block charge transfer disappear. Flat-fielding problems, especially severe in the near-IR because of the complexity of the sky spectrum, completely disappear even for badly-behaved devices. The flat fields can also be generated on the fly with median techniques looking directly at the background, since the background is dynamic and a given horizontal pixel is sky most of the time. These techniques have been developed for the four-shooter scanning surveys at Palomar by Schneider and Gunn (cf. Schneider et al. 1989) and work very well.

Finally, the technique facilitates multicolor photometry with good time response if the focal plane is big enough; our focal plane puts six columns of five chips each along the scan direction, with five different filters (see the discussions in Chapters 2 and 4) for our five bands u' , g' , r' , i' , and z' . The total elapsed time across the array is about 5.6 minutes.

The difficulties with TDI are not negligible, but are, we believe, surmountable. The optical design must have very low distortion; this is addressed in Chapter 2, and we have a design which is excellent from this viewpoint.

Second, the chips must be exquisitely aligned rotationally, so that stars traverse columns very accurately. For a 2048 x 2048 chip, if we want no more than 0.25 pixel error (which is easy to measure) the rotational alignment must be better than 1 arcminute. This is not, in fact, difficult to do, but one must be careful. We will try to do at least factor of two better.

Exposure time

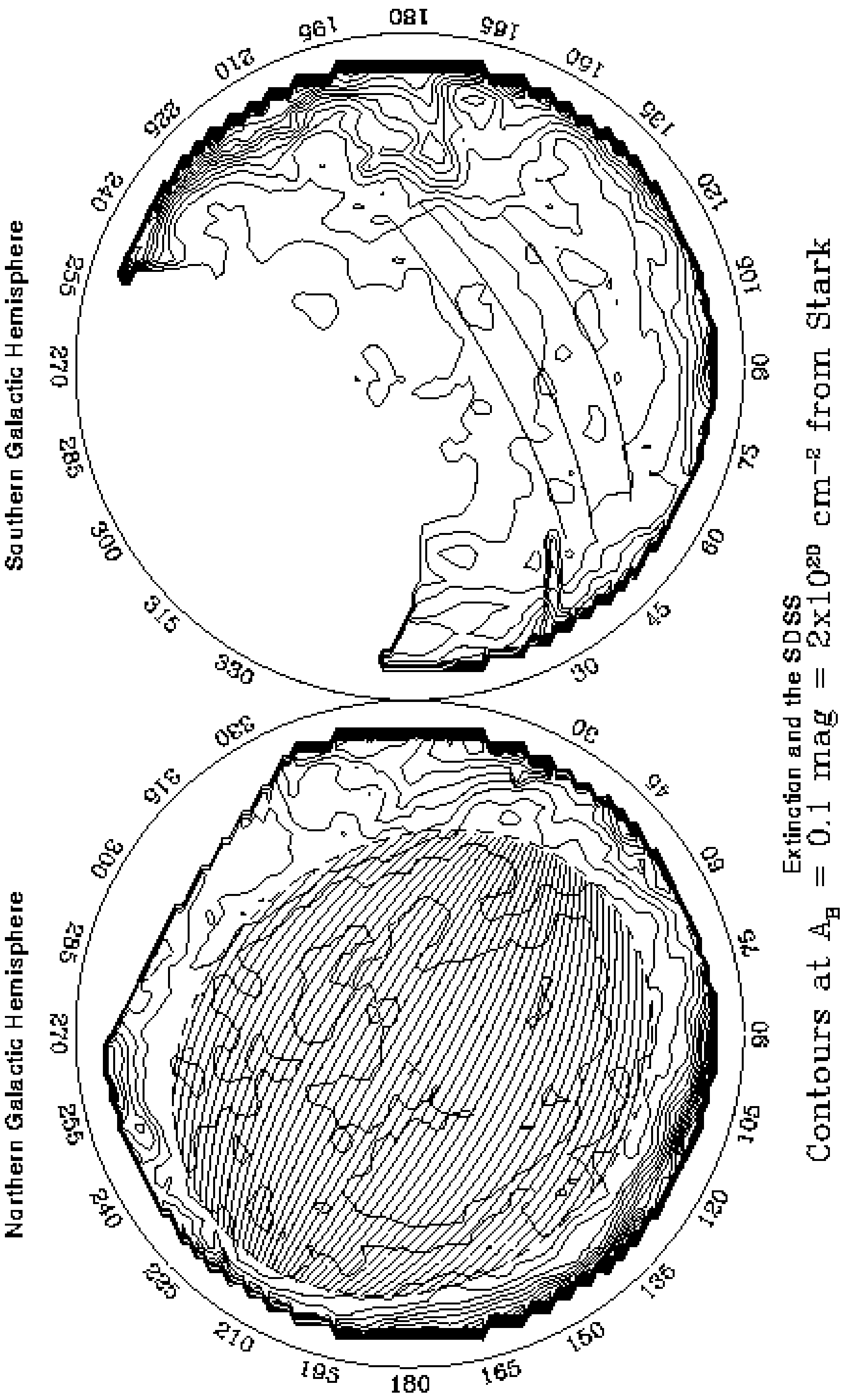
The issue of exposure time is, to some extent, a matter of taste. We will see that we get a quite good signal-to-noise ratio on galaxy images near the spectroscopic limit, and could, in principle, back off a little on our proposed integration time, but there are some technical limitations. If we scan at the sidereal rate, 15 arcseconds/second, (which is *not* accomplished by simply parking the telescope for any declination except 0°) the time taken by a star to cross the 13.5 arcminute field of one CCD is 55 seconds, and so that is the exposure time for that rate. At that rate, the line rate for each chip is 37.5 lines/second, and the pixel rate about 77 kHz. Tektronix/SITe kindly splits the serial register into two halves, so we need read only half the 2048 pixels (plus 20 extended register pixels and 20 overscan pixels), through each amplifier, for a rate of about 38 kHz. This is nearly optimal; the read noise rises like the square root of the rate when one goes much faster than this, and the read noise is already a major contributor to the noise at u' ; for a rate twice sidereal, the flux is down a factor of 2 (to about 19 per pixel in the sky) in u' , the nominal noise is up to 7 electrons, and the readout noise completely dominates. Since our u' sensitivity is none too good anyway, this quickly becomes catastrophic. There are other problems, such as suitably accurate and sufficiently inexpensive 16-bit A/D converters not being fast enough. So going significantly faster is problematic. At the sidereal rate, the imaging survey will take about 20-25% of the survey time, so going faster does not really save very much in any case. Going slower *would* have considerable impact upon the time-to-completion, not so much because of the time added but because we expect conditions to be good enough for the imaging survey (good seeing *and* photometric) not more than about a quarter of the time. Thus a rate near sidereal, and consequently an exposure time of about a minute, appears to be about optimal; we will assume the sidereal rate throughout this document.

Planning the scans: The survey footprint and coordinate system

For a given duration for the survey, we wish to maximize simultaneously the total number of galaxies, the survey depth, the linear scale projected on the sky, and the degree of completeness. The nominal survey objective is to obtain redshifts for 10^{6} galaxies. The largest contiguous solid angle that can be surveyed from one hemisphere that excludes the Galactic plane is about pi steradians, corresponding to a cone with an opening angle of 120°. Within this area there are ~ 10^6 galaxies brighter than r' = 18.15 ; this limit is about 5 times deeper than the current B = 15.5 surveys, and there is good sensitivity to large-scale structure to z ~ 0.2.

The geometric boundaries of the survey region need to be considered in some detail. The cone b > 30° seems natural, but there are severe problems with it. The latitude of the site is 32.8°, which carries the southern boundary of the survey to an altitude of 24.6deg at transit (delta = -32.6°). At that altitude differential refraction effects are catastrophic. If we apply the prescription of Burstein and Heiles (1978) for Galactic extinction to the Heiles (1975) HI maps with or without his `residual' galaxy count maps (Heiles 1976) we also see that a region centered on the pole is *not* the best if one is attempting to use the largest contiguous region of low extinction. In particular, the region near longitude 0° has much higher extinction than the one in the anticenter direction, and one should tip the region by several degrees in that direction. It hurts not at all from the extinction point of view to tip it northward in declination as well, nor to make it slightly elliptical so that its declination extent is not quite so great as the extent in right ascension. We have finally provisionally decided on an elliptical region centered at 12^{h} 20^{m} , +32.8° (so that it passes overhead), whose minor axis is the meridian at that right ascension, with extent +-55° in declination. The major axis is the great circle perpendicular to that, and the extent is +- 65°; it extends from about 7^{h} 6^{m} to about 17^{h} 34^{m} . The most southerly declination is -22°, so if one stays near the meridian when working in the South, the minimum altitude is about 35°, which is manageable. Tipping to earlier right ascension has another benefit which is probably fully as important as the absorption; the site suffers from the same monsoon season as Kitt Peak, and pushing the `prime time' a little earlier in the year is a boon, not to mention the fact that the nights are longer. The survey footprint as presently defined is compared with the extinction contours calculated from the HI column densities measured by Stark et al. (1992) in Figure 1.1. We have tentative plans to rotate the ellipse by ~ 20° to match regions of low extinction better, and will use the extinction contours derived from the COBE/DIRBE maps by Schlegel (1995). The detailed considerations are summarized by Kent (1996).

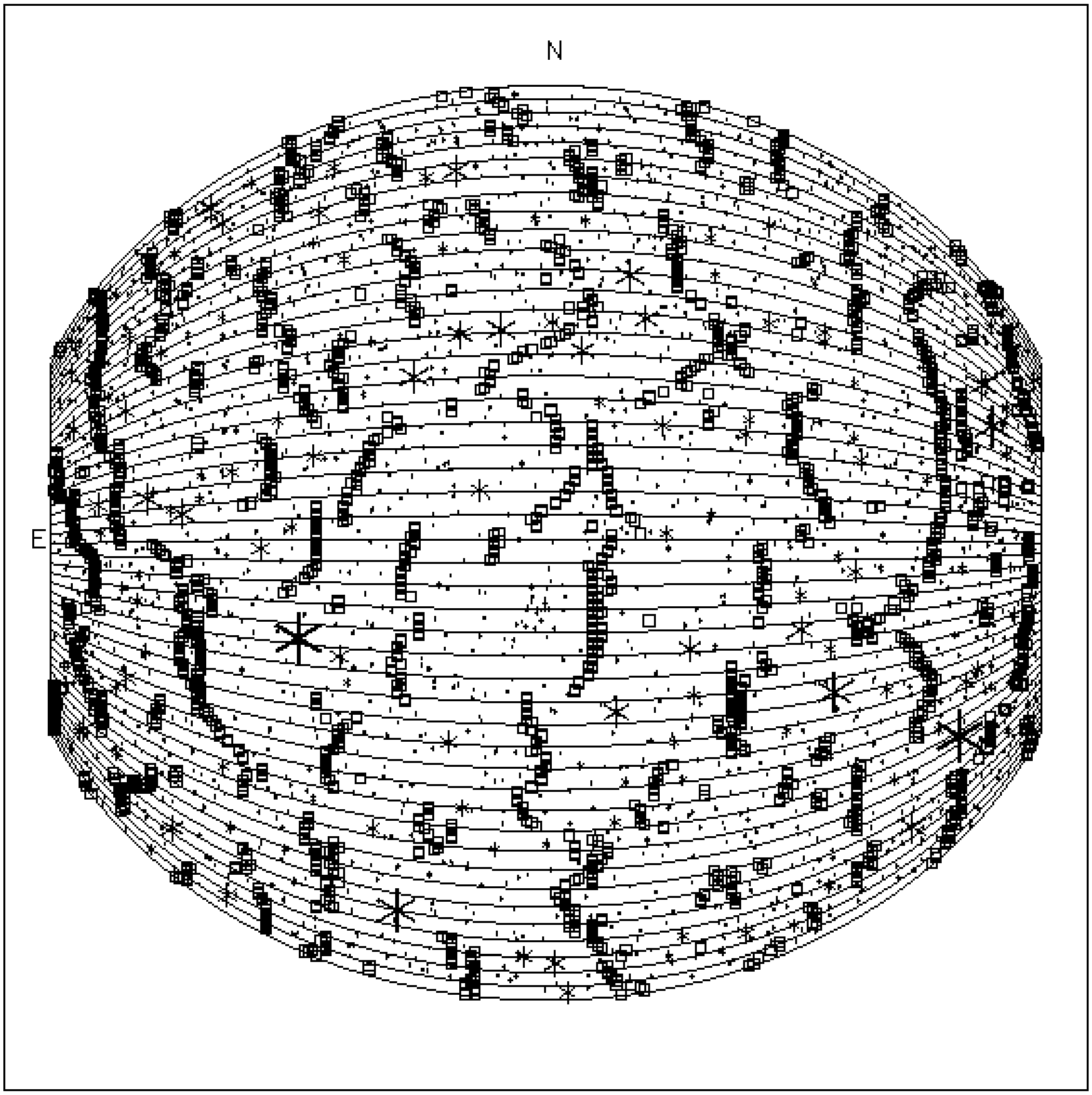
Figure 1.1



The footprints of the Northern and Southern SDSS surveys. The tracks for the photometric survey are shown by heavy lines. The contours show the extinction measured from the HI column density.

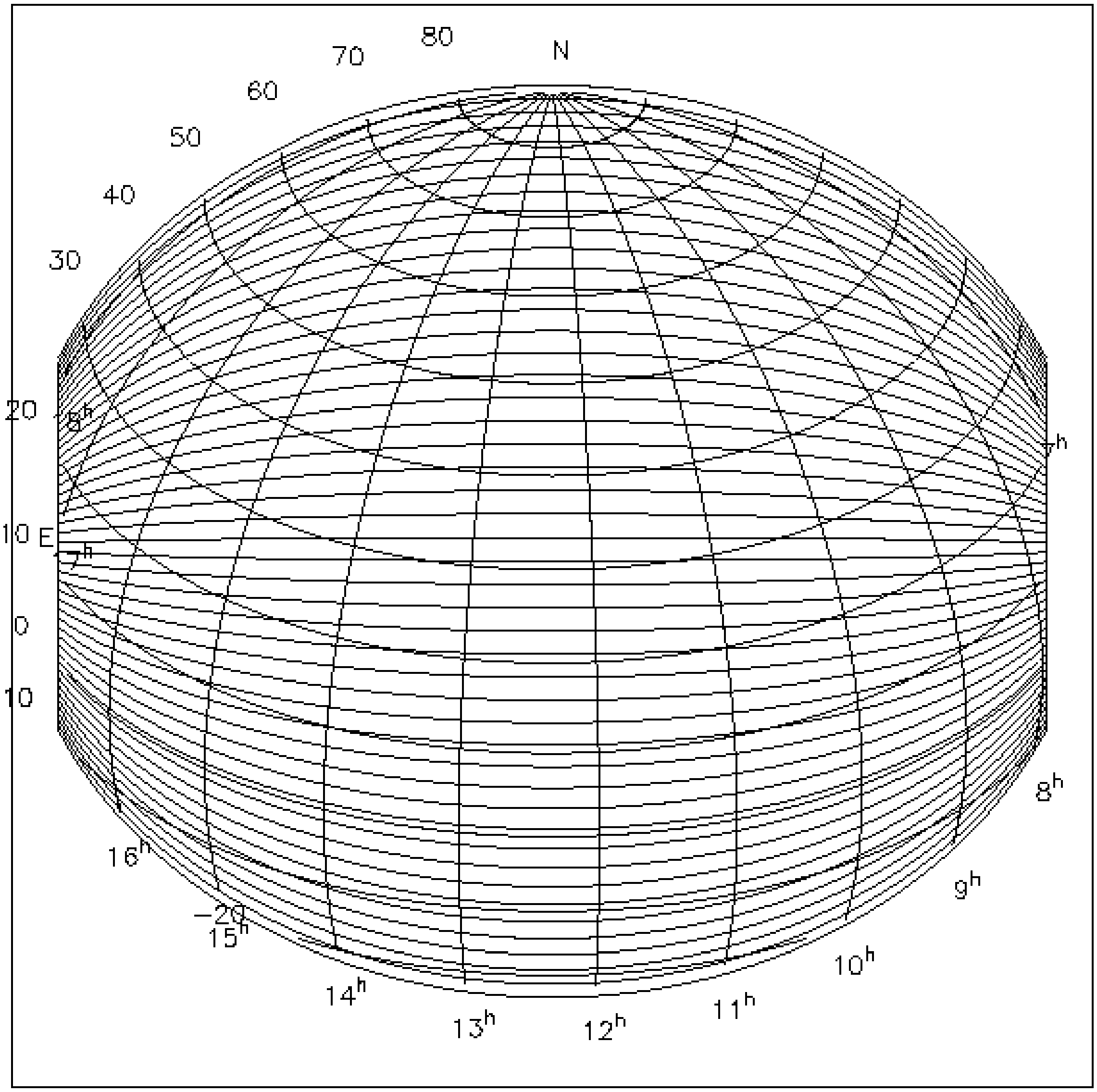
We should spend some time here considering the strategy "in the large" for the photometric survey; i.e. how we in fact scan to cover the survey region. We defined this above as an elliptical area on the sky, 110° by 130°. It is clear that we must scan along *great circles* in the sky to minimize transit-time differences across the imaging CCD array, which with the telescope/camera design presented in Chapters 2 and 4 is 2.5° square; the projection errors from the finite size of the array are only just negligible for our field, and we cannot deviate appreciably from the great circle requirement. It seems reasonable to scan as nearly as we can in right ascension, to minimize the motion of the telescope and the change in airmass in a given scan. This suggests that we arrange the scanning in the following manner: (see Figures 1.1, 1.2 and 1.3).

Figure 1.2



Projection on the sky of SDSS survey area. `N' is the North Celestial Pole. The stripe longitude lambda is measured from the survey central meridian positive to the east along the great circles perpendicular to that meridian; the latitude eta is measured along that meridian to the relevant great circle, positive to the north. The plot is centered on lambda , eta = 0,0. The positions of the stars in the Yale Catalogue of Bright Stars are shown down to about 7th magnitude. The largest symbols represent 0^m stars and the smallest stars of 5^m and fainter. The Monitor Telescope patches are shown as squares and are placed 6 per 15° of lambda to avoid bright stars.

Figure 1.3



Right ascension-declination grid superposed on the survey area. The projection is the same as in Figure 1.2.

The meridian which passes through the center of the survey area, $12^h\ 20^m$, defines the *central meridian* of the scan. The great circle perpendicular to it passing through the survey center at delta =32.8deg; is the *survey equator*. A point in the survey region is defined by a *survey latitude eta* which is the angle between the survey equator and the great circle passing through the point perpendicular to the survey meridian, and a *survey longitude lambda* measured positively eastward from the survey meridian to the point along that great circle. Note that the nature of the constant longitude and latitude curves is backwards from the usual; the constant latitude curves are great circles, and the constant longitude curves are circles centered on the *survey poles*, in this case an *east* pole and a *west* one, at delta = 0 , alpha = $18^h\ 20^m$ and $6^h\ 20^m$. The constant latitude curves are the scan tracks. A single scan, called a *strip*, must be combined with another to make a filled *stripe* about 2.5 degrees wide. The constant latitude lines converge toward the survey poles, so the two strips in a stripe *cannot* be centered on great circles. Instead, a suitable track in the array, about a half-chip width away from the center (cf. Figure 2.2) tracks the great circle which is the *stripe equator*; the other strip then uses the other one.

The strips need not be scanned their full length at one time; if one does, they are $8^h\ 40^m$ long at the survey equator, decreasing only very slowly as one goes to higher or lower survey latitude. There are 45 stripes, 90 strips, in the survey; this is one more than the arithmetic would suggest, but having a fiducial stripe *centered* on declination 0°, where one *can* just park the telescope, will be very valuable indeed. There is clearly a lot of overlap in the stripes as one approaches the survey poles; this amounts to about 28% over the area of the survey (expressed as the fraction of the area of *sky* covered twice). At the ends of the stripes closest to the poles there is, in fact, a small region near the center of each stripe which is covered by *three* stripes. The southernmost and northernmost stripes dip to 35° altitude on the survey meridian and should be done very close to the meridian under the very best conditions; otherwise, it seems prudent for the sake of keeping the data most useful as one progresses to build contiguously out from the survey equator. It may have an unfortunate effect on the spectroscopic tiling strategy (Chapter 8) to have three such growing regions, but it is probably necessary for the sake of the homogeneity of the survey.

Pixel Size and Object Classification

The specific tradeoff of pixel size versus field size is addressed by fixing the pixel size first and then stuffing the focal plane with enough CCDs to use the entire field, which should have been designed as large as possible. Actually, it does little good to pack the CCDs across the scan direction much closer than their sizes, since one must always make two passes to cover the sky. The tradeoff here is how much sky one wishes to cover twice, since sampling twice gives one all manner of useful data. We have here, largely for mechanical and economic reasons, chosen to have a quite small overlap from one scan to the next, only about 8%.

For identifying galaxies and studying galaxy morphology, it is essential that the pixel size be small enough to take advantage of the best seeing. Morphological typing of galaxies is important in a number of applications, and we consider the morphological and color data which will emerge from this survey an absolutely necessary part of the survey, without which the redshift data are far less valuable.

The required pixel size is proportional to the size of the seeing disk (by which we mean here the total image size, as contributed to by seeing, optics, tracking, etc. -- but clearly one wants it to be *dominated* by seeing), and typical seeing varies considerably with site, and with telescope at a given site. At the 4-m Mayall telescope the median full width at half-maximum is about 1 arcsecond, and at the Multiple Mirror Telescope the median is about 0.8 arcsecond. One should sample the seeing disk with no fewer than 2 pixels per full width at half-maximum, and a scale of 0.4 arcsecond per pixel then provides proper sampling for seeing that is characteristic of the Multiple Mirror Telescope site, and, we believe, for Apache Point. Our error budget is such that we will attain 1-arcsecond images when the seeing is 0.8 arcseconds, and our sampling will be more than adequate unless the seeing is *very* good.

The classification of images as either stars or galaxies depends on the pixel scale, the seeing, and the signal-to-noise ratio per pixel. For this purpose one wants to use the largest pixels possible while still being able to recognize the most compact galaxies at the survey limit. As a secondary consideration, the pixels should be small enough that star contamination does not seriously impair galaxy identification or photometry. Given arbitrarily high signal-to-noise ratio and perfect flat-fielding, even poorly resolved galaxies can be distinguished from stars, so one must find the best compromise between pixel scale and integration time. The dominant source of noise is statistical fluctuations in the night sky. The dominant source of systematic error in the night sky is a combination of bad flat-fielding, interference from nearby objects, and scattered light from bright stars. In crowded fields (e.g., in a cluster of galaxies or at low Galactic latitude) the latter two are serious; with CCDs of the quality obtainable today and especially in TDI mode, flat-fielding errors are completely negligible. It would appear that one could get by with somewhat bigger pixels for this problem than that of morphology, since differences between galaxy profiles (particularly small de Vaucouleurs ones) and stars are most sensitively detected at about 1 seeing diameter away from the center; having *smaller* pixels, however, is clearly all to the good so long as there is enough charge so that shot noise dominates the read noise.

The most difficult galaxies to identify are compact ellipticals (because they look like stars) and low surface brightness objects (not because they are difficult to classify, but because they are hard to *find*). Consider the former. Without moon, the night sky has a surface brightness $\mu_r = 20.8$ (Thuan-Gunn (1976) r system). Galaxy profiles can be measured to a surface brightness $\mu_r = 24$ mag arcsecond^{-2} reliably but systematic errors start to dominate at fainter levels. Photometry of bright field galaxies shows that most have a mean surface brightness within the $\mu_r = 24$ mag arcsecond^{-2} isophote in the range $\mu_{24} = 21.5$ to 22.5 , nearly independent of morphological type (Kent 1985). At a minimum, the survey must identify objects at μ_{24}= 21.5 reliably (compact ellipticals like M32 have μ_{24}<= 21.0).

We have done fairly extensive simulations of the survey imaging data (cf. Chapter 9) to investigate questions like these; For example, a compact elliptical with an effective radius of about 1.5 pixels (0.6 arcsecond) at z=0.216 , at which redshift it is slightly beyond the spectroscopic survey limit (r'=18.15). Tests with the MIRAGE object finder and classifier reject this object's being a star at the 20 sigma level, and a magnitude fainter and 20% smaller, at z=0.288 , still at 10 sigma . At a redshift of 0.576 in the red, at which point the effective radius is 0.35 arcsecond and the total S/N a factor of three smaller than the case at the survey limit in r' , it is still 4.5 sigma away from a star. Indeed, at the survey limit the difference between a de Vaucouleurs law classification and an exponential is 3.2 sigma , so simple classifiers can do very well at this brightness level, which is not at all surprising when one looks at the images in the simulations. Indeed the software which reduces the imaging data works this well (Chapter 10).

At the other extreme, that of low surface brightness galaxies, we should be able to do very well indeed; objects of the size of the very low surface brightness galaxy Malin 1 (see Impey and Bothun 1989) will be detectable a factor of 10 fainter, partly because of our high sensitivity and partly because of the excellent flat-fielding afforded by TDI scanning.

Besides galaxy magnitudes, the photometric survey will provide the largest and most homogeneous multicolor photometry of galaxies and stars by far. The uses of a multicolor survey include: 1) Eliminate an important bias due to the K-correction; 2) Use color as a substitute for morphology at larger redshifts to determine what types of galaxies are being measured; 3) Determine redshifts of faint galaxies using multicolor photometry (Connolly et al. 1995) Provide color information for a low-redshift sample of galaxies that can be used for comparison with high-redshift galaxies to study their evolution; 5) Identify QSOs via their color; 6) Identify many interesting classes of stars, including evolved Population II objects, white dwarfs, subdwarfs, very hot stars, carbon stars, and (dare we say?) very cool objects on the main sequence and below; 7) Calibrate Galactic extinction via its effect on the colors of stars, galaxies, and QSOs; 8) Discover previously unknown classes of objects by their colors; 9) Do optical identification and photometry of objects drawn from catalogs in other wavebands (X-ray, radio).

The Spectroscopic Survey

The spectroscopic survey is specified by the selection procedure: the angular sky coverage; nature of the sampling (e.g. filled versus unfilled); minimum fiber separation; and details of the spectroscopy (wavelength coverage, resolution, minimum signal-to-noise ratio, fiber diameter, number of object fibers, number of sky fibers). The various choices depend upon the program objectives: one would use a different strategy if the objective were to collect the largest number of redshifts than if the objective were, say, to measure the galaxy clustering correlation length to the highest precision. Our goal is to allow the greatest range of analyses to be carried out with the spectroscopic sample, and thus we plan to observe complete, filled, well-defined sample. While this is perhaps not the most useful or efficient observing strategy for some specific analysis, is likely to provide the best sample overall for understanding the structure in detail, the relationship of galaxy morphology with that structure, and certainly for understanding almost all dynamical aspects of the structure. Of course, as our scientific goals include measuring clustering on the largest scales, we do put great emphasis on making the spectroscopic selection as uniform as possible, in a way that allows the selection function with respect to redshift to be measured in as model-independent a way as possible.

The remainder of this section gives a general outline of how one might approach the optimum strategy. The principal conclusions are:

1. A wide-angle shallow survey is preferred over a narrow deep one. The obvious selection is a cone centered on the North Galactic pole; the widest practical angle is a cone of opening angle about 120°. We have

seen, however, that the details of the location of our site, the weather, and the distribution of Galactic extinction favor a slightly elliptical region not exactly centered on the NGP.

2. Galaxies should be selected to have a magnitude inside 3 arcseconds of roughly r' < 19.5, corresponding approximately to a limiting total (actually, Petrosian) magnitude of r' = 18.15. This results from the happy coincidence of the availability of ~ 10^6 galaxies in pi steradians in the north polar cap at that brightness level and the fact that the average brightness at that level over a reasonable fiber diameter is comparable to that of the sky. There are several possible small variants on this theme which determine the exact completeness criteria for the sample, but this should be the rough limit.
3. The fiber diameter should be of the order of 3 arcseconds. This is driven by the requirement of getting a reasonable fraction of the galaxy's light in the fiber and not being overwhelmed by sky. This turns out to be quite easy to do; this corresponds to a diameter (in mm) for which identified vendors can deliver fibers of very high optical quality.
4. The spectrograph should have a resolution (full width at half-maximum) of no worse than 10Å (i.e. 3.0Å per pixel) for measuring redshifts, or 5Å (1.5Å per pixel) for velocity dispersion measurements. We will in fact manage to do somewhat better than this.
5. The usable spectral coverage should be at least 4600 - 8250Å (see below). We will cover 3900 - 9200Å with two double spectrographs.
6. The exposures should reach a signal-to-noise ratio of at least 13 per Å. We will comfortably make this at the limit; this defines our 45 minute spectroscopic exposure time, and determines almost directly the time-to-completion of the survey.

The Magnitude Limit and the Selection Criteria

Is it worthwhile to go fainter than about r' = 18.2 ? Some elementary considerations might suggest yes. If we measure galaxies to a limiting brightness s , then the exposure time tau needed to reach some fiducial signal-to-noise ratio varies as tau ~ s^{-1} (if the sky background is negligible), but the number of galaxies varies as N ~ $s^{-1.25}$. Hence the number of galaxies that can be measured in a fixed length of time varies as N ~ $s^{-0.25}$. Thus, even though the exposure times are longer for fainter sources, we can more than compensate by stuffing more fibers into the focal plane. Nevertheless, several competing factors make it undesirable to go much fainter, and the evaluation of the optimum limiting magnitude is relatively complex.

The main limit on going fainter is competition from the sky background. The dark night sky at our site has a measured V -band surface brightness of about 21.7 mag arcsecond $^{-2}$, or about 21.2 in our r' band. For point sources, the optimum limiting magnitude in the sense discussed above comes approximately where the sky and object contribute equally. The optimum fiber diameter for very faint stars (i.e., the size that maximizes the signal-to-noise ratio) is about 1.5 times the seeing full width at half-maximum. If we want to work in seeing as bad as 1.5 arcseconds, the fiber diameter should be at least 3 arcseconds. (For extended sources, the optimum diameter is larger.) The sky intensity within an area of 3 arcseconds diameter is equivalent to the light from a r' = 19.1 star. Since 65% of the light from a star enters the fiber with this seeing, the limit for stellar objects is about r' = 18.6 . For point sources, the time required to reach a given signal-to-noise ratio for a source of flux s changes rather abruptly from a s^{-1} relation to a s^{-2} one at this brightness level.

The case for extended objects is more complicated. A characteristic break point in efficiency again occurs where the light from a galaxy is comparable to the night sky. For a representative low surface brightness spiral, this occurs at about r' = 18 , or perhaps a little fainter. The change is not so abrupt, and the inefficiency as one goes fainter is compensated to some extent for extended objects since the fainter ones are smaller and one gets an increasing fraction of the flux into the aperture. At a cutoff of about 18 in r' the flux fraction grows something like $s^{-1/2}$, so a tau ~ s^{-1} behavior is also applicable to objects fainter than the sky. Of course, the same considerations apply for bright objects, and one does not do nearly as *well* for bright objects as the s^{-1} relation would suggest. Our combination of r' magnitude and surface brightness cuts yields a sample with a sharp cut-off in light down the 3" fiber at 19.5, which is only a few tenths of a magnitude fainter than the light from the sky.

Other factors argue against going much fainter than r' = 18.2 . First, even at that magnitude, the number of fibers needed in the telescope field is of order 600 and management of that many fibers will be a challenge. Second, as will be shown below, exposure times are still quite long. Third, although the most efficient survey in some particular sense might be one that is narrow-angle but deep, in that case most information on large-scale structure will be derived from objects that are relatively faint and therefore difficult to study in other ways. Fourth, unless the survey covers a large solid angle, the largest linear dimension is along the line-of-sight for which information on correlation functions is more difficult to interpret because of the wide range of galaxy luminosities, and the inevitable aliasing, or crosstalk, between the survey volume and the structures within it (e.g., Kaiser & Peacock 1991). Finally, and perhaps technically most important, the subtraction of the sky spectra is easy when the sky and object are comparable and we can get by with relatively few (we are considering a number like 20) sky fibers, while if we go much fainter the problem becomes very much harder, and at the faintest levels we might be driven to use one sky fiber for each object fiber and chopping, which is used for the Norris spectrograph at Palomar (although many workers do not need large numbers of sky fibers even when doing very faint object spectroscopy; cf. Wyse and Gilmore 1992).

The Filling Factor

Does it pay to carry out an unfilled survey? There are two possible strategies. First, one might observe galaxies in noncontiguous fields. One might have gaps of order the telescope field diameter of 3°, in which case the corresponding linear gaps are of order 16 h^{-1} Mpc at z=0.1 , roughly the distance scale over which we are trying to measure structure. A second possibility is that if one were restricted by the number of fibers that could be placed in each field, then one could observe only every second or third galaxy. The main effect of incomplete sampling is that we lose information on the smallest scales, and in particular we will be less able to identify poor clusters, groups, and pairs for mass studies. With our proposed dense-sampling strategy, we are limited only by the finite number of galaxies, and since we are firmly of the opinion that the best statistical methods for analyzing the structure have yet to be invented, it is worth preserving as much detail as possible. The suggested volume is roughly $3 \times 10^7 h^{-3}$ Mpc^3, or about 2000 30 Mpc `bubble' volumes, so if correlations are weak or absent on larger scales, many statistical questions can be answered at the 2% level or so.

Spectral Resolution and Range

A good starting point for choosing the spectral resolution is the Center for Astrophysics redshift survey (Tonry and Davis 1979). Spectra from that survey cover 4300 - 6900 Å with a resolution of about 5Å (1.8Å per pixel). The spectrograph does not have good blue sensitivity. Redshifts are determined either from absorption lines or emission lines -- in both cases only a few lines contribute most of the signal. In absorption, three features are dominant: the Mg triplet lambda 5180, Ca lambda 5270, and the Na I doublet lambda 5890. In emission, H alpha is the strongest (and often the only) line. The strongest lines blueward of the sensitivity limit at zero redshift are the Ca II K and H lines at 3933, 3969Å and the G band 4300Å in absorption, and [OII] 3727Å in emission.

The spectral sampling is set by the requirement that broadened absorption lines be well resolved. Although there is a large range in galaxy velocity dispersions, most of the objects in the survey will be moderately luminous galaxies (typically spirals). Nearby spirals like M31 have velocity dispersions that range from 100 to 150 km/s, and we will take 125 km/s as a working value, which corresponds to 5.1Å full width at half-maximum at a central wavelength of 5200Å. Given a finite number of pixels available and the reality of read noise for CCDs, the best resolution is that at which these lines are just resolved. Lower resolution reduces the signal-to-noise ratio because the lines are diluted with continuum light; higher resolution limits the spectral range and hence the number of lines available for velocity and dispersion measurements. Our spectrographs (see Chapter 7) will have a 3 Å projected aperture width which resolves 5.1 Å and fully samples a line with 6 Å FWHM, and does so with 3 pixels per resolution element on the CCD.

A spectral range of about 5100 - 6600 Å in the galaxy rest frame covers many of the major emission and absorption lines. Significant numbers of galaxies will be detected out to a redshift of 0.25, so the upper wavelength limit should be at least 8250Å for H alpha . Alternatively, we might insist on covering [O II] lambda

3727 at z = 0 , and the range lambda lambda 3700 - 6600 , sampled at 1.4Å per pixel, would have the same velocity resolution. In particular, the absorption at H and K is so strong that it is often easier to get the redshift for a faint, high redshift object in which they appear in one's band than a brighter nearby one in which they do not. Since we will have good sensitivity in the blue, the blue cutoff should include them if possible, at least at all but the very smallest redshifts. As described in Chapter 7, we will use two double spectrographs, one side of each covering the range lambda lambda 3900 - 6100 , the other lambda lambda 5900 - 9100 , each range sampled with 2048 pixels. Thus the CaII K line lies within the range for all redshifts and lambda 3727 for redshifts larger than about 0.05.

Provided that the spectral resolution is sufficient to resolve the absorption lines, the minimum signal-to-noise ratio needed to derive a redshift depends mainly on the strength of the absorption lines. For convenience, the signal-to-noise ratio per Å of spectral continuum will be quoted. For an elliptical galaxy with strong lines, spectra obtained in the Center for Astrophysics redshift survey show that one can measure a reliable redshift if the signal-to-noise ratio per Å is at least 8, i.e., one needs to collect 64 object photons $Å^{-1}$ assuming that the noise is dominated by photon statistics from the source. This number must be increased, however, if sky background and/or readout noise is significant. A big problem for some galaxies is that they have weak absorption lines (presumably because they have a significant amount of light from early-type stars) and yet do not have strong H alpha emission. In these cases one may need 2 or 3 times as many photons to derive an absorption-line redshift. We adopt as a guide the goal of obtaining spectra with S/N of 15 per Å. Simulated galaxy and quasar spectra and the sensitivity estimates in Chapter 7 indicate that we can in fact reach this goal with exposures of somewhat less than a one hour with our telescope and spectrographs.

Dealing with Differential Refraction

The 3° field is big enough that differential refraction effects across it are serious, and the 3-arcsecond fibers small enough that chromatic differential refraction is serious. The observing strategy must cope with these issues, and we must resolve at the outset whether they are so serious that we cannot proceed.

With our choice of survey area, and the (probably reasonable) requirement that fields along the Southern and Northern survey borders are taken at meridian crossing, the minimum altitude reached is 34.5deg (+- 30^m hour angle). The difference in position between an image at 4000Å and one at 9000Å at that altitude is 1.9 arcseconds for the mean atmospheric conditions at Apache Point. The center of the image over this band is at about 5100Å, independent of altitude. In comparison, at an altitude of 25°, the lowest if we had used the b > 30° region, the difference is 2.9 arcseconds, almost the full diameter of the fiber. The flux loss when a typical faint galaxy image is displaced by 1 arcsecond in a 3 arcsecond fiber has been investigated in the simulations and ranges from 7 to 20%. A starlike object loses about 16% for a similar displacement in 1 arcsecond seeing, and 20% in 3 arcsecond seeing; the numbers for galaxies are not very sensitive to the seeing. The displacement goes roughly like 1/ lambda^2 - 1/ lambda$_c^2$. Thus the spectrum for an image centered on the fiber at 5100Å is multiplied by a function which looks like

$$T = 1 - L_e[(\lambda/\lambda_c)^{-2} - 1]^2[(\lambda_e/\lambda_c)^{-2} - 1]^{-2},$$

where L_e is the loss fraction at the ends, and lambda$_c$ is the central wavelength in the sense that the displacement at the ends is symmetric with respect to it; lambda$_e$ is the wavelength at either end. The average loss over the spectrum is about a third of the loss at the ends, but that may or may not be reassuring depending on where the strong spectral features are on which one is basing the redshift.

The situation is actually not too bad for galaxies; at the redshifts where bright galaxies are dropping out of the sample, about 0.2, H and K are near 5000Å where the losses are smallest; even lambda 3727, having only lately entered the range, is in a place where the losses are about 24% of those at the ends. H alpha does not fare so well, and is at a place where the losses are 80% of those at the ends at this redshift, but at zero redshift, the loss is only 25% of the end losses, and we will be called upon to use H alpha mostly for low-redshift dwarfs (which also tend to be large on the sky, so the end losses are small.) All in all, the average loss, which varies between 3% and 7%, is something we can live with. The effect as a signal loss is small compared to the increase in sky brightness at this altitude, which is 1.45 times brighter in the continuum than the sky at 1.2 airmasses where the simulations were made, and will require 22% longer exposures to reach the same signal-to noise ratio if the galaxy and sky contributions are about equal. The argument can be made that we should increase the exposure times by of order 25% at these low altitudes if we wish to keep the survey as homogeneous as possible. It is at these same places near the boundaries of the survey where the Galactic absorption is highest, so the problem gets even worse. We discuss this problem in the next section.

A possibility that we are strongly considering is to take 4 very short (one minute) "auxiliary" spectra for every plate, with the telescope offset +- 1.5 arcsecond (a fiber radius) in directions that correspond to the altitude and azimuth in the middle of the exposure. If the very low-level vertical charge-transfer efficiency of the spectroscopic CCDs is high enough (the single-pixel signal is only a few electrons), we can bin the data *on the chip* into, say, 10 bins of 200 pixels each, which would have higher signal than in a single line in the full exposure. This would allow us to reconstruct in some detail the effect of chromatic differential refraction on the spectrum, and would require only about 5 minutes of overhead. One could also use this information to check the accuracy of the astrometry and plate drilling. Even if we cut the exposure time by that much, the increased information would probably be worth the small hit in the signal-to-noise ratio, and may allow us to do spectrophotometry at the level of 5% accuracy.

An effect about which we can do almost nothing is the fact that the spectra of extended objects is different in different places, and the fiber looks at different places at different wavelengths. To the extent that this affects the *continuum* the procedure outlined in the previous paragraph will serve, but not for differences in the strengths of features on the scale of the wavelength resolution.

The other effect of differential refraction is that the images of objects move relative to one another as the exposure progresses, and, of course, the mean positions depend on the mean hour angle. With a spectroscopic field center at an altitude of 36.5° on the N-S boundaries of the survey, the refraction difference from the top of the field to the bottom is 8.1", in the sense that the image is compressed with respect to the sky. The image is also compressed laterally because it is higher in the sky, but the effect is smaller, 2.7" (the ratio is the square of the sine of the altitude, so near the zenith refraction results in a change of scale alone.) We can adjust the scale of the image at the telescope in real time, but if we drill the plate neglecting refraction the errors are too large; for the mean scale, corresponding to a shrinkage of 5.4 arcseconds in 3 degrees, the images are as much as 2.7 arcseconds from the fiber centers, a disastrous difference. Thus one must assume an hour angle, calculate the refraction, and drill the plate to suit. In the course of an hour exposure, the refraction changes (most importantly, the refraction pattern rotates with respect to the field) and at 35° altitude the maximum image motion going from the meridian to an hour over, assuming that the exposure is guided in the middle of the field and the rotation is controlled to minimize the error at the edge, is about 0.8 arcseconds. It is a little less if one starts 30 minutes east and goes to 30 minutes west, but not much. Thus if one drills the plate for the mean position, the errors are at most 0.4 arcsecond, about 24 microns. The scale can be adjusted to make the maximum vertical component of the error about half that. (We obviously care most about the vertical component because the error is there compounding the chromatic effects). The effect of this on the spectra is very small, and corresponds to a seeing degradation of about 0.13 arcsecond in the root-sum-square sense (i.e. going from 1.0 to 1.01 arcseconds.) Horizontally, one gets essentially the full effect, but this is equivalent to 0.26 arcsecond added in quadrature with the seeing, or going from 1" to 1.03" . We are in this happy position *only* if the plate is drilled for the hour angle range actually used, and if the drilling tolerances are smaller than these errors. Happily, the latter seems to be possible, but the former requires very careful scheduling.

Dealing with Galactic Absorption

There are very compelling reasons to attempt to make a survey which is as homogeneous as possible *outside the Galaxy*, i.e. one in which the reddening-corrected limit is at least roughly constant over the survey area. The average color excess at the boundaries of the survey is about 0.08 magnitude, but varies widely. Our photometric survey will itself produce the best reddening and extinction data which have ever been obtained. It is unfortunately not going to be possible to measure the effects of reddening in the photometric survey on a timescale short enough to be able to use this in spectroscopic target selection. Indeed, we plan to measure reddening using hot stars. We will need to obtain spectra of these stars in order to confirm their spectral types, which means that we will not be able to measure the reddening until the spectroscopy has been done. Therefore we will correct our object catalogs for reddening following Burstein and Heiles (1978) or updates, such as the DIRBE maps (following Schlegel 1995).

There will be in the survey area about 4000 hot white dwarfs and O and B subdwarfs brighter than g'=20.5 , which can be distinguished by their colors and which will yield individual accurate reddening values, and a like number of cooler but still useful horizontal branch stars. There will be more than a million F and G subdwarfs, a hundred per square degree, which will be useful at least statistically, and, of course, the galaxy counts themselves, which will go deep enough that several almost independent samples (from the point of view of spatial clustering statistics) can be chosen at different brightness levels. The color distribution of faint galaxies can also be used to measure the reddening.

If one is to compensate for extinction which dims an object by a factor f and attempt to reach the same signal-to-noise ratio as for an object with no extinction, the exposure time is increased by a factor between f^{-1} for very bright objects and f^{-2} for faint ones, all other things being equal. If the extinction over the survey region goes roughly as the secant of some equivalent latitude, then a *very* crude representation of the correct exposure time for faint objects, including the effects of both extinction and sky brightness, referenced to the exposure time at the zenith at the survey center, is something like

$$T \approx 1 + \alpha r^2 + \beta y^2 .$$

Where

$$\alpha = \frac{4A_p}{1+2A_p} = 0.30$$

and

$$\beta = \frac{\csc(a_{min} - 1)}{2} = 0.37.$$

Here r is the angular radius of the point in question from the survey center in units of the radius to the boundary in that direction, y is the angular distance of the point from the survey equator in units of the total survey latitude extent, A_p is the Galactic extinction near the pole, and a_{min} is the minimum altitude at the survey boundary. The first correction term represents the effect of Galactic extinction, the second the brightening of the night sky. This assumes that the exposure is done at small hour angle, and is calculated for a galaxy at the survey limit where the total galaxy signal is, at the zenith, about equal to that of the sky. The values given are for our proposed survey geometry. The mean value of T over the survey region is about 1.25, and at the northern and southern boundaries near the central meridian of the survey is about 1.65, which represents essentially the full dynamic range of the exposure times. Thus near the survey center, presumed done near the zenith, the exposure time should be 20% smaller than mean, and near the boundaries on the survey meridian 32% longer; the limiting magnitude, however calculated, should be adjusted faintward by an amount equal to the Galactic extinction in the relevant waveband.

It is important to note that even if the model one uses for the absorption is not correct in detail, these corrections will help and should be applied. One can correct for residual effects in large-scale structure analyses by means of the selection function, but the smaller the corrections the smaller the remaining systematic uncertainties they leave behind. There need be no guesswork at all about the sky brightness correction, since it is monitored continuously by the guiding system (cf. Chapter 7).

The Test Year

We plan to operate the survey instrument and software for one year to test all systems thoroughly before actually beginning the survey. This is an expensive proposition, but we consider it absolutely necessary for several reasons.

The value of the survey will be severely compromised if it is not homogeneous, so we feel it imperative that insofar as possible, *nothing* be changed that influences the data quality once the survey is underway. For a system this complex, it seems inevitable that there will be troubles and inefficiencies which need fixing, and a year seems a reasonable timescale to accomplish this. There are several aspects about which we are particularly concerned, which include:

Details of the performance of the camera system, which include tracking algorithms, the scale-determining algorithms discussed in Chapters 5 and 7, ghost images, the effects of stars bright enough to saturate the CCDs, the focus servo, and others which will arise as engineering is completed.

Photometric and astrometric calibration. The former will require the setup of a network of either photometric or spectrophotometric standards before the survey begins (for this reason we plan the monitor telescope to go into operation even well before the test year), and we must test in detail the rather unconventional approach to calibration proposed here (Chapter 5). The astrometry *can* be refined as we progress in the survey, but we must know that it is good enough at the beginning that the errors have negligible effect on the placement of the spectroscopic fibers. The photometry must be good enough that there is negligible effect on the selection of the spectroscopic sample.

All the detailed algorithms which make up the pipelines with which we are reducing the data need to be tested with real data. We have made out extensive simulations of the SDSS data with the express purpose of testing these algorithms (cf. Chapter 9), but there are sure to be any number of senses in which these simulations are not realistic enough to test the pipelines thoroughly.

The selection criteria for galaxies for the spectroscopic sample must be thoroughly tested on real survey imaging data, and the statistics appropriate to the sample investigated. This is perhaps the most serious algorithmic problem we must address during the test year, and the one thing we must not under any circumstances change once we are underway.

The selection criteria for QSOs must be similarly tested; this is in a sense a more difficult problem, since the selection criteria are designed to separate stars from QSOs in a statistical sense, and one must balance the need for a homogeneous survey in the face of a position-dependent star-to-QSO ratio (and a position-dependent stellar population), which is certain to have a large impact on the success rate of QSO selection, against the need for an efficient classifier.

The survey time to completion depends on a number of factors over which we have some control, such as the efficiency with which we can change spectroscopic plates, and switch observing modes when conditions dictate. We will learn from experience during the test year how to observe with as little overhead as possible.

Survey Time-to-Completion

The telescope instrumentation has been designed so that switching between the photometric and spectroscopic programs can be accomplished quickly during a night. The photometric program will place much greater demands on the quality of observing conditions: we expect to operate the camera only during clear, moonless periods with seeing better than one arcsecond. Since one of the functions of the photometric program is to identify galaxies for the spectroscopic observations, it is necessary that the imaging survey always be in the lead as far as sky coverage is concerned.

A 130° strip is covered at the sidereal rate in 8.6 hours, and the total scanning time to cover the whole survey field once is about 700 hours. There are approximately 1800 spectroscopic fields in the survey area. With 45 minutes of exposure time, we estimate that it will require a bit more than an hour per setting with all the overheads; then the spectroscopy alone will require about 2000 hours. The right ascension range of the survey is about 10 hours, with the extremes at pretty high declination, so we can work easily 2 hours over and perhaps three. Thus we can work about 60% of the year. If we assume that 50% of all nights are spectroscopically clear, 60% of the time on those nights is sufficiently dark (clearly a nearly-new moon low in the sky is not a problem), on 60% of those nights the survey region is accessible, and that there are 8 hours per night, we find that we can work 535 hours per year. The spectroscopic survey will thus take about 3.7 years, and the photometric survey about 1.3 years, for a total of 5 years.

The elliptical region suggested above has two more advantages from the point of view of hastening the completion of the survey. It is clearly advantageous insofar as time-to-completion is concerned to limit declination coverage and increase right ascension coverage, because at any time of the year one must thereby do less and has longer to do it; thus the elliptical region will require about 10% less time than a circular one of the same area centered at the same place. In addition, the move to earlier right ascension is a move to longer nights at a given limiting hour angle. We may well wish to cover a somewhat larger area with the imaging than the one in which we plan to do spectroscopy, precisely to check the reddening; this will also help with the tiling (see Chapter 8).

It is implicit in the discussion above that the telescope be operating nearly all the time; there are certainly a few nights around full moon when there is *no* time when it is dark, but in general the survey must operate whenever the sky is dark for long enough for anything useful to be done. In addition, we will see in the following section that, at least at the beginning of the survey, we will be able to go directly from the Northern survey discussed above to the deep Southern survey with no idle time. This will change as the survey progresses, because the declination extent of the Northern survey at the right ascension extremes is small, and as those areas are finished, the hour angle gap between the Northern and the Southern regions will grow too large to bridge. This has negligible effect on the time estimates above, but will open the possibility of other projects during those times.

The presence of only a small amount of moonlight will necessitate stopping operation, but we anticipate that some `bright time' will be used at the beginning for calibration exercises (letting the Monitor Telescope catch up with the secondary star patches) and consistency checks on the photometry. Another bright time exercise will be to make scans perpendicular to the great circles, to tie the photometric and astrometric systems together by more than the (relatively small) overlaps between stripes. The need for this time will ease considerably as the survey progresses.

Deep Survey in the South Galactic Cap

In the autumn the North Galactic cap is inaccessible, but a substantial fraction of the South Galactic cap is visible. In principle, the primary survey strategy could be employed in the South, but this involves a relatively small augmentation of the data base. We prefer the option of a deep survey in a smaller area; this provides new and complementary scientific opportunities. Perhaps as important as the considerably greater depth is the opportunity to investigate time-variable phenomena thoroughly.

A deep survey is most efficiently done in a single, long, thin stripe. The Southern imaging survey consists of two interlaced areas that are repeatedly scanned. These scans will be co-added to produce a single, deeper image, and differences between scans will be used to identify variable objects. For example, if the survey is centered on the celestial equator and restricted to regions with less than 0.1 magnitudes of reddening, then we can go from about $20^h\,40^m$ to about $5^h\,0^m$, a total of about 125°, or 8.3 hours scanning time. However, our first priority is the survey in the North, and thus we will observe in the South only in the time that none of the Northern survey area is available. This restricts us to a total of 90°, or 6 hours scanning time. The stripe is covered completely, therefore, in two strips, 12 hours.

The total area of this stripe is 225 deg^2 . The time per year available to it (assuming a 70-30 split with the Northern survey) is 260 hours, with the same ground rules as used above. If we devote a somewhat larger fraction of the time to imaging, say 40%, we will have about 540 hours over five years, covering the field about

45 times and obtaining a limiting flux 2^m fainter than that of the Northern survey. The corresponding 5:1 limiting magnitudes are 24.4, 25.3, 25.1 24.4, and 22.9 in u' , g' , r' , i' , and z' . This schedule makes the mean interval between scans during the season about 20 days, but these will, of course, be concentrated into the dark half of the month and into stretches of good weather. The opportunity to discover supernovae and other variable objects at hitherto unexplored brightness levels is superb (see below), and we may well want to increase the fraction of time devoted to imaging still further in order to improve time coverage, even though the data in some of the extra scans might be of lower quality. A given field will occasionally be observed off-meridian, so as to break up aliasing in searches for periodic variability. Also, for many variable objects one can profitably squeeze the moon a bit, since the r' band is not too badly affected by a little moonlight, and i' and z' can stand a great deal. It is also the case that light cloud cover can be accommodated, since we do not depend on a single pass for photometry; the ability to calibrate scans in less-than-photometric transparency to scans obtained under photometric conditions is the most important factor in allowing us to increase the fraction of imaging time in the South.

The Southern survey has other practical advantages. Since the great circle for its (single) stripe is the celestial equator, the telescope can be parked and act as a transit instrument, which will result in increased stability and superb astrometric accuracy. This, combined with the long-term multiple coverage, will produce a large volume of excellent proper-motion data. In addition, the wide range in right ascension means that much of the spectroscopy can be performed near the meridian.

Finally, we will obtain photometry and galaxy spectroscopy (using the same selection criteria as in the North) in two further stripes in the South (a single pass each), along the great circles that go through alpha = 0^h, delta = +15° and alpha = 0^h, delta = -10°. The addition of these two stripes greatly increases the number of independent baselines of the largest extent, which will be of great use in measuring structure on the largest scales.

Although there certainly exist photometric surveys which go as deep or deeper than the Southern survey proposed here, there is no survey in existence or planned which goes so deep over so large an area of sky, nor are there any over any appreciable area with accurate five-color photometry. We view this survey as an indispensable bridge between the Northern shallow survey and the very deep pencil-beam surveys which will be possible with the new generation of very large telescopes and which go two magnitudes deeper yet.

The principal scientific drivers for a deep, multi-color imaging survey are a search for high-redshift clusters (to $z > 1$), and a characterization of the faint galaxy population, in order to probe the local population of very low luminosity and very low surface brightness galaxies and to study the evolution of galaxies and galaxy clustering. Object identifications can be made from radio and X-ray catalogs; often the most dramatic objects in these bands turn out to have optical counterparts at 24th magnitude and fainter. With the multiple exposures, variability studies will be done at brighter magnitudes. Variability will bring to light classes of AGNs and QSOs that would not be distinguished by their colors alone. Moreover, characterization of the variability of AGNs has much to teach us about the geometry of the broad-line region. Variable stars will also be discovered in this survey in great numbers, at quite faint magnitudes, although the SDSS will not compete with surveys such as MACHO and OGLE for stars brighter than 19th magnitude. Five years is long enough, and our astrometry will be good enough, that we will be able to measure proper motions for large numbers of stars (including perhaps faint nearby stars with *very* high proper motions, hundreds of milli-arcsec/year). Finally, this survey will undoubtedly uncover large numbers of supernovae (we *conservatively* estimate a minimum of 35 per observing season), allowing us to characterize the statistical properties of supernovae quite well, in particular the rate of supernovae in various types of galaxies. If we can reduce the data on a short enough time scale (see below), we can let the community know about those that are discovered, for intensive follow-up on other telescopes.

For the spectroscopic time in the South, we envision a combination of several projects; the following are some examples. The number density of quasars on the sky is ~ 240 per spectroscopic field to r' = 19.5 (i.e. one magnitude fainter than the nominal limit of the quasar survey in the North). We can indeed push the spectroscopy to these fainter magnitude because quasars, as emission-line objects, do not require the signal-to-noise ratio to get a redshift that an absorption-line galaxy requires. We estimate that two hours will be required per field to go this deep, probably broken into two one-hour exposures on different nights to minimize refraction effects. This survey will be invaluable for studying large-scale structure in the quasar distribution. By going deeper in magnitude, we probably will not substantially increase the *redshift* out to which quasars are seen (indeed, quasars are being discovered now at z > 4 that easily pass the Northern spectroscopic survey limits), but we will increase the *sampling* of the quasar density field, and denser sampling allows much more detailed characterization of clustering. In addition, this faint quasar survey will substantially extend the dynamic range for studies of the evolution of the quasar luminosity function, and it will provide information about faint peculiar stars and compact galaxies, which will inevitably find their way into the spectroscopic target list.

This survey of faint QSO candidates occupies only ~ 1/2 the fibers for a total of ~ 50 spectroscopic fields, leaving a large amount of time for other spectroscopic projects. One possibility is to target very blue galaxies appreciably fainter than the galaxy spectroscopic limit in the North; the fraction of galaxies with strong emission lines is quite high if one chooses blue enough objects. Another way to think about spectroscopy in the South is to realize that in the entire Southern stripe, there are roughly 440,000 objects (stars, galaxies and quasars) with 3" aperture magnitude brighter than 19.5 in r' . This is of the same order of magnitude as the total number of objects we *could* target spectroscopically in the time allowed (assuming that we keep to the 45-minute exposure times per object used in the North). This suggests that we observe *all* of these objects spectroscopically; this project would encompass essentially all spectroscopic projects envisioned for the South. It would prove to be a tremendous boon for studies of stellar populations in our own Galaxy, and is bound to turn up all sorts of interesting and unexpected serendipitous objects. It will be absolutely invaluable for understanding the consequences of our selection criteria for quasars, and will allow us to probe in detail how well our star-galaxy separator works (and explore the population of compact galaxies in a way the Northern survey will be unable to do). There is no reason to do this survey of roughly 400,000 objects in a contiguous stripe, of course, and considerations of Galactic structure studies will probably drive us to spread the fields between the three Southern stripes which will be surveyed photometrically (as mentioned above, we still have to survey the two "outrigger" stripes spectroscopically using the same criteria as in the Northern survey).

The strategy of the Southern survey is much more flexible than that of the Northern survey. We do not need to use the same strategy from one year to the next, and we can imagine using the telescope in a variety of ways to address specific scientific problems. It might be worthwhile, for example, to devote a few hours of imaging time to a complete photometric map of M31.

Complete analysis of the Southern photometric data will require two pieces of rather tricky software which are not necessary for the Northern survey: co-adding of frames to go deep, and subtraction of frames in order to uncover faint variable objects. However, the results of addition and subtraction analyses are *not* needed to select spectroscopic targets: all of the spectroscopic targets will be chosen on the basis of photometry reduced exactly as it is in the North. Searches for variable objects suitable for spectroscopy can be done quite effectively at the catalog level, and do not require the much more sophisticated frame subtraction software. Our plan is to start the Southern survey with the same data reduction software that will be in place for the Northern data; no additional software will be needed to run the Southern survey (it should be mentioned that running this survey over repeated scans of the same area is a wonderful test of the robustness of our algorithms, although we hope to find any obvious deficiencies in our algorithms during the test year). As the survey progresses, we will develop and refine the frame addition and subtraction software, which will allow us to exploit the full potential of the Southern data. Because nearly all of the Southern spectroscopic targets will be selected on the basis of the first photometric scan, the turn-around requirements on data processing are much less severe than they are in the North. However, it is clearly valuable to have fast turn-around so that we can notify the world at large about interesting transient phenomena, in time for follow-up observations. Our standard data reduction system should typically allow turn-around in a week or less. For supernovae one would ideally like to have even faster analysis, preferably on the mountain. We are not actively working to develop such capability, but we will provide a "Y-fork" on the data acquisition system so that a group with the hardware and software resources to carry out such an analysis can connect to the data stream.

References

Burstein, D., & Heiles, C., 1978, *Ap. J.* **225**, 40.

Connolly, A.J., Csabai, I., Szalay, A.S., Koo, D.C., Kron, R.G., & Munn, J.A. 1995, AJ 110, 2655.

Heiles, C., 1976, *Ap. J.* **204**, 379.

Heiles, C., 1975, *Astron. and Ap. Suppl* **20**, 37.

Impey, C., & G. Bothun, 1989, *Ap. J.* **341**, 89.

Kaiser, N., & Peacock, J. A. 1991, *ApJ* **379**, 482.

Kent, S., 1985, *Ap. J. Suppl.* **59**, 115.

Kent, S. 1996, "Defining the Northern Survey Area", SDSS Internal Memo.

Schlegel, D. 1995, PhD Thesis, University of California.

Schneider, D., M. Schmidt, & J. Gunn, 1989, *A. J.* **98**, 1507.

Stark, A.A., Gammie, C.F., Wilson, R.W., Bally, J., Linke, R.A., Heiles, C.E., & Hurwitz, M. 1992, ApJSuppl 79, 77.

Thuan, T. X., & J. Gunn, 1976, *P. A. S. P.* **88**, 543.

Tonry, J., & M. Davis, 1979, *A. J.* **84**, 1511.

Wyse, R. F. G., & Gilmore, G. 1992, *M.N.R.A.S.* **257**, 1

The 2.5-meter Telescope

Optical Design

General Considerations

The desiderata are a telescope of about 2.5 meter aperture, of a scale and focal ratio well adapted both to fibers for a many-fiber spectrograph, and to the pixel size of available large-format CCDs for a time-delay-and-integrate (TDI) imaging survey. The telescope should have a field which is as large as practicable, and which should lend itself to the use of fiber-fed spectrographs. The size and throughput of the overall system should allow the efficient acquisition of spectra of galaxies at our spectroscopic limit of r' = 18.15 . As outlined in this section, we believe that we have succeeded in producing a design for such an instrument which is simple and yields excellent performance, with a focal plane well-matched to the unusual requirements of TDI imaging. The ray trace image sizes are less than 0.6 arcseconds RMS over a three-degree field at the design wavelength, and are less than 0.7 arcseconds RMS in any of our filter passbands over the imaging wavelength range of 3500 to 9300Å. The lateral color is very small, of order 5 microns (0.1 arcseconds) over the whole field over the spectroscopic wavelength range of 4000 to 9000Å. The dominant aberration in the spectroscopic mode (with its demand for wideband imaging) is longitudinal color, which creates images larger than 1 arcsecond RMS (to be compared to the 3 arcsecond fibers we will use) only at extreme field angles at the ends of the spectrum.

While wide-field optical designs exist for such focal ratios (the Baker-Paul three-mirror design, for example, and more recent variants, and the Schmidt, of course) they all suffer from either excessive length for a 2.5 meter aperture or an inaccessible focal plane, which renders the fiber spectroscopy difficult. Early design efforts revealed that there appeared to be some hope (which we realize below) that a more-or-less conventional Ritchey-Chrétien-like optical system would deliver the requisite performance while working with a focal ratio about f/5, which we need to match our chosen detectors.

After quite extensive exploration, it was determined that the CCD division of Tektronix was the only viable supplier for focal plane arrays which satisfied the requirements for quantity, large format, and broadband sensitivity. After our contract with Tektronix was well under way, an independent entity, Scientific Imaging Technologies (SITe), purchased the CCD division and its contracts. The need for good ultraviolet response dictated the use of thinned backside-treated devices; the UV fluxes are so low that very low readout noise was also required. With 24 micron pixels, the SITe devices provide excellent focal plane image sampling at about f/5 with our expected image size of about 1 arcsecond FWHM. This combination yields 0.4 arcsecond pixels, about 2.5 pixels per FWHM, and about 1 percent total power in a gaussian star image beyond the Nyquist frequency.

Efficient fibers are available in diameters of 100 to 600 microns which preserve well focal ratios faster than about f/7, and which suffer hardly any degradation at focal ratios around f/5. Fibers which subtend about three arcseconds are needed to cover the bright parts of galaxies in the brightness range of interest. The scale at f/5, 60.6 microns per arcsecond, requires 180 micron fibers, which have excellent optical performance if handled and terminated properly and also give good sampling for the 24 micron pixels of the large Tektronix/SITe imaging arrays at the demagnified focus of the spectrographs. The focal ratio, therefore, was fixed at f/5.0.

As pointed out in the design report for the Swope and Irenée DuPont telescopes by Bowen and Vaughan (1973) it is possible to design a Ritchey-Chrétien telescope with a flat field by making the curvatures of the primary and secondary mirrors the same, which yields zero Petzval curvature in the focal plane. Since a Gascoigne astigmatism corrector is required, and since this element introduces a bit of positive field curvature, the design needs to deviate a little from this prescription, but not by very much. This results in a final focal ratio of just under twice the primary ratio, depending a little on the back focal distance, and a very large secondary, about half the primary diameter for the field sizes obtainable with f/4 primaries (about 3 degrees). For large telescopes one would like to get around the limitation of slow primaries and large secondaries. These desiderata were met in the design of the 2.5-meter DuPont telescope with the introduction of only moderate field curvature.

The Design

The instrument described here uses the same philosophy taken to even faster primary and overall f/ratio. Our requirements are rather unusual for an astronomical instrument, since we must accomplish TDI imaging over a large field. This requires that field of view distortion be carefully controlled, since either a change of scale or a differential deviation from conformality of the mapping of the sky onto the focal plane across a chip translates immediately into image degradation. It is easy to show, in fact, that for the pixel sizes and CCD sizes of interest, *no* axially symmetric optical design is satisfactory for large enough field angles. For a three-degree field, troubles with mapping a sphere onto the focal plane cause image degradation of the order of 0.15 arcsecond, which with 0.5 arcsecond images is negligible, but the errors grow like the square of the field diameter. The desired map (onto a flat focal plane) is one which creates a Mercator-like projection of the sky onto the focal plane, with both parallels of `latitude' (i.e. curves of constant angular separation from a great circle through the center of the field in the scan direction) and meridians mapped into straight lines, but no such map is axially symmetric, and anamorphic optics, which would be required for such a map, have not been investigated. The errors for a zero-distortion design are in any case not excessive with our field. Conventional Ritchey designs have two orders of magnitude too much distortion for this application.

In addition, conventional Ritchey designs with Gascoigne correctors this fast have unacceptably large lateral color, both for this imaging application and (especially) for the fiber spectroscopy. We were therefore compelled to go to somewhat more complex systems, and have evolved a design with a two-element refracting corrector which has excellent performance. It makes use of the fact that the astigmatism correction of a Gascoigne plate goes as the square of the distance from the focal plane for a given strength, while the lateral color and distortion only go linearly. Thus a pair of plates, one of the usual form and of weak power placed some distance from the focal plane, and another the negative of the usual form of n times the strength of the first placed 1/ n^{th} the distance of the first from the focus, can correct astigmatism while introducing no lateral color or distortion. Distortion remains at a level (12 microns over the field of the camera) set by the order of the aspheric used for the second corrector element and can in principle be removed (or specified) exactly; the lateral color is less than 10 microns peak-to-peak over the whole field over the spectral range of the spectrograph, and is a negligible contribution to the image diameter for any filter or field location in the camera.

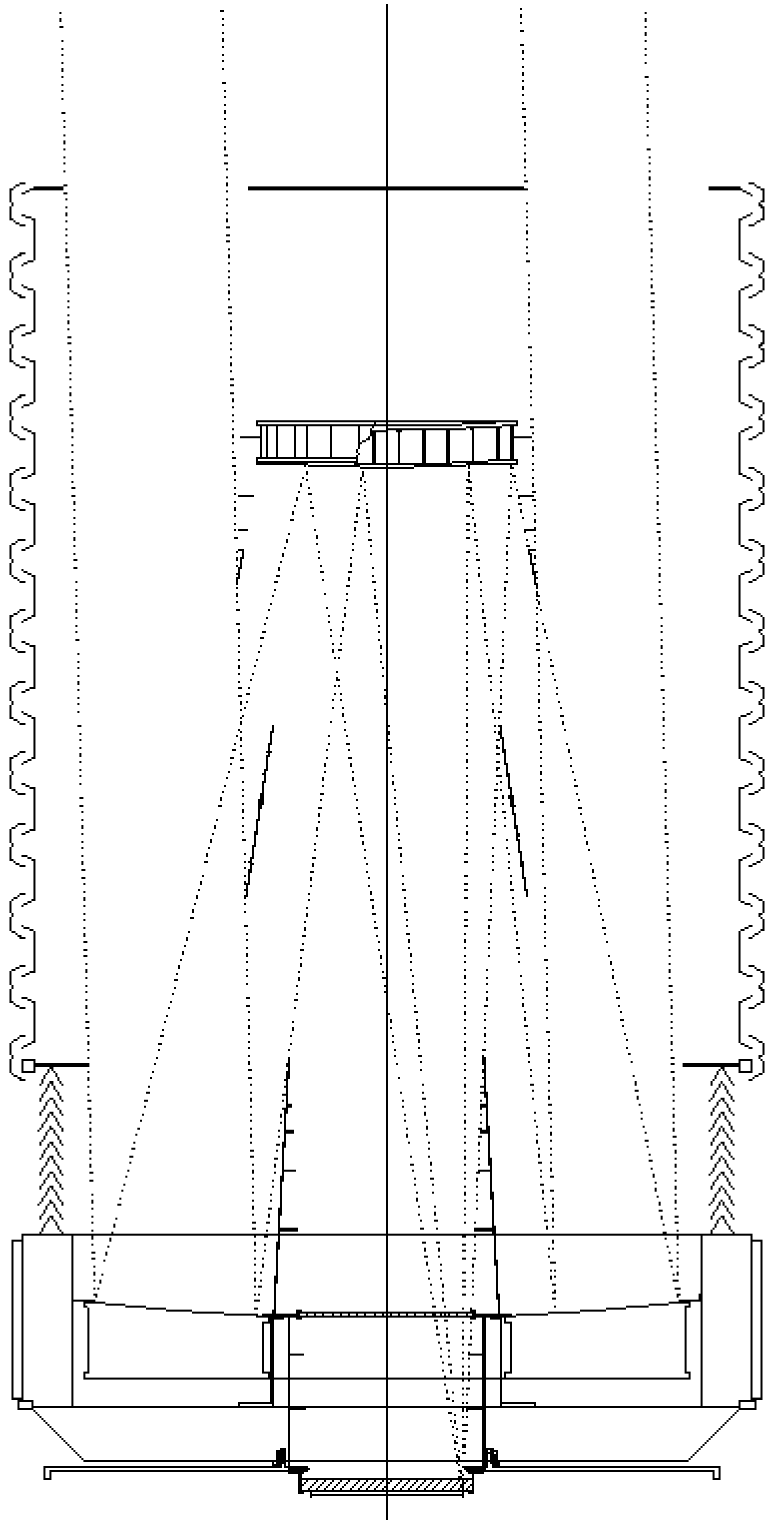
The design presented here has almost zero distortion in the sense that the radius in the focal plane is proportional, to high accuracy, to the field angle (not its sine or tangent); zero distortion for most wide-field imaging is defined for the condition that the radius in the focal plane is proportional to the *tangent* of that angle, which results in faithful representations of figures on *planes*, but we wish as faithfully as possible to image figures on a sphere onto a surface which is almost planar. For this case a compromise is necessary between the wishes for constant scale in the sense that meridians have constant linear separation in the focal plane, and the desire that parallels of latitude do likewise. The optimum case depends somewhat on the aspect ratio of the field and is somewhere between the radius in the focal plane going like the sine of the input angle and its tangent. For a square focal plane, which is close to the situation at hand, the radius approximately proportional to the angle itself is best, and we have made this choice. The errors can be minimized by clocking different chips at different rates to correspond to the local scale along the columns, but we have chosen for reasons of noise reduction and simplicity in the data system to clock all CCDs synchronously. Our design results for the best compromise tracking rate in worst-case image smearing along the columns of 0.06 arcseconds, 3 microns, or 0.14 pixels over the imaging array. Stars do not quite follow straight trajectories in the focal plane, but this is compensated for by a slight rotation of the outer chips, amounting to about .006 degrees at the corners, and causes an error of only about 0.24 pixel if uncompensated.

The design uses a 2.5-meter f/2.25 primary with a 1.08-meter secondary, which with its baffles (1.30-meter dia.) obscures 27% of the incoming beam. The central hole in the primary is 1.17 meter in diameter. Cassegrain telescopes with fields this large are notoriously difficult to baffle, but a variant on the "Venetian blind" baffling system used in the DuPont design will work well here. The optical layout and the current baffle design is shown in Figure 2.1; the rays in that figure are for an object at the extreme edge of the 3deg field. The output f/ratio is 5.0, and the focal plane is 0.76 meters behind the vertex of the primary in order to clear the cell and allow room for the instruments. The telescope is quite short-coupled, with the secondary only 3.6 meters in front of the primary. The corrector consists of two aspheric fused quartz elements, as discussed above. The first (`Gascoigne') element is approximately coincident with the vertex of the primary mirror; there will be two interchangeable rear correctors, a very thick one associated with the camera (and in fact an intimate part of its mechanical design -- see Chapter 4), and a much thinner one for use with the spectrograph. The top surface of the second element in the camera configuration is just 58 mm from the focus, and for the camera element the back surface of the filters, which are cemented to the corrector, is 8 mm from the focus. The spectrographic configuration has 34 mm of working space behind the second corrector. The 3° field is 0.65 meters in diameter; over this field, the focal surface of the camera configuration is described to reasonable accuracy as a simple quartic in the radius; it is flat to within about +- 0.2 mm over the inner 0.5 m, and rises rapidly about 2 mm at the edge.

The CCDs for the imager will be mounted to conform to the focal surface, which requires a tilt of just under a degree at the edge of the field. There is one further complication in the design, namely that the CCDs as produced are slightly convex, with a reasonably well controlled radius of about 2.2 meters. The best fit plane results in focus errors of about 100 μ RMS, which at f/5 corresponds to an image degradation of about 20 μ. We have chosen *not* to live with this, but instead to correct this curvature individually for each chip with weak field flatteners cemented to the rear face of the corrector. This is more to attempt to keep the point spread functions reasonably constant over a chip than a fundamental discomfort with global focus errors this large, but we shall see that for some chips in the array there are unavoidable significant variations anyway.

The spectroscopic optical configuration is similar; the two share the primary, secondary, and Gascoigne corrector, but the last corrector is substantially different. It is much thinner (which substantially reduces longitudinal color), quite strongly curved, and a bit farther from the focal plane than the one for the camera. Its design was optimized for lateral color, which is better than 4 microns RMS over the whole field while maintaining polychromatic images better than 1 arcsecond RMS diameter. The primary-secondary spacing is also slightly different; the difference is well within the secondary focus range. The spectroscopic design violates the `telecentric' condition that the focal plane be perpendicular to the central ray in each image. This means that the fibers are not placed perpendicular to the focal plane. We will use drilled plug-plates for the fibers, and the most straightforward way to deal with the lack of telecentricity, as we shall see, is to drill the plug-plates for the fibers while the plates are deformed slightly. This will be discussed further in Section 7.6.0.1.

Figure 2.1



The baffles and optical layout of the SDSS telescope. The outer interlocking "C"-shaped baffles which form the upper tube are carried by an independently mounted and driven wind baffle mechanism and take the major wind loads on the telescope. The corrector system are the two thin plates close to the focal plane. The rays shown are from the edge of the 3° field.

Table 2.1a: The Optical Design for the SDSS Telescope, Camera Mode

sur	**c**	**s**	*glass*	a_2	a_4	a_6	a_8	**k**	*clr dia*
1	-8.889e-5	0.0	-air	0.0	0.0	3.81e-22	-1.52e-29	-1.285	2500

2	-1.390e-4	-3646.14	air	0.0	0.0	1.79e-19	0.0	-11.97	1080
3	0.0	3621.59	fq	2.321e-5	-1.173e-10	-7.87e-17	1.59e-22	0.0	722
4	0.0	12.0	air	0.0	0.0	0.0	0.0	0.0	721
5	0.0	714.00	fq	-2.732e-4	2.056e-9	-5.81e-15	1.75e-20	0.0	657
6	0.0	45.00	bk7	0.0	0.0	0.0	0.0	0.0	652
7	0.0	5.00	air	0.0	0.0	0.0	0.0	0.0	651
8	0.0	8.00	air	0.0	0.0	0.0	0.0	0.0	651

Table 2.1b: The Optical Design for the SDSS Telescope, Spectroscopic Mode

sur	**c**	**s**	***glass***	**a_2**	**a_4**	**a_6**	**a_8**	**k**	***clr dia***
1	-8.889e-5	0.0	-air	0.0	0.0	3.81e-22	-1.52e-29	-1.285	2500
2	-1.390e-4	-3644.46	air	0.0	0.0	1.79e-19	0.0	-11.97	1080
3	0.0	3619.91	fq	2.321e-5	-1.173e-10	-7.87e-17	1.59e-22	0.0	722
4	0.0	12.0	air	0.0	0.0	0.0	0.0	0.0	721
5	-4.307e-4	672.64	fq	0.0	0.0	0.0	0.0	0.0	657
6	0.0	10.00	air	-7.747e-5	-4.123e-10	-6.53e-15	5.23e-20	0.0	656
7	0.0	86.61	air	0.0	0.0	0.0	0.0	0.0	653

Table 2.1: Optical Design of the SDSS Telescope

In the above tables, c are the curvatures, positive if concave right. k are the conic constants ($k = 0$ is a sphere, $k = -1$ a paraboloid, $k < -1$ a hyperboloid, $-1 < k < 0$ a prolate ellipsoid, and $k > 0$ an oblate ellipsoid; generally, $k = -e^2$). s are the spacings in millimeters from the previous surface, positive if to the right. `Glass' is the material following the surface. The sign of `glass' changes for reflections and is positive for rightward-moving rays, negative for left. a_2, a_4, a_6, and a_8 are the aspheric coefficients for polynomial aspherics, where the general form of the surface is

$$t = c(h^2 + (k+1)t_c^2)/2 + (a_2 h^2 + a_4 h^4 + a_6 h^6),$$

where t_c is the solution to the conic surface equation

$$t_c = c(h^2 + (k+1)t_c^2)/2.$$

The index for fused quartz (fq) is 1.46415 at 4700 Å.

The camera design is specified in Table 2.1a, and the spectroscopic design in Table 2.1b. The primary mirror is f/2.25, and the overall system is f/5.0. In this final version, the camera design has been optimized for the chosen distribution of filters over the focal plane, and the details of the performance, and that of the spectroscopic design, are discussed below. The overall scale is 60.4 microns/arcsecond; one pixel is 0.403 arcsecond = 24 microns. There are a total of 6 glass-air surfaces, the pupil radius is 1250.00 mm with a 625 mm central obscuration, taken (not quite correctly) at the primary, and the first conjugate is at infinity.

The primary is almost hyperbolic, with about a wave and a half of 6th and 8th order flattening at the edge; the secondary is likewise, with about two waves of 6th order steepening. For interferometric tests of the primary, we used a null lens verified using a technique developed by J. Burge (Steward Observatory Mirror Lab) which uses a computer generated hologram. Interferometric testing of the secondary uses an innovative and powerful holographic test against an almost matching sphere, also conceived by Burge. The first corrector plate is straightforward, and, under contract with Contraves Inc., has been aspherized, figured and polished using a null test. The second corrector, in imaging mode, is very strong, with an aspheric sagitta of more than 8 millimeters, but the accuracy required is not very high by optical standards and the surface has been monitored with sufficient accuracy with a simple profilometer during fabrication. The contract for this second corrector plate for the camera was completed by Loomis Custom Optics with excellent results. Custom Optics also drilled the camera mounting holes (see Chapter 4) in the flat side of this corrector, and the corrector underwent application of a custom striped antireflection coating which matches the camera photometric bands. The second spectrographic corrector was aspherized, figured and polished under contract with Tinsley Laboratories.

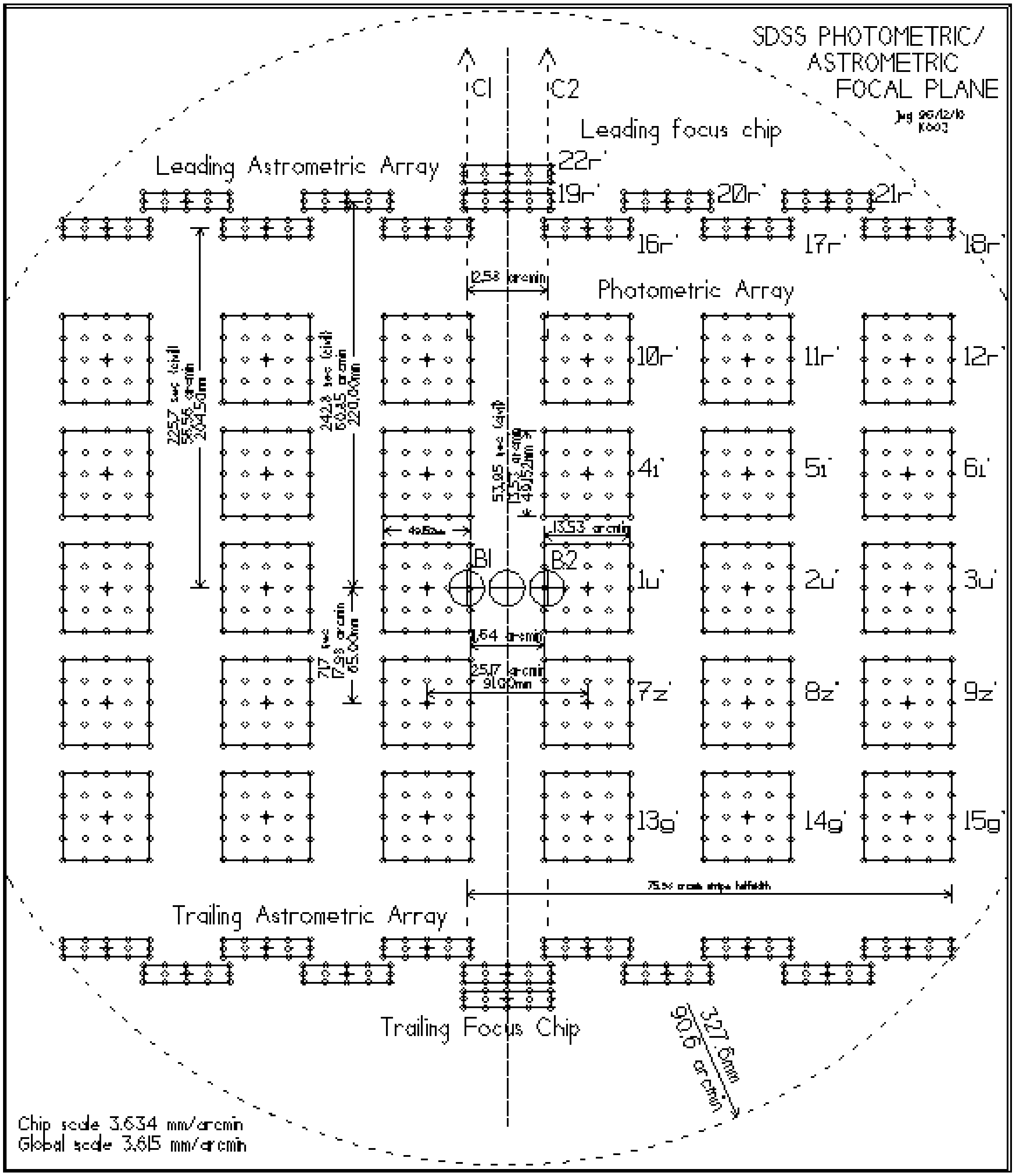
The Performance of the Imaging Design

The discussion of the optical performance of the camera configuration is a bit complicated because of the complexity of the focal plane, with different filters and field flatteners in different locations and the effect of distortion on the final TDI image quality, so we will approach a full discussion gradually. A series of simple monochromatic traces of the camera system without the individual field flatteners is presented in Table 2.2a. Here the focus (distance behind the dummy surface 8, which is the nominal 8 mm back focal distance behind the last (filter) element) f_b , the image height h at that focus, and the RMS image diameter epsilon is tabulated for each of 8 field angles from the center to the edge of the imaging field (which is somewhat smaller than the whole spectroscopic field) for the 5 colors which are the effective wavelengths for the five filters we propose. The last four field angles correspond to the outer corners of the outermost chip in some row, and angles which are not actually reached at a given color with the camera design (see Figure 2.2) are prefixed with a ` * '. The images are degraded somewhat at the very edge, where the radial field curvature is maximum, by the finite (flat) area of the chips, but this effect will be evaluated in detail along with TDI and polychromatic effects below.

The form of the focal surface at 4760Å is presented next, in Table 2.2b. delta is the total longitudinal focal deviation from a plane at the indicated angle, ht the height in the focal surface, lindev the deviation from a best-fit straight line relating the input angle to the height, and lin ht that linear relation. The errors (maximum of about 4 μ) simply reflect the order of the fit; it is clear that the distortion is controlled by the local slope of this surface, and can be made to vanish (or take on any reasonable form) exactly. The corrector is close enough to the focus that there is little repercussion on the image quality or the lateral color when it is modified slightly.

To discuss the quality of the images as accurately as possible requires a much more detailed analysis, which we summarize briefly here. The optical layout of the camera focal plane is sketched in Figure 2.2, which shows the locations of the 30 2048x2048 "photometric" CCDs, the 22 2048x400 "astrometric" chips, and the two 2048x400 focus sensors. The filters u' ($\text{lambda}_{\text{eff}}$ = 3540 Å), g' ($\text{lambda}_{\text{eff}}$ = 4760 Å), r' ($\text{lambda}_{\text{eff}}$ = 6280 Å), i' ($\text{lambda}_{\text{eff}}$ = 7690 Å), and z' ($\text{lambda}_{\text{eff}}$ = 9250 Å), with which each field is identified, are fully described in Chapter 4. (The camera is right-left reflection symmetric and the lower astrometric/focus array is the mirror image of the upper array). The direction of the TDI scan is upwards in this diagram, so a given star first encounters an astrometric device, then an r' chip, then a i' chip, and so on until, 485 seconds later, it encounters the final astrometric chip. The large circle is the full 90.4 arcminute radius field. Each small circle within the CCD outlines is a point for which the system has been traced with five wavelengths chosen such that each is the mean wavelength of the corresponding quintile of the filter response; thus each has equal weight in the final image. Cemented to the flat back surface of the final corrector is the colored glass component of the filter and to that the field flattener, whose radius of curvature is chosen from a set of four which spans the requirements and on whose back surface is coated the short-pass part of the filter. The central thickness of the filter/flattener sandwich is constant at 5.00 mm.

Figure 2.2



The field of the SDSS Camera. The photometric CCDs are arranged in 6 identical columns of 5 2048x 2048 chips, each with one of the five primary filter bands of the survey. Leading and trailing these columns are arrays of 12 2048x 400 chips for astrometric calibration and focus. The fields discussed in the text are identified by number and their corresponding filter. One stripe of the survey is composed of two TDI scans, one centered on the column labeled C1 through the boresight B1, the other on the column labeled C2 through B2.

The final images (five per CCD) are composed of the five individual monochromatic images and, because TDI integrates along a column, of the five images along a CCD column, taking account in the first instance of any lateral color shifts and in the second of any residual distortion perpendicular to the column and any residual distortion and scale error along the column. The images are defocused to lie in the best-fitting focal surface with the mean curvature of the CCDs for each subfield (tilt and piston are fitted). The input angles along the column accurately represent images at successive equally spaced time intervals in TDI mode, and the geometry on the sky for TDI is accurately modeled. The final images are shown in Figure 2.3. Here each row of images is the model of TDI output for the array as labeled in Figure 2.2; thus the bottom row consists of five images each across fields 13, 14, and 15, the next 7, 8, and 9, and so forth. The top two rows are the astrometric fields 16, 17, 18, and 19, 20, 21. The images for the focus array are discussed and shown later. The spacing between successive closely spaced images in the mosaic is 3 arcseconds. The bottom panel shows the images as delivered by the design optical system; the top as convolved with 0.8 arcsecond Gaussian seeing, which we believe will be the worst seeing we will have to cope with in the photometric part of the survey (see Chapter 1). The PSFs were generated by fitting discrete Zernike polynomials to the slope errors in the system and using those fits to generate intercepts in the desired focal plane for 1200 rays for each of the 25 images which go into the polychromatic TDI composite. Those rays were simply binned in 0.05 arcsecond pixels to generate the intensities for the greyscale images.

The situation is summarized quantitatively in Table 2.2c, where each row lists the properties of one field (detector). The identifying numbers are as in Figure 2.2. The table lists for each field the field center in millimeters measured from the optical axis (-y is the TDI scan direction) the size of the CCD for that field, the filter, the field flattener curvature in units of 10^{-3}mm^{-1} (ffc3), the CCD curvature in units of 10^{-4}mm^{-1} (ccd4), the vertical scale in that field (mm/arcmin), the RMS focus error in microns over the CCD caused by mismatch between the final best focal surface and the curved CCD surface, the residual field curvature in units of 10^{-4} mm^{-1} (dc4), and the minimum (em) and maximum (eM) RMS image diameters over the field in microns. We should perhaps comment on the residual curvature; the overall scale is as listed in Table 2.2b, 3.623mm/arcmin, but the field flatteners change the scale locally for each chip to a number close to 3.6431, which was the "scan scale" -- i.e. the assumed tracking rate. Changes in this scale from chip to chip, and color to color, represent errors in the TDI images, and the field flattener curvatures are chosen for the best compromise between keeping the scale constant and matching the focal curvature. Scale errors are in general much more serious for image quality than focal errors, so there is usually some residual curvature.

The results indicate that for the *photometric* array, the maximum RMS image diameters are for the ultraviolet fields reaching 0.68 arcseconds for the outermost one. The increase from 27 to 41 microns from the monochromatic to the full TDI polychromatic treatment is mostly due to longitudinal color, with tiny contributions from defocus, lateral color, and substantial ones from TDI effects. Images almost as large are seen at the field extremes at the other end of the spectrum in z' , where they reach 39 microns, 0.65 arcsec RMS. The other images are of order 0.5 arcsec or better over the whole field. The problems in z' are just due to the extreme wavelength; the optimization of the system involves balancing the color effects at the wavelength extremes, and because the polychromatic effects at u' are so large the optimization was `tilted' toward the ultraviolet.

The images for the *astrometric* chips are almost as good except for the outer half of field 18, the outermost of the second rank of CCDs, where the images reach two-thirds of an arcsecond in RMS diameter.

Recall for all of these results that for a roughly gaussian image, the RMS diameter is 1.20 times the FWHM, so by that measure the images are correspondingly better.

The images for the focus chips (field 22) are quite good, about 0.38 arcsecond RMS diameter, and with almost no variation over the field, so even though the focus sensors (see Chapter 4 for a fuller discussion of the focusing scheme and illustrations of the focus chip images through focus) are near the outer edge of the field, there is no compromise in image quality for them. The fact that the dominant residual aberration at this point is third-order astigmatism means that the out-of-focus images are decidedly elliptical and this can be used to refine the focus servo.

Summary of Telescope Optical Performance for the Imaging Mode (k003)

Table 2.2a: Parameters for best focus at each color at each radius

Angle	lambda	f_b	h	epsilon
arcmin	Å	mm	mm	mm
0.00	4760	-0.361	-0.000	0.018
30.00	4760	-0.088	-108.452	0.015
45.00	4760	0.052	-162.685	0.013
60.00	4760	-0.084	-216.913	0.010
70.00	4760	-0.458	-253.069	0.011
73.00	4760	-0.632	-263.918	0.011
82.00	4760	-1.380	-296.474	0.015
*90.00	4760	-2.399	-325.431	0.021
0.00	3540	-0.322	-0.000	0.017
30.00	3540	0.031	-108.435	0.014
45.00	3540	0.267	-162.666	0.013
60.00	3540	0.260	-216.895	0.019
70.00	3540	-0.014	-253.053	0.027
*73.00	3540	-0.156	-263.902	0.030
*82.00	3540	-0.803	-296.461	0.037
*90.00	3540	-1.731	-325.417	0.043
0.00	6280	-0.384	-0.000	0.018
30.00	6280	-0.154	-108.461	0.017
45.00	6280	-0.065	-162.695	0.017
60.00	6280	-0.268	-216.922	0.019
70.00	6280	-0.696	-253.077	0.022
73.00	6280	-0.887	-263.925	0.023
82.00	6280	-1.687	-296.481	0.029
*90.00	6280	-2.756	-325.438	0.037
0.00	7690	-0.396	-0.000	0.018
30.00	7690	-0.187	-108.465	0.018
45.00	7690	-0.124	-162.700	0.020
60.00	7690	-0.361	-216.927	0.024
70.00	7690	-0.815	-253.081	0.029
73.00	7690	-1.015	-263.929	0.031
*82.00	7690	-1.841	-296.485	0.038
*90.00	7690	-2.934	-325.442	0.048
0.00	9250	-0.405	-0.000	0.018
30.00	9250	-0.212	-108.468	0.019
45.00	9250	-0.168	-162.704	0.022
60.00	9250	-0.431	-216.931	0.028

70.00	9250	-0.905	-253.084	0.034
73.00	9250	-1.111	-263.932	0.036
*82.00	9250	-1.957	-296.488	0.045
*90.00	9250	-3.069	-325.444	0.057

* These field angles are not reached in these colors for the photometric array.

Table 2.2b: Focal surface, lambda=4760Å, scale=3.61519 mm/arcmin

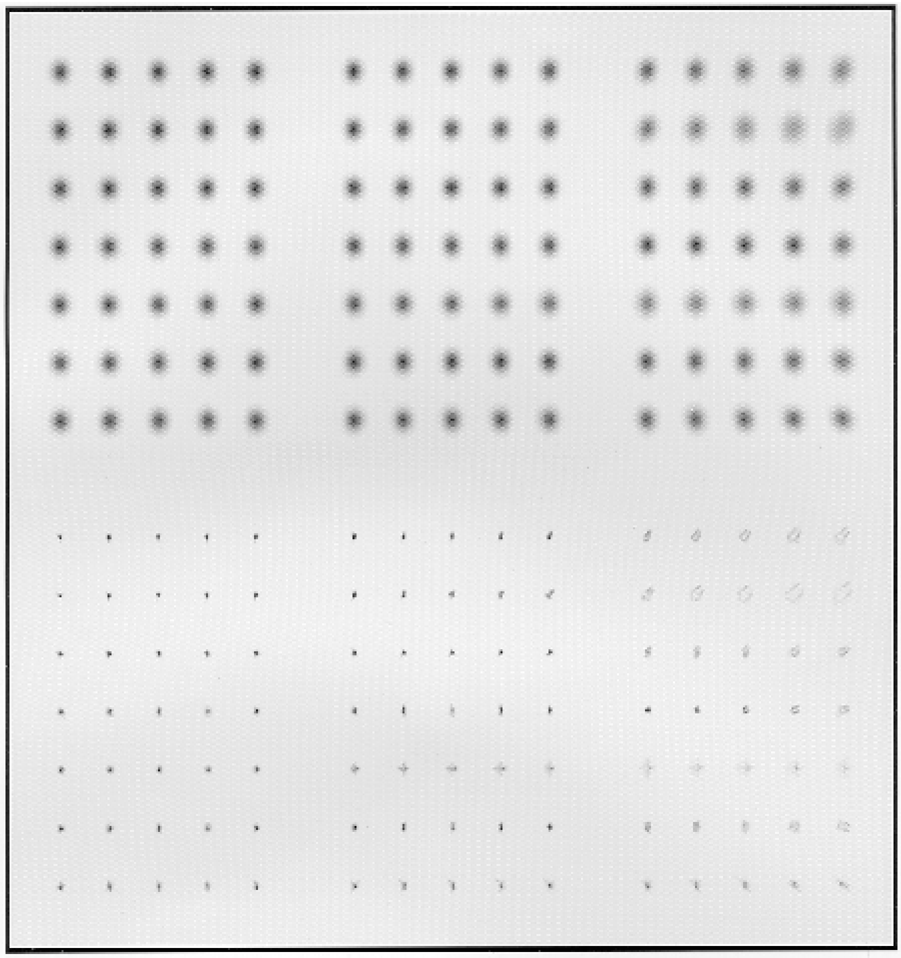
Angle	**delta**	**ht**	**lindev**
arcmin	**mm**	**mm**	**mm**
0.0	-0.361	-0.000	-0.000
30.0	-0.088	-108.452	0.005
45.0	0.052	-162.685	0.003
60.0	-0.084	-216.913	0.009
70.0	-0.458	-253.069	0.012
73.0	-0.632	-263.918	0.011
82.0	-1.380	-296.474	-0.001
90.0	-2.399	-325.431	-0.027

Table 2.2c: Summary of TDI Images, Scan Scale=3.6343 mm/arcmin

fld	**x**	**y**	**size**	**fil**	**ffc3**	**ccd4**	**vscl**	**dfoc**	**dc4**	**em**	**eM**
1	45.5	0.0	49.2x 49.2	u'	1.32	4.3	-3.6311	3	0.2	17	40
2	136.5	0.0	49.2x 49.2	u'	1.32	4.3	-3.6326	10	-0.5	19	29
3	227.5	0.0	49.2x 49.2	u'	1.32	4.3	-3.6329	30	-1.5	30	41
4	45.5	65.0	49.2x 49.2	i'	1.40	4.3	-3.6350	4	0.2	18	21
5	136.5	65.0	49.2x 49.2	i'	1.40	4.3	-3.6353	13	-0.4	20	24
6	227.5	65.0	49.2x 49.2	i'	1.40	4.3	-3.6339	30	-1.3	25	35
7	45.5	-65.0	49.2x 49.2	z'	1.40	4.3	-3.6350	4	0.1	18	21
8	136.5	-65.0	49.2x 49.2	z'	1.40	4.3	-3.6351	13	-0.5	20	26
9	227.5	-65.0	49.2x 49.2	z'	1.50	4.3	-3.6348	26	-1.1	28	39
10	45.5	130.0	49.2x 49.2	r'	1.30	4.3	-3.6342	12	-0.5	18	19
11	136.5	130.0	49.2x 49.2	r'	1.30	4.3	-3.6338	25	-1.0	20	23
12	227.5	130.0	49.2x 49.2	r'	1.50	4.3	-3.6348	37	-1.4	27	32
13	45.5	-130.0	49.2x 49.2	g'	1.30	4.3	-3.6347	11	-0.4	19	21
14	136.5	-130.0	49.2x 49.2	g'	1.30	4.3	-3.6344	24	-0.9	21	23
15	227.5	-130.0	49.2x 49.2	g'	1.40	4.3	-3.6345	40	-1.6	27	32
16	45.5	204.5	49.2x 9.6	r'	1.50	4.3	-3.6351	2	-0.1	19	21
17	136.5	204.5	49.2x 9.6	r'	1.50	4.3	-3.6347	10	-0.8	22	26
18	227.5	204.5	49.2x 9.6	r'	1.80	4.3	-3.6358	20	-1.4	30	40
19	0.0	220.0	49.2x 9.6	r'	1.50	4.3	-3.6350	1	-0.0	20	21
20	91.0	220.0	49.2x 9.6	r'	1.50	4.3	-3.6349	6	-0.4	21	24

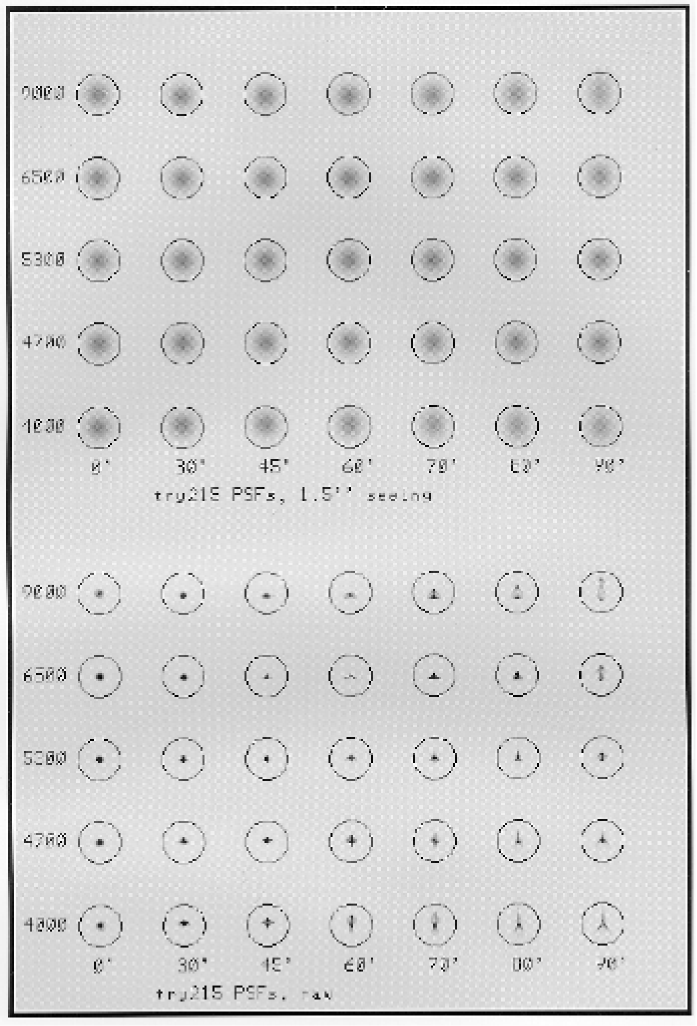
21	182.0	220.0	49.2x 9.6	r'	1.60	4.3	-3.6351	17	-1.2	26	33
22	0.0	235.5	49.2x 9.6	r'	1.50	4.3	-3.6350	1	-0.2	22	23

Figure 2.3



PSFs for the SDSS camera. These are shown as produced by the optical system alone (bottom) and convolved with 0.8 arcsecond Gaussian seeing (top). Each row of five closely spaced images are from one field, with the fields from one row of CCDs arranged side-by-side. Thus the bottom three sets are from fields 13,14, and 15, and the top three, 19, 20, and 21. (See Figure 2.2). The images in a set are spaced 3 arcseconds apart.

Figure 2.4



Spectroscopic PSFs for the SDSS telescope. These are as seen in the spectroscopic configuration on the mean focal surface. The field angles are (from left to right) 0, 30, 45, 60, 70, 80, and 90 arcminutes, and the wavelengths from top to bottom 4000, 4700, 5300, 6500, and 9000 Å.

Summary of Telescope Optical Performance for the Spectrographic Mode (kmg001)

Table 2.3a: Average focus for all colors at each radius (fibers)

Angle	**lambda**	**f_b**	**h/dh**	**D**	**epsilon**
arcmin	**A**	**mm**	**mm**	**mm**	**mm**
0.00	5300	-0.007	0.000	0.000	0.029
30.00	5300	-0.143	-108.818	0.000	0.026
45.00	5300	-0.424	-163.322	0.000	0.024
60.00	5300	-0.978	-217.855	0.000	0.025
70.00	5300	-1.536	-254.241	0.000	0.027
80.00	5300	-2.265	-290.713	0.000	0.026
90.00	5300	-3.203	-327.372	0.000	0.025
0.00	4000	-0.007	0.000	-0.135	0.036
30.00	4000	-0.143	0.004	-0.081	0.030

45.00	4000	-0.424	0.005	-0.015	0.025
60.00	4000	-0.978	0.005	0.076	0.028
70.00	4000	-1.536	0.004	0.148	0.036
80.00	4000	-2.265	0.002	0.231	0.049
90.00	4000	-3.203	-0.004	0.325	0.065
0.00	9000	-0.007	0.000	0.131	0.036
30.00	9000	-0.143	-0.004	0.078	0.029
45.00	9000	-0.424	-0.004	0.014	0.026
60.00	9000	-0.978	-0.004	-0.074	0.036
70.00	9000	-1.536	-0.004	-0.145	0.046
80.00	9000	-2.265	-0.002	-0.226	0.056
90.00	9000	-3.203	0.003	-0.317	0.068
0.00	4600	-0.007	-0.000	-0.058	0.031
30.00	4600	-0.143	0.002	-0.035	0.027
45.00	4600	-0.424	0.002	-0.006	0.024
60.00	4600	-0.978	0.002	0.033	0.025
70.00	4600	-1.536	0.002	0.065	0.027
80.00	4600	-2.265	0.001	0.101	0.030
90.00	4600	-3.203	-0.001	0.141	0.035
0.00	6500	-0.007	-0.000	0.062	0.031
30.00	6500	-0.143	-0.002	0.037	0.027
45.00	6500	-0.424	-0.002	0.007	0.024
60.00	6500	-0.978	-0.002	-0.035	0.029
70.00	6500	-1.536	-0.002	-0.068	0.034
80.00	6500	-2.265	-0.001	-0.106	0.036
90.00	6500	-3.203	0.002	-0.149	0.040

Table 2.3b: Average focal surface: scale=3.62730
mean exit pupil at -5174 mm

Angle	**Focus**	**ht**	**hlindev**	**yp**	**dyp**
arcmin	**mm**	**mm**	**mm**	**rad**	**rad**
0.0	-0.006	0.000	0.000	0.0000	0.0000
30.0	-0.143	-108.818	0.184	-0.0280	-0.0247
45.0	-0.423	-163.322	0.185	-0.0396	-0.0324
60.0	-0.978	-217.855	0.158	-0.0477	-0.0347
70.0	-1.536	-254.241	0.113	-0.0508	-0.0329
80.0	-2.265	-290.713	-0.017	-0.0523	-0.0293
90.0	-3.203	-327.372	-0.331	-0.0538	-0.0258

The Performance of the Spectrographic Design

In Table 2.3 data are presented which are relevant to the spectrographic mode. At each of seven field angles from the center to the extreme edge, the focal properties are given on a surface which represents the average focal surface over the spectrograph wavelength range of 3900 Å to 9200 Å (the spectrograph is described fully in Chapter 7). The first set of entries is for 5300 Å, which roughly centers the range of index variations for the spectrograph. The height differences (in the sense of the height at the given wavelength minus the height at 5300 Å, which are tabulated after the 5300 Å entries), thus represent lateral color, which is seen to be 10 microns total, +- 5 microns, or less over the whole field, and are even somewhat smaller at the edges where the images are larger. The D s are longitudinal deviations from best focus at that wavelength, and the epsilon s are RMS image diameters at the compromise focus. It is only at the wavelength extremes and at the extreme edge of the field that the RMS diameters of the compromise images exceed 1 arcsecond; the best focus images there are substantially less than 1 arcsecond (45 and 40 microns at 4000Å and 9000Å, respectively, at 90 arcminutes radius), and the increase is due solely to longitudinal color. The 72 micron worst-case RMS diameter is still much smaller than the 180 micron fibers, however, and the effect on throughput is not large; we discuss the issue more fully below. The details of the average focal surface are presented next: the sagitta of the focal surface, the mean height (here just the average of the 4000 Å height and the 9000 Å one, and presumably where one will drill the fiber hole), the deviation from a linear relation with the field angle (it is seen here that the different final corrector form, chosen to yield the best polychromatic images, results in quite serious distortion, but this is of no importance for the spectrograph), the direction cosine of the central ray measured from the direction of the axis, and the *difference* between this angle and the angle which the normal to the focal surface makes with the axis. This last entry is the angle with which the fiber hole must be drilled into a plate which conforms to the focal surface. The maximum value is about 2.0 degrees, compared to the 5.7 degree half-angle input cone at f/5. The losses, even into the f/4 input beam of the spectrograph, are large enough to be important, and we will compensate for it by drilling the holes into a deformed plate, as will be discussed in detail in Chapter 7.

As will be discussed further in Chapter 1, we expect to use the best-seeing time for imaging and do fiber spectroscopy under less-good conditions; average seeing for the latter will probably be in the range 1.2-1.5 arcseconds. Differential refraction at the ends of the spectrum at the maximum zenith angle (55 degrees) is just under +- 1 arcsecond from the central wavelength image at the altitude of the site. With 1.5-arcsecond Gaussian seeing, a 3-arcsecond fiber at the field edge collects 95% and 92% of the light at 4000 Å and 9000 Å, respectively, when centered on the image, and at worst 65% and 67% when decentered by 1 arcsecond. This is not substantially worse than the situation in the center of the field at the central wavelength, where the centered number is 98% and the 1-arcsecond offset number is 72%. For extended objects, we collect a smaller fraction of the light, of course, but the *differential* between the center and edge is smaller.

Greyscale PSFs for the field angles in the table are shown in Figure 2.4, and convolved with 1.5 arcsecond Gaussian seeing; the images are 6 arcseconds apart in the mosaic, and the circles are three arcseconds in diameter, the input diameter of the fibers.

Mechanical Design and Performance

General Considerations

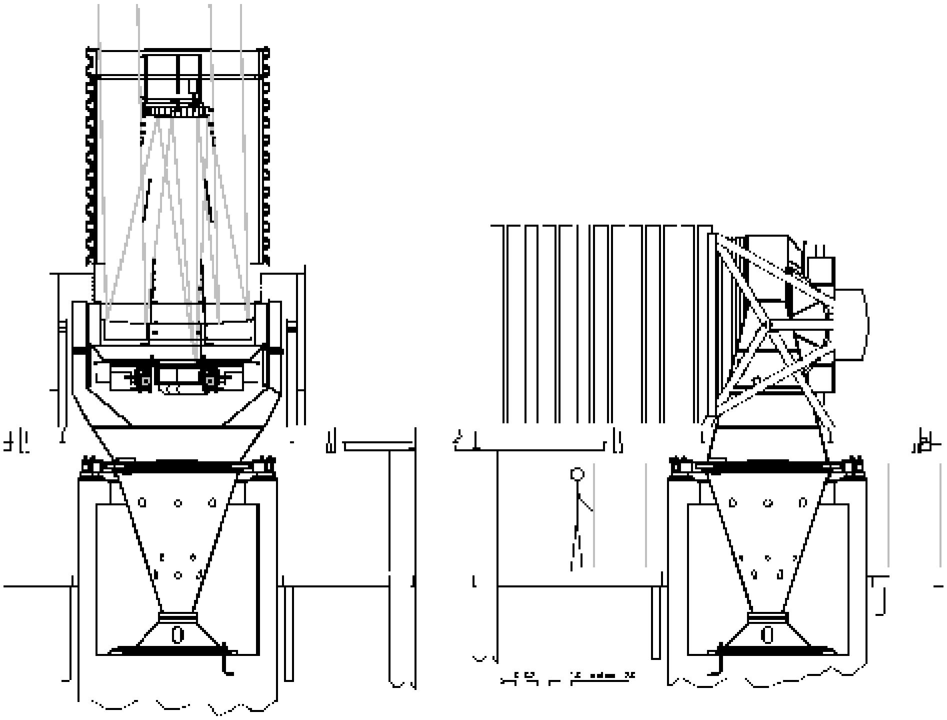
The telescope is an altitude-azimuth design similar to the Apache Point Observatory (APO) and Wisconsin, Indiana, Yale, NOAO (WIYN) 3.5 m telescopes. (Mannery et al. 1986a,b, Gunnels 1990a, Johns and Pilachowski 1990). This design takes full advantage of lightweight mirror technology resulting in a telescope with low inertia, low friction, and mechanical simplicity.

We have elected both for cost reasons and for thermal performance to use a roll-away enclosure. The telescope is protected from the wind and stray light by means of an independently mounted and driven baffle which is coaxial with and encloses the telescope.

Structure

The telescope optics support structure (OSS) consists of the primary support structure (PSS) and the secondary truss (Figure 2.5). The PSS is a steel weldment that supports the primary mirror and couples the OSS to the fork. The one-piece construction of the PSS has a higher stiffness to weight ratio and is lower in cost than the more traditional detachable mirror cell. The secondary space truss controls five of the degrees of freedom of the secondary mirror directly. With adequate tension in the secondary vanes, the rotation mode of the secondary about its optical axis can be kept above 10 Hz. The square secondary frame is efficient at resisting this tension.

Figure 2.5



Two views of the 2.5-meter telescope. The mechanical design is essentially a scaled version of the WIYN 3.5-m instrument. The wind baffle and the light baffles are shown in relation to rays from the edge of the 3° field of view.

The eight metering elements of the secondary truss are graphite fiber reinforced epoxy tubes. This material has about 2.3 times the stiffness to mass ratio of steel. This confers the following benefits:

- A substantial amount of mass is removed from the truss without degrading its static deflection or lowest natural frequency.
- The reduced moment of inertia reduces the susceptibility of the telescope to wind-induced tracking errors.
- The diameters of the truss elements are reduced without lowering their natural frequency. This, in turn, decreases the wind loads on the truss.
- The reduced mass of the truss moves the center of gravity of the OSS forward. This allows the altitude bearings to be located lower on the OSS and increased the clearance for instruments mounted behind the PSS.

Tubes and other linear structural shapes of graphite fiber reinforced epoxy have a much lower coefficient of thermal expansion than steel in the long direction. Thus, another benefit is improved metering of the primary/secondary separation with temperature changes.

Bearings and Drives

The moving mass of the SDSS 2.5-m telescope is 15,500 kg, which is light enough that exotic bearing technology is hardly required. We have chosen to use precision rolling-element bearings and friction drives throughout. These elements have all been manufactured and installed at the site and we have reasonably reliable measurements of the errors associated with their manufacture.

A pair of 2.54 meter diameter, hardened and ground drive segments are mounted on the sides of the PSS next to the fork. The measured high frequency (greater than eight cycles/revolution) runout of the drive segments is less than 100 nm RMS. Motor driven capstans, friction-coupled to each disk, provide balanced altitude drive torques and minimize windup of the PSS (Gunnels, 1990b).

The telescope azimuth structure consists of the fork and the azimuth cone. At the apex of the azimuth cone is a spherical roller bearing that supports the weight of the telescope. At the upper end of the cone is a 2.54 meter diameter disk, with a hardened and ground outer surface. Its high frequency runout is 220 nm RMS. This disk is guided by four roller assemblies, two of which are motor driven. These rollers, with the bearing at the cone apex, define the telescope azimuth axis.

Incremental encoders are friction coupled to the large disks. Readily available encoders (ROD 800, Heidenhain Corp., Elk Grove Village, Illinois) with a reduction ratio of 25:1 give 3.6 milliarcsecond resolution on the sky and allow slew rates higher than 4 °/s. Absolute axis encoding is provided by magnesensors (Sony Corporation, Park Ridge, New Jersey). These devices detect the position of magnets mounted to the drive disks and generate a signal repeatable to about 1 micron or better than 0.2" on the sky. Each axis is controlled by a third order dc position control system (Schier, 1990).

We have measured the dynamic performance of the APO 3.5-m telescope and find that the telescope is quite stiff with locked rotor resonance frequencies (the natural frequency of the telescope about an axis with the drive motor shafts locked) of 7.8 and 11.7 Hz for the azimuth and altitude axes respectively. Scaling laws indicate that a similar 2.5-m telescope should have 40% higher frequencies.

Rolling element bearings are used for each axis. These bearings require little maintenance, are low in friction, and generate negligible heat during operation. The measured high frequency radial run-out of the spherical roller bearing used as the lower azimuth bearing for the SDSS 2.5-m telescope is less than 310 nm RMS. This corresponds to a contribution of 23 milliarcsecond to the RMS tracking error for the telescope. The high frequency radial run-out of the altitude bearings is 51 nm RMS. They are estimated to contribute less than 6 milliarcsecond RMS tracking error.

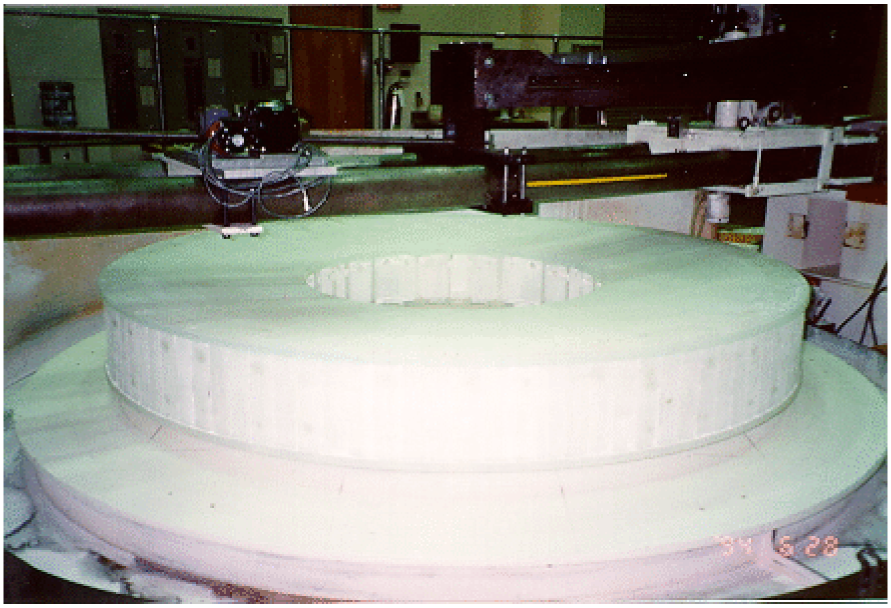
Optics: Support and Thermal control

The mirror is a borosilicate honeycomb (Figure 2.6) and was cast by Hextek Corporation (Tucson) in July 1992. The casting technique is similar to that developed at the University of Arizona Mirror Lab, except that the furnace is not rotated. The first casting attempt failed during annealing and cracks were found in the blank when the oven was opened. The causes of the failure were identified and corrected and the mirror was reheated in January. After a successful anneal, the blank was cleaned and inspected and found to be of excellent quality with low residual stresses. The Optical Sciences Center at the University of Arizona generated, figured and polished the mirror, which was delivered, aluminized, to APO in July 1996.

The primary mirror is supported on air pistons using elastomeric, low-friction rolling seals for both axial and transverse supports. Three stiff load cells serve as axial hard-points. Simple servo systems act to control the pressure provided to those air pistons in the 120° sector associated with each load cell so that the unsupported mirror weight applied to the load cell is less than 10 N. Each axial hard-point is positioned axially with a motor driven lead screw. This allows control of primary piston and tilt.

As the telescope changes elevation from zenith to horizon, the secondary will sag about 500 microns with respect to the primary optical axis. This decollimation is corrected by actively translating the secondary so that its vertex remains on the optical axis of the primary and correcting the tilt of the secondary as necessary.

Figure 2.6



Primary mirror at the Optical Sciences Center. The mirror was cast of Ohara E6 borosilicate glass by the Hextek Corporation.

We will be using aluminum plug-plates to position the optical fibers for the multifiber spectrograph. The plates will be drilled for the predicted temperature of use. However, just before use, the image scale will have to be matched to the fibers. This will be done by translating the primary axially and refocusing the secondary. A 10°C temperature mismatch can be corrected by translations of 2.4 mm and 2.0 mm for the primary and secondary respectively.

The borosilicate glass used in the primary mirror is a low thermal expansion material with an expansion coefficient of 2.8×10^{-6}/°C . To prevent thermal distortion of the mirror and mirror seeing from significantly degrading image quality, the temperature of the mirror must be uniform to 0.2°C. (This figure is larger than that established for other telescope using similar mirrors because the resolution of this wide-field optical system is moderate, though adequate to the task.)

Several active temperature control systems have been developed to perform these tasks (Johns and Pilachowski 1990; Siegmund et al. 1990; Lloyd-Hart 1990). We plan to use the rather simple system that is implemented on the ARC 3.5-m mirror. Extensive instrumentation of this mirror and balancing of the ventilation system has demonstrated that the performance of this system will be more than adequate for the 2.5-m telescope optics.

The secondary mirror is supported by three whiffletrees. Each whiffletree is positioned by an axial lead screw driven by a stepper motor via a harmonic drive. This provides control of mirror piston and tilt. The axial step size is 53 nm. This gives a resolution of 7.8 milliarcseconds on the focal plane. The lead screw/bearing assembly is specified to have an accuracy of 250 nm. The performance of these actuators on the 3.5-m telescope secondary has been measured. The differential motion across the diameter of the mirror is 310 nm RMS. This amounts to 57 mas RMS 2-dimensional motion on the sky for the 2.5-m. We expect that much smoother motions, albeit over a smaller range, can be obtained with piezo-electric actuators. These will be mounted in series with the lead screw actuators and used to correct focus during imaging. Similar drives would do nicely to control primary piston and tilt, although it is not yet clear that such high precision will be necessary for that mirror.

The secondary mirror is a borosilicate hot gas fusion blank manufactured by Hextek Corporation. The temperature control requirement to avoid degrading image quality corresponds to a uniformity of 0.2°C. The mass of the secondary is 10% of the primary, and the ventilation flow rate and heat transfer rates are reduced by the same factor. The need to bring cooling fluid to the secondary can be avoided by radiating the excess heat to the sky.

We plan to install in the primary and secondary mirrors a temperature measurement system that we have developed for the 3.5-m telescope to debug mirror temperature control systems. This system uses integrated circuit temperature sensors (Analog Devices AD590) to measure temperatures at one to two hundred locations throughout the mirror. These data are used to adjust flow rates of ventilating air to equalize thermal time constants throughout the mirror. The system, being tested on the 3.5-m mirror, has an accuracy of better than 1°C, a resolution of 5 m°C, and a calibration drift rate of 14 m°C/month RMS.

Temperature changes and, most dramatically, temperature differences between the face and back plate of a mirror, affect its power. This causes a scale change at the telescope focus and contributes to astrometric error, if not corrected in the analysis of the image data. The temperature measurement system described above will monitor changes in the power of the optics and allow the necessary corrections to be made.

The Instrument Rotator

A large diameter instrument rotator covers the back of the mirror cell. The spectrographs, camera and fiber plug-plate cartridges will mount to this rotator. Since the spectrographs will corotate with the plug-plates, this eliminates most of the flexing of the fibers that might occur during an integration and should allow for better sky subtraction. Also, as the fibers will be less than 2 meters long, additional benefits include reduction of light loss and materials cost.

The angular accuracy required for the rotator is reduced from that needed for the axes by the ratio of the telescope focal length to the field radius, a factor of 34. This degree of accuracy is quite straightforward to achieve, but is by no means a negligible task. The bearing is a Rotek four-point contact ball bearing. Its high frequency (greater than eight cycles/revolution) lateral runout was measured at about 170 nm RMS per axis at an altitude of 0°. The encoder and friction drive for the rotator are similar to those used for the axes. The drive disk for the rotator is about 2.80 meters in diameter and its high frequency runout is 1.0 μm RMS.

It is necessary, in order to control the image scale of the telescope, to monitor the distance from the focal plane to the vertex of the primary mirror very accurately (25 microns) . The load path connecting these locations is very stiff and we are likely to have excellent control of the temperature of this material as part of the primary mirror temperature control system.

Light Baffles

The roll-off enclosure for the telescope is very compact and has a low cross-section for wind loading, both of which reduce the mass and cost of the enclosure base. However, it leaves the telescope completely exposed to the wind and to light sources. These problems are addressed by the wind baffle that closely surrounds the telescope but has a separate low-precision drive system and transfers wind loads to the stationary portion of the telescope building. The wind baffle has a square cross-section that fits closely around the square secondary frame of the telescope.

The sides of the wind baffle, fabricated under contract with CVE Machining, are covered with wind-permeable panels. The panels (H. H. Robertson model 5100), with 25% equivalent open area, consist of interlocking "C" cross-section elements. Light paths through the panel require scattering from a minimum of two surfaces. Thus, with suitable coatings, the panel can be made quite light opaque.

The sky-facing end of the wind baffle contains an annular opening formed by a central disk and a panel with a circular opening (both supported by the wind baffle frame). This opening provides clearance for light from the 3° field of view to reach the telescope entrance pupil. The wind baffle blocks light rays that would otherwise have to be intercepted by the other baffles and prevents direct illumination of the primary mirror by sources more than 27° from the boresight (Figure 2.1).

The inner baffles consist of the secondary baffle (in front of the secondary), the primary baffle (extending through the primary center hole), and the conical baffle (suspended between the primary and secondary). The conical baffle is not usually present in a two mirror telescope design. It is necessary here to avoid unacceptably large central obscuration that would otherwise be the consequence of this wide field optical design. The design shown has a central obscuration of 27%. The conical baffle adds an additional 2%.

We have contracted for an analysis of the baffle design using the Advanced System Analysis Package (ASAP). Preliminary designs have been analyzed using this code and indicate quite satisfactory performance, and we expect that this design will achieve over the whole range of incidence angles greater than about 30° off axis ratios of incident flux to flux in the focal plane (PSNIT) better than 2x 10^{-6} . This translates, for example, to a scattering contribution to the night sky in the focal plane of about 26.0 mag/arsec ^{2} by a quarter moon 40 degrees off axis. This is 2% of the *dark* night sky, and less than half a percent of the moonlit night sky. The ASAP software is being used for the scattered light analysis of many other projects, including the Near IR Camera Multiobject Spectrograph (NICMOS), a second generation Hubble Space Telescope instrument.

Wind Baffle Design

The sides of the wind baffle are covered with panels that are 25% porous to the wind. Water tunnel studies indicate that with this porosity, the flow speed around the telescope secondary is reduced to about 1/3 of the free stream flow speed while the flushing time for fluid inside the baffle is still a very rapid 15 to 30 seconds. Water tunnel studies show that flow passing through the channels in the panel diffuses rapidly, i.e., on scales of 0.1 m. No spatially persistent high-velocity jets that might cause wind-induced tracking error are observed.

Centered on the telescope azimuth axis and flush with the telescope enclosure floor and the top of the telescope fork base is a motor-driven circular floor panel that follows the motions of the azimuth axis of the telescope. It supports the wind baffle altitude drive, drives it in azimuth and provides rotating floor space around the telescope for the storage of the camera when a fiber optic cartridge is on the telescope.

Traditional telescope enclosures act as cavity radiators. The net radiation imbalance with a clear sky is roughly 100 watts/m ^{2} of horizontal projected area. In a well-designed telescope enclosure with low thermal inertia and minimal heat sources, this power (kilowatts for a typical slit size) comes from conduction from the air within the telescope chamber, i.e., via the production of colder than ambient air. This cold air can cause image degradation should it enter the telescope light path.

The wind baffle reduces radiative coupling of the telescope OSS to the sky by minimizing the area of the opening at the end of the telescope. The outer surfaces of the wind baffle are covered with low emissivity aluminum foil tape. Wind baffle surfaces that are well coupled to the sky but which must be black in the visible to absorb scattered light are covered with black chrome tape, a selective emitter with high emissivity in the visible and near IR and low emissivity in the thermal IR. This solution does not work inside the wind baffle since the surface emissivity would have to be nearly zero to reduce the emissivity observed through the opening into a cavity. However, the wind baffle is small enough (unlike a conventional telescope enclosure) that it is practical to circulate fluid through pipes thermally coupled to the wind baffle panels. The temperature of the fluid will be controlled to be the same as the ambient air temperature. Approximately 350 watts of heat will be required.

The control system for each of the two axes of the wind baffle is similar to the control system for the APO 3.5-m telescope enclosure. In this system, a velocity feed-forward signal is obtained from the azimuth axis tachometer. A linear variable differential transformer (LVDT) is mounted so as to produce a signal proportional to the difference in azimuth of the enclosure and telescope. These signals control two dc motors that drive two of the four wheels that support the 100 ton rotating telescope enclosure. Each motor is driven by a separate velocity servo that is closed about a tachometer on the motor shaft. With this system it has been straightforward to control the telescope enclosure orientation to 1 milliradian or better.

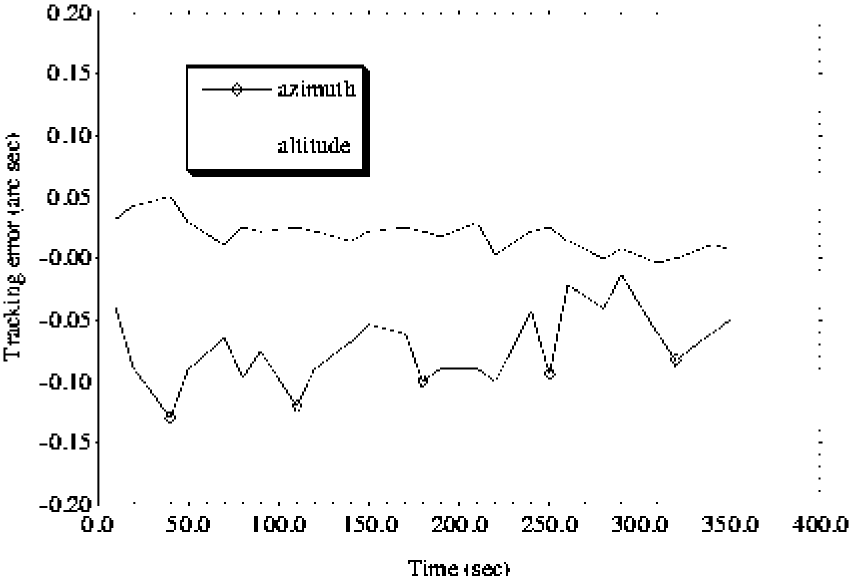
Tracking

The wind baffle reduces wind loading on the telescope OSS by a factor of ten. With care taken in the design of drive and encoding systems, bearings, and the structure, the result should be a telescope with very low wind-induced tracking error that performs at the level necessary to achieve our goals for astrometric accuracy.

Figure 2.7 demonstrates that this level of performance is feasible. These data were obtained on the ARC 3.5m from an intensified CCD camera sampling at intervals of about 13 seconds and integrating for about 1 second while the telescope was tracking open loop. Wind speed was low and the telescope was pointed southeast at an elevation of 60°. These data are uncorrected for seeing effects but are from a 5 minute interval during which the tracking was particularly smooth. The error about a linear drift is better than 100 mas peak-to-peak.

Wind-induced tracking error scales as $L^{2/3}$, where L is a typical scale length (Ulich 1988). Thus, the 2.5-m telescope, which is similar in topology to the 3.5-m telescope, should be about 20% less susceptible to wind-induced tracking error. The use of graphite fiber reinforced epoxy in the secondary truss should result in an additional improvement of similar magnitude.

Figure 2.7

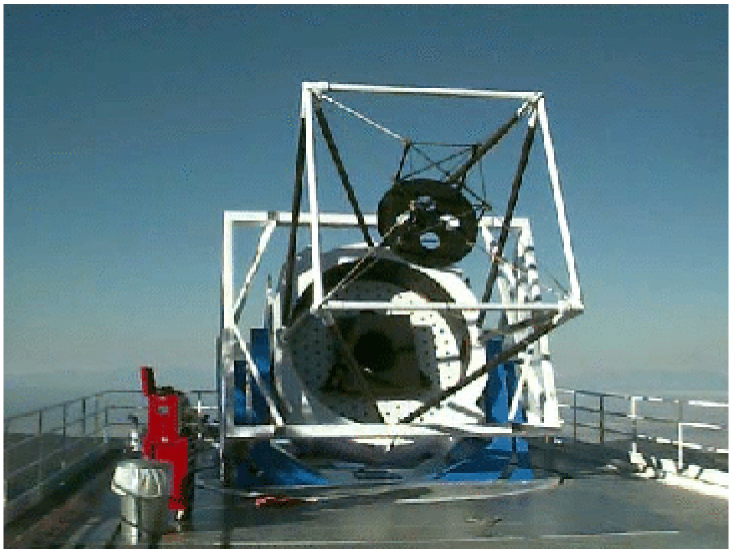


Tracking error in altitude and azimuth for the 3.5-m ARC telescope. The error about a linear drift is better than 100 mas peak-to-peak, and is not much worse about a constant.

The Enclosure

During spectroscopic observations, it will be necessary to change plug-plates cartridges about once per hour, and the plug-plate cartridges are expected to have a mass of about 100 kg. For this, and other reasons, the telescope is mounted so that we have access to it via a level track, but still have sufficient ground clearance not to degrade the seeing.

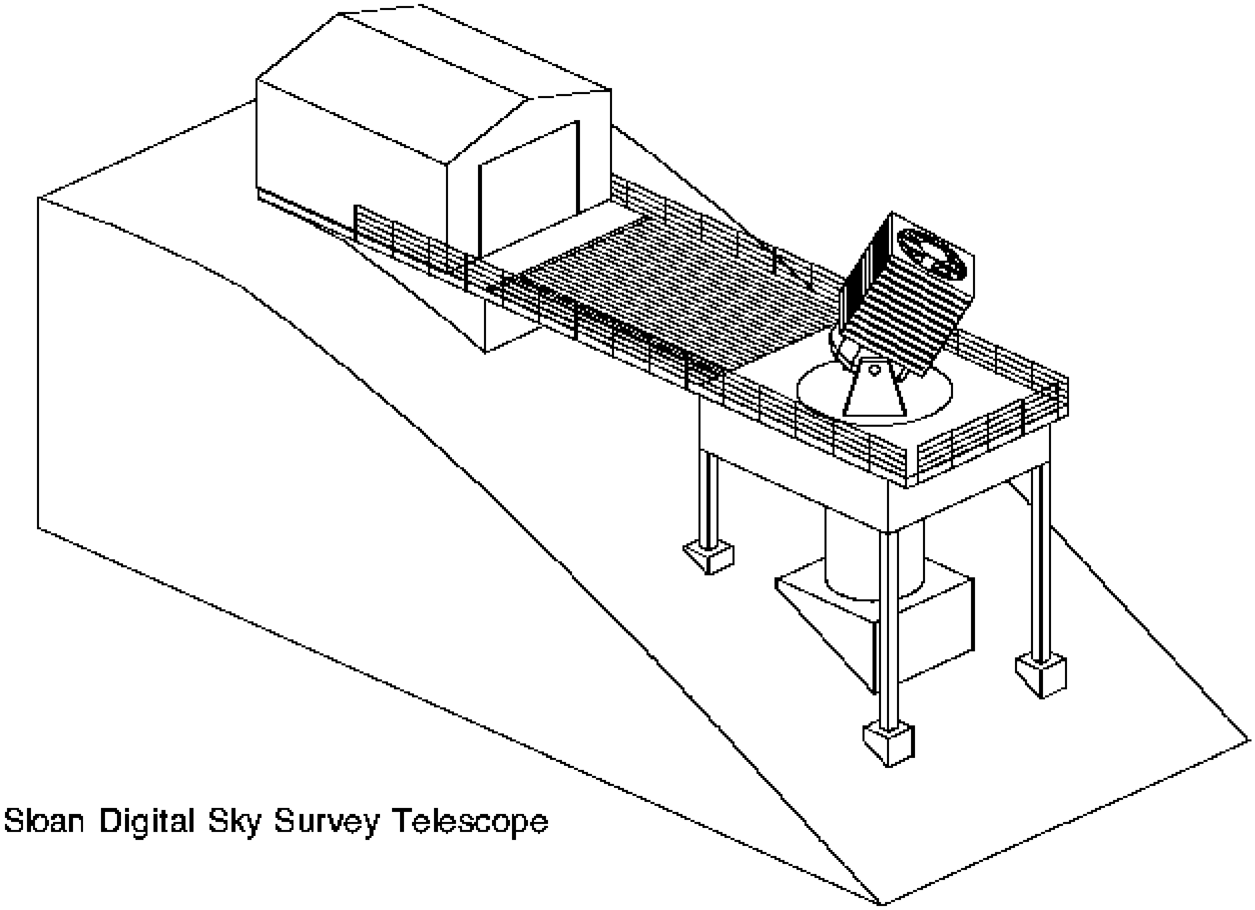
Figure 2.8



The 2.5-m telescope as installed in October, 1995. Centered on the telescope azimuth axis is a circular floor panel that rotates with the telescope. Access to the telescope is along a horizontal ramp from the support building through the telescope enclosure (at the right edge of the photo) which is in its open position.

The telescope site is about 90 meters (300 feet) south of the existing 3.5-m telescope enclosure at the Apache Point Observatory, about 20 meters (70 feet) west-southwest of the ridge top, i.e. in the prevailing upwind direction, and allows the telescope to be located at the same level as the ridge top and still be above the trees, 9 to 12 meters high (see Figure 2.8). The support building on the ridge top, near the telescope, will be used for plate plugging and the storage of plug-plate cartridges. With the support building level with the telescope, plug-plate cartridges can be wheeled easily between the support building and the telescope.

Figure 2.9



Perspective view of the telescope and enclosure with the enclosure rolled off.

The telescope enclosure is a roll-away rectangular frame structure mounted on wheels, and is shown in a perspective drawing in Figure 2.9. During observations, it is rolled downwind from the telescope to the top of the ridge. Large doors on either end of the enclosure are opened during this operation to prevent interference with the telescope and to reduce the wind-load in the direction of motion.

The roll-off approach has several advantages. The thermal ones have been recognized for some time. There are, of course, associated difficulties, particularly wind loading and stray light as discussed above, but for our application the advantages seem compelling enough to compensate for the problems of operating the telescope in the open, to wit:

- The telescope enclosure is considerably smaller and less expensive at $570k than a conventional enclosure since it does not need to accommodate the entire volume swept out by the telescope.
- The telescope enclosure wake is minimal, especially when open. It is important to avoid buildings with large wakes at sites (like APO) with other telescopes, since air overcooled by conduction to the ground is transported throughout the wake volume by turbulence. If the light path to a telescope passes through such a wake, image quality degradation will result.
- Commercially available bridge crane end trucks and electrical cable reel systems provide the transport system and power for the telescope enclosure. The doors at the ends of the enclosure are commercially available, as well. For conventional enclosures, the enclosure rotation and shutter drive systems are special designs.
- When realuminization of the primary mirror is required, a hoist mounted to the ridge beam serves to lift the primary mirror out of the telescope. The enclosure is used to transport the mirror over the bed of a truck waiting at the ridge top. This eliminates the need to provide a large pathway for the primary mirror within the structure of a conventional enclosure.

- Traditional dome-induced seeing is eliminated. Sources of heat, radiative cooling, and thermal inertia are reduced and the flushing of the volume surrounding the telescope is improved.
- The tallest trees at the site are located just below the ridge top and near the proposed site. If the telescope were placed at the ridge top, it would be downwind of these trees and the turbulence they produce.
- With the enclosure track built up the slope, only the columns near the telescope need be long. Much of the track is near ground level, allowing better access for maintenance.

Personnel safety is assured by allowing control of the enclosure motions only from a location that provides a good view of the operation. The perimeter of the platform surrounding the telescope is protected by a railing. The outdoor portion of the deck is covered with galvanized steel grating to provide personnel safety but minimize snow buildup.

The design provides several options for handling drive system failures. Power to the building is backed up by an emergency generator. The telescope and wind baffle drive systems allow them to be moved with hand-operated winches into an orientation allowing the enclosure to be closed. The force required to move the enclosure (with the doors open) directly into a 25 m/s (55 mph) wind is 7000 N (1500 lbs). As a back up, this force is supplied by a pair of electric wire rope winches powered by a small general purpose gasoline generator.

References

Barlow D.J., Blanco D.R., and Poyner A.D., 1987, "Tracking a 150 ton altitude-azimuth telescope to sub-arcsecond accuracy", in *Structural Mechanics of Optical Systems II*, A.E. Hatheway, ed., SPIE 748, 17.

Bowen I.S., and Vaughan A.H., 1973, *Applied Optics*, **12**, 1430.

Davison W., and Ulich B.L., 1982, "Performance of the Multiple Mirror Telescope (MMT): II. Mechanical Properties of the MMT", in *International Conference on Advanced Technology Optical Telescopes*, G. Burbidge and L.D. Barr, Ed., SPIE 332, 9.

Gunnels, S.M., 1990a, "Detail design problems and their solutions: Apache Point Observatory 3.5-m telescope", in *Advanced Technology Optical Telescopes IV*, L.D. Barr, Ed., SPIE 1236, 854.

Gunnels, S.M., 1990b, "Direct Friction Drives", *Magellan Project Report No. 18*.

Johns, M. W., and Pilachowski, C., 1990, "WIYN 3.5-m telescope project", in *Advanced Technology Optical Telescopes IV*, Barr L.D., Ed., SPIE, 1236, 2.

Johns, M. W., 1991, "Azimuth Bearing Study for the WIYN 3.5 Meter Telescope", Wisconsin, Indiana, Yale, NOAO 3.5 Meter Telescope Internal Report, 1991, WODC 02-06-01.

Lloyd-Hart, M., 1990, "System for precise temperature sensing and thermal control of borosilicate honeycomb mirrors during polishing and testing", in *Advanced Technology Optical Telescopes IV*, L.D. Barr, Ed., SPIE 1236, 844.

Mannery E.J., Siegmund W.A., Balick B., and Gunnels S., 1986, "Design of the Apache Point Observatory 3.5 m Telescope IV. Optics Support and Azimuth Structures", in *Advanced Technology Optical Telescopes III*, L.D. Barr, Ed., SPIE 628, 397.

Mannery E.J., Siegmund W.A., Balick B., and Gunnels S., 1986, "Design of the Apache Point Observatory 3.5 m Telescope IV. Primary Mirror Support System", in *Advanced Technology Optical Telescopes III*, L.D. Barr, Ed., SPIE 628, 390.

Poyner A.D., Montgomery J.W., and Ulich B.L., 1986, "MMT Pointing and Tracking", in *Advanced Technology Optical Telescopes III*, L.D. Barr, Ed., SPIE 628, 9.

Schier J.A., 1990, "Torque Perturbations in the Magellan Main Drive Motors", *Magellan Project Report No. 16*.

Siegmund, W.A., Stepp, L.M. and Lauroesch, J., 1990, "Temperature Control of Large Honeycomb Mirrors", in *Advanced Technology Optical Telescopes IV*, L.D. Barr, Ed., SPIE 1236, 834.

Ulich B.L., 1988, "Overview of Acquisition, Tracking, and Pointing System Technologies", in *Acquisition, Tracking, and Pointing II*, SPIE 887, 22.

Site

General Considerations

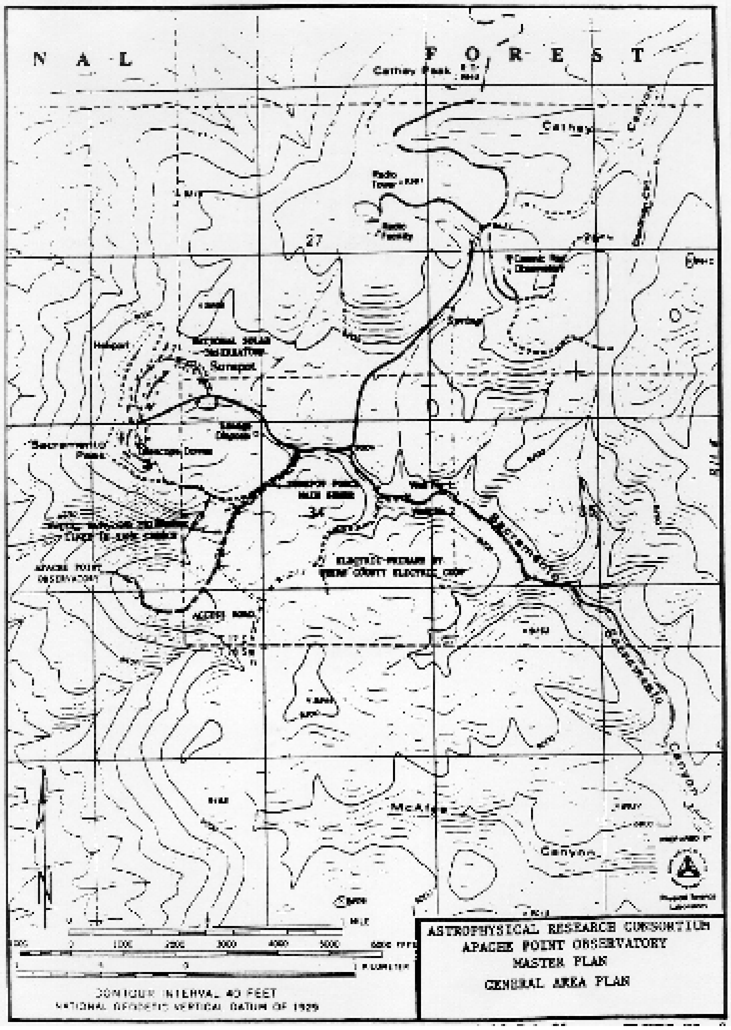
We propose to conduct the survey at the Apache Point Observatory (APO) on Sacramento Peak in New Mexico, a site which meets all of the requirements and most of the desiderata of the SDSS. The site is on the continental U.S., is well-developed and is readily accessible. The weather, seeing and sky brightness at this site are excellent, as necessary for the successful conduct of the SDSS in our lifetimes. Three of the SDSS institutions (Chicago, Princeton and Washington) have been involved, through their membership in the Astrophysical Research Consortium (ARC), in the development of this site, which has already in place the support infrastructure needed to run the survey and has a Forest-Service approved place to build the 2.5 m telescope.

Several major redshift surveys, including the SDSS, are currently planned or are underway. Most of these are to be conducted in the Southern hemisphere, and the comparison of the results of these surveys with those of the Northern hemisphere SDSS will greatly enhance the value of both sets of surveys. There are at present no ongoing or planned redshift surveys in the Northern hemisphere which are at all comparable with the SDSS in terms of angular coverage or number of redshifts to be obtained.

The technological challenge of the camera and the data system, and the drilling of the spectroscopic fiber plug plates, all mandate location of the SDSS at a site easily accessible for the member institutions. Location of the survey at APO makes it logistically possible to carry out the major part of the data reduction at Fermilab. We also hope to be able to handle the transfer of data products to the SDSS member institutions over the Internet or its successor by the time the survey is underway, and this may not be possible from a more remote site.

Location and Topography

Figure 3.1



Contour map of the site. The Apache Point area and the Sacramento Peak Solar Observatory are shown.

The APO site lies in the Sacramento Mountains of south-central New Mexico within the Lincoln National Forest, and is about half a mile south of the Sacramento Peak station of the National Solar Observatory. The latitude and longitude are 32° 46' 50" N, and 105° 49' 12" W, and the elevation is about 9180 feet (2800 meters). A map of the site area is shown in Figure 3.1, and includes the locations of the Cosmic Ray Site, the National Solar Observatory, and the roads in the area. APO is located on a promontory which projects westward from the main ridgeline.

The summit area itself is fairly flat, gently sloping, and lightly wooded.

Since the main ridgeline is the western boundary of the Sacramento Mountains, only low foothills lie between Apache Point and the Tularosa Basin, 10 km to the west.

The prevailing winds for most of the year encounter the Apache Point Site as the first obstacle, and there are thus no upwind sources of turbulence or warm air. The turbulent boundary layer is locally quite thin.

Access to the observatory site is necessarily from the east since deep and steep-walled canyons bound the promontory in the other directions. An 1800-meter paved road connecting to State Highway 6563 has been in place for several years to provide access to the ARC 3.5 m telescope.

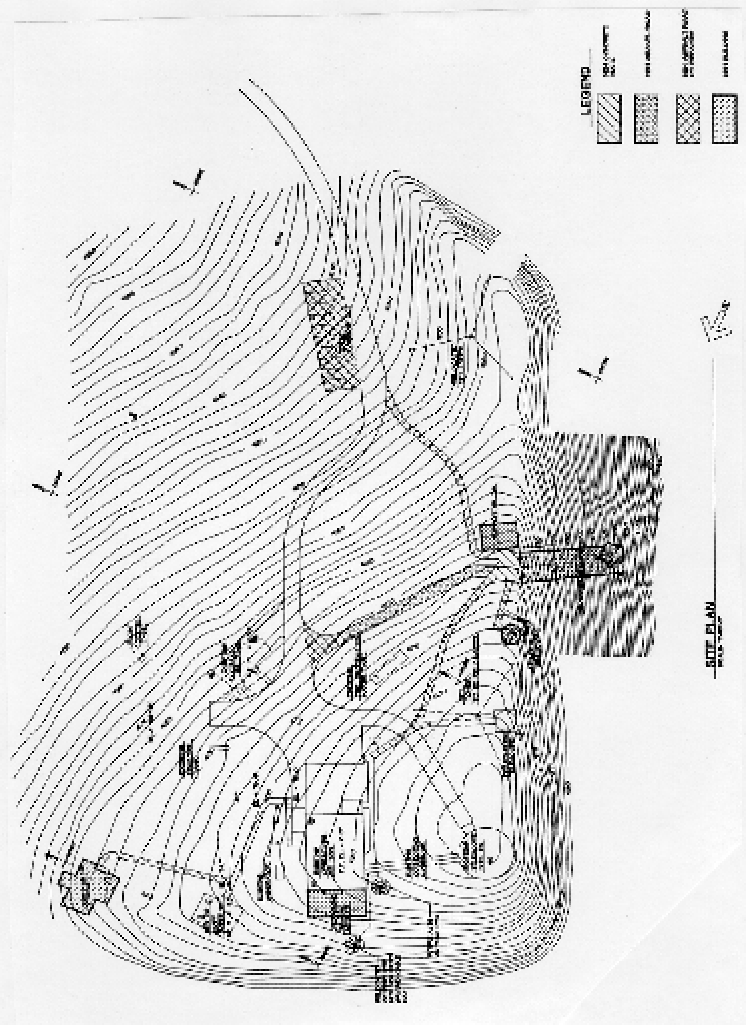
Site Development

Early planning by the Astrophysical Research Consortium, in consultation with the U.S. Forest Service, allocated two sites for major telescopes at Apache Point in addition to the 3.5-m site, where development is complete. The Board of the Astrophysical Research Consortium has undertaken the management of the construction phase at Apache Point for the SDSS, and the subsequent conduct of the SDSS, using one of these two sites.

In addition to the telescope itself, there is SDSS supporting construction needed at the site, but ample space for these associated activities is available. Consequently, in August 1992, after soliciting proposals from a number of highly qualified design firms, M3 Engineering and Technology of Tucson, Arizona was selected to provide design services for the necessary new developments at the site to accommodate the SDSS. M3 Engineering has extensive experience in the design of astronomical observatories in the southwestern United States and completed the design for the new survey facilities in February 1993. Construction was completed that same year.

The development for the 3.5-m site includes the telescope enclosure, an operations building, a utility building and a dormitory (Figure 3.2). The operations building contains 350 m² (3700 ft²) of floor area and includes a computer room, electronics shop, darkroom, kitchen, lounge and library. The SDSS will be able to share many of these facilities. In particular, the present computer room contains adequate space for the needs of both telescopes. About 132 m² (1424 ft²) of floor area was added to the operations building. The existing library and electronics shop were moved to this addition. The space thereby vacated will be used in support of the survey as a control room and workshop. Also, three new offices were added for survey personnel and the original darkroom was remodeled for use as a clean room for servicing the CCD camera.

Figure 3.2



Site plan for the SDSS project. The 2.5 meter telescope building, the plugging support building (all shown cross-hatched) and operations building (new addition shown cross-hatched) are shown. The new dormitory building (cross-hatched) is at the upper left.

The utility building contains a two-car garage and a 50 m^2 (500 ft^2) machine shop.

The original dormitory consists of three motel-type rooms, each with two beds. A second dormitory was added about 30 meters east of the existing dormitory, and contains six single rooms and a kitchen.

The 2.5-m telescope is located about 90 meters (300 feet) south of the 3.5-m telescope enclosure. The site is 20 meters (70 feet) west-southwest of the ridge top. The 2.5-m support building is near the telescope just beyond the open position of the telescope enclosure. Its main function is to provide a space for the plugging of plug-plates and for the storage of plug-plate cartridges in the plugging queue or the telescope queue. Inevitably, some telescope and instrument maintenance will occur in the support building because of its proximity to the telescope.

Near the 2.5-m support building is the Monitor Telescope enclosure (see Chapter 5). Since this telescope does not require excellent image quality, it is located near the support building for maximum convenience and sheltered in a small commercial dome.

A driveway links the 2.5-m telescope with the rest of the site. It is used to deliver the plug-plates to the support building, to supply consumables and for the occasional removal of the primary mirror for realuminization. This driveway is not paved so that the heat storage and absorption characteristics of the soil near the telescope are not degraded. Sidewalks connect the support building to the operations building and the lower parking lot, and provide access for people and wheeled carts.

As part of the same construction contract, a 1.0-m telescope enclosure was built for New Mexico State University (NMSU). Under an agreement with NMSU, ARC is providing a telescope site and some services and facilities to NMSU. The terms of the agreement provide that NMSU pay its share of all costs associated with construction and operation of the telescope and that management and funding is independent of ARC. The 1.0-m telescope site is located between the 3.5-m and monitor telescope sites. A 30 meter high tower which supports microthermal sensors and wind speed and direction indicators is located 60 meters (200 feet) south of the support building.

We have not made significant changes or additions to the other infrastructure at the site. These facilities include a 185 kVA emergency power backup generator and a 20,000 gallon water storage tank and 500 gpm pump for fire protection. Electrical power comes to the site via overhead 3-phase conductors with a grounded counterpoise for lightning protection. These conductors are owned and maintained by Otero County Electrical Cooperative. Domestic water is supplied to the site from wells belonging to the National Solar Observatory (NSO), for which ARC is charged at local Cloudcroft rates.

US West provides 25 phone lines to the site. We currently have four voice lines for general communications, one line dedicated to FAX service, two lines dedicated to modem use, and 1 circuit dedicated to a leased T1 (1.544 Mbit/second) line to Las Cruces, New Mexico for Internet access. These lines come from a microwave tower through NSO to APO, at APO expense. White Sands Missile Range has 1 voice circuit and 4 data circuits. There are three new circuits currently awaiting installation.

Weather and Sky Brightness

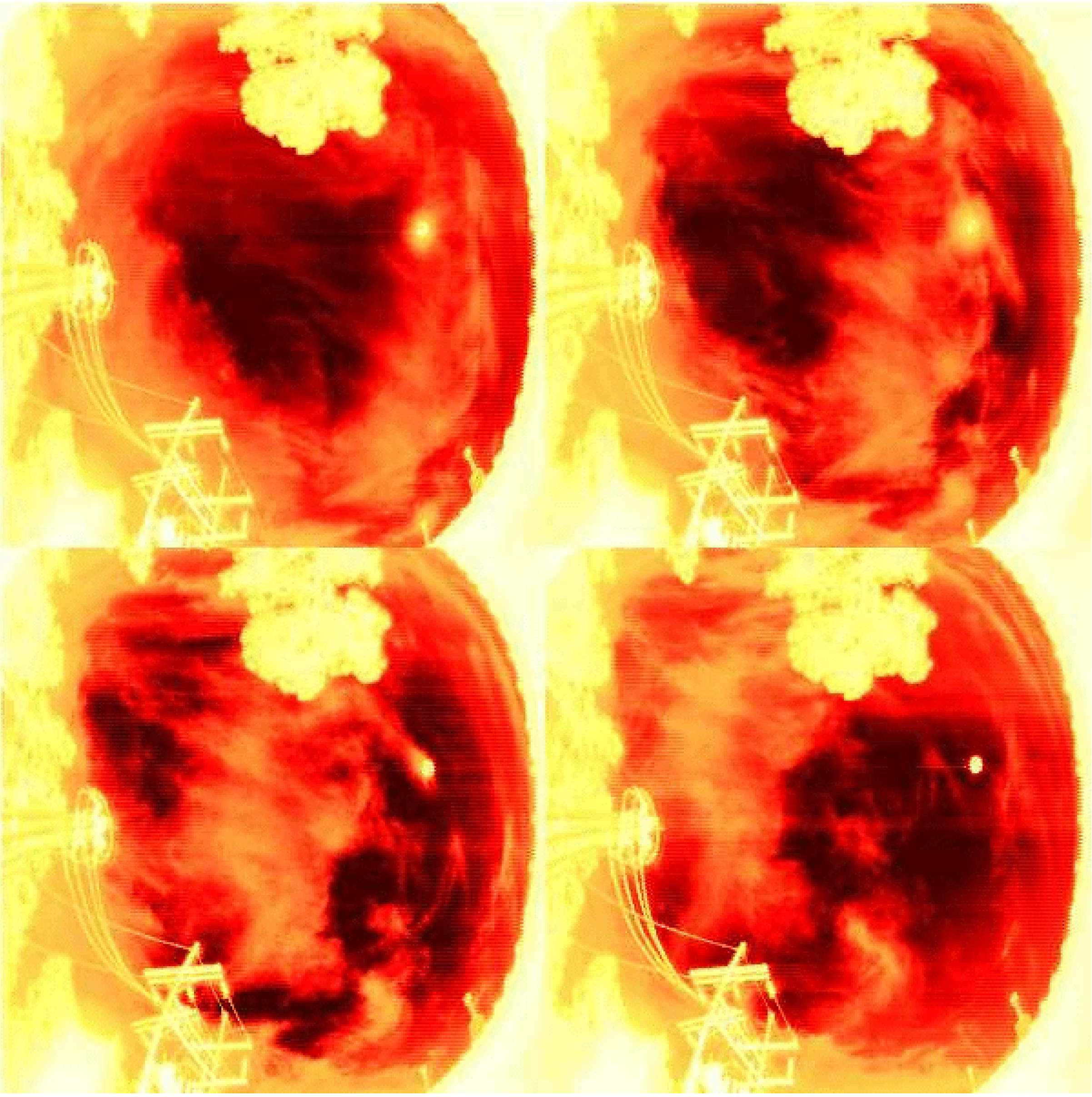
The Apache Point climate is relatively mild, particularly considering its 2800 meter elevation. Winds strong enough to interfere with normal observatory operations are relatively rare and are, in any case, often associated with thunderstorms. Rarely should access to the observatory site be impaired by heavy snowfall.

The weather patterns at APO are similar to those at Kitt Peak, though the summer `monsoon' is not quite so severe. At KPNO, from 1966 through 1986 67% of the nights were useful, and perhaps about half of these were photometric. Similar statistics apply to Apache Point. These statistics confirm the results of other studies of cloud cover made over the past 30 years.

Over a year, 70% of the night time is clear, best in October-November (>85 %) and worst in July-August (50-60%). The percentage of nights that remain perfectly cloudless and photometric throughout the night averages about 35% (50% in October-November, <19% in July-August). These numbers are comparable to or better than optical sites at Mt. Hamilton, Mt. Palomar, Mt. Wilson (California), Kitt Peak (Arizona), and Mt. Locke (Texas) (Smith and Salisbury 1962).

The precipitable water vapor should be as low as possible because of the effect of water absorption on i' - and z' -band photometry and spectrophotometry near the long-wavelength limit of the spectrographs. This quantity ranges from less than 2 mm to 5 mm seasonally, and is as low as any major continental observing site except Mt. Lemmon (Arizona), as would be expected at an altitude of 2800 meters.

Figure 3.3

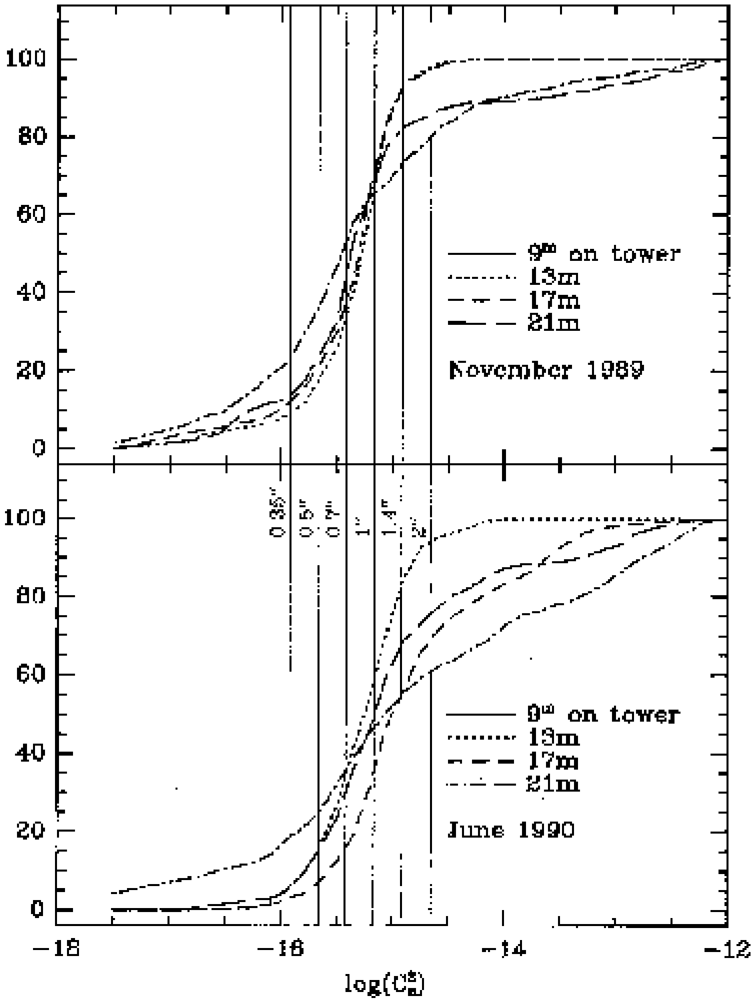


Four images of daytime cloud cover at APO. The observations are taken four minutes apart by the 10 μm cloud scanner. The cloud motion from WSW can be seen. The brightest feature, above and to the right of center, is the Sun.

Since early 1994, an imaging cloud scanner (Hull et al. 1995) has been in operation at the site to provide rapid, reliable, quantitative measures of cloud cover. This scanner uses a liquid-nitrogen cooled HgCdTe infrared detector which maps the sky over a 135°x135° area centered on the zenith every ten minutes with an angular resolution of 0.9°. The scanner operates at 10 μm and is highly sensitive to emission from clouds. An example of four daytime images is shown in Figure 3.3. The cloud motion can be seen, as can the scattering of sunlight, which enlarges the solar image, as clouds cross the Sun. The emission measured by the cloud scanner can now be accessed by the Monitor Telescope data reduction software (see Chapter 5). Data are now being gathered for a study of the correlation between atmospheric extinction measured by the MT and 10 μm emission measured by the cloud scanner.

The average visual sky brightness (Schneeberger et al. 1979) is 21.9 V mag per square arcsecond, not including the darker, midwinter nights (December-February). The site is protected from growth in light pollution by White Sands Missile Range and National Monument to the west, the Lincoln National Forest to the east, the Mescalero Apache Reservation to the north, and the Fort Bliss Military Reservation to the south. The growth of nearby Alamogordo is expected to be limited by availability of water; in any case, the City of Alamogordo in 1990 passed a lighting ordinance, very favorable to astronomy, which controls the types of lighting which can be used in public places and places strict control on lighting after 11:00 PM. In 1978, El Paso and Alamogordo contributed 10% of the sky brightness at 85° zenith angle and 2% of the sky brightness at 45% zenith angle, at the respective azimuth angles of 180° and 270°. Garstang (1989) has made light pollution projections based on a population model (without account of changes in lighting ordinances) and projects that the sky at the zenith will increase in brightness by $0^{m}.03$ from 1990 to 2000.

Figure 3.4



Microthermal sensor data from APO. Cumulative distributions for the refractive index structure coefficient C_n^2 for all nights in November 1989 and June 1990 measured at at the four levels on the tower (9.2,12.6,16.8,and 21.2m) are shown. The vertical bars labeled with the FWHM seeing diameter are derived using an atmospheric model (Anderson and Morales 1991).

Seeing

Figure 3.5

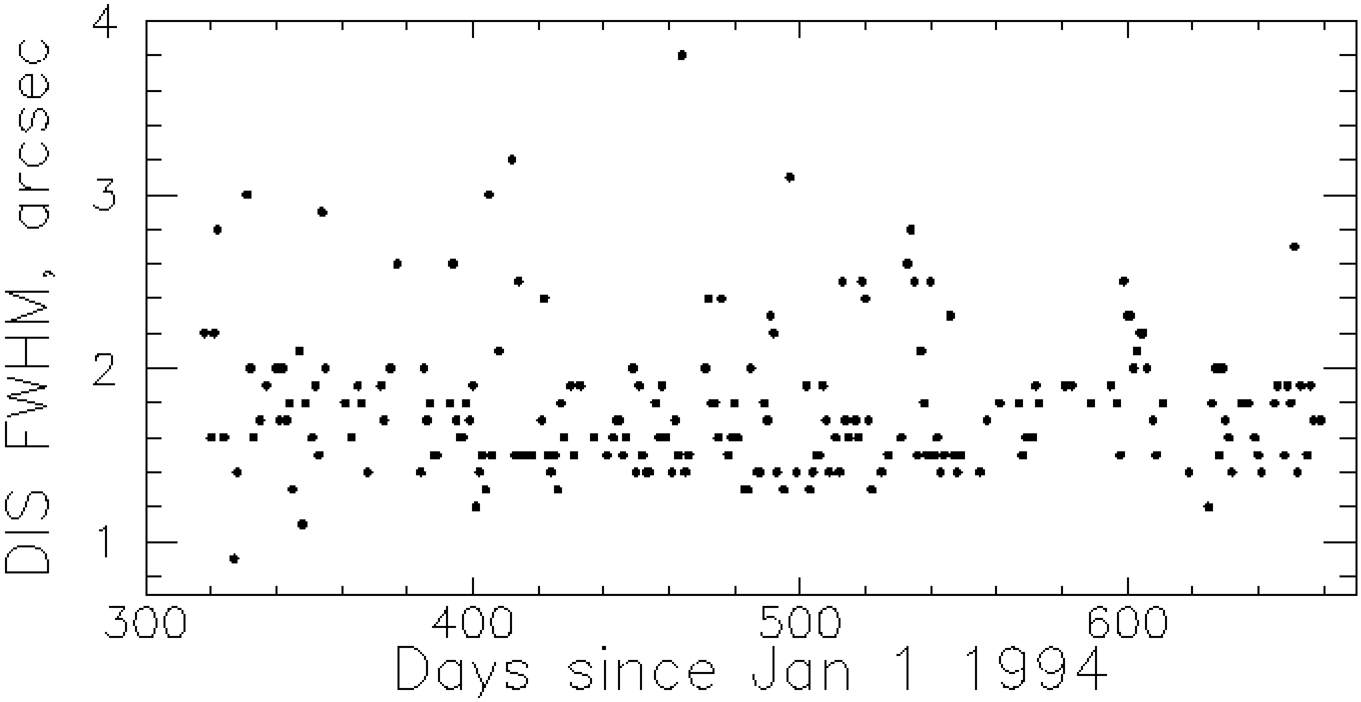


Image size at g measured by the Digital Image Spectrograph (DIS). The data were taken between November 1994 and October 1995.

All optical telescopes require clear, dark skies, but the operation of the SDSS further requires that about 25% of the clear time be photometric and have seeing <= 1" . The ARC telescopes have been located at APO because it fulfils these requirements. The National Solar Observatory, located at Sacramento Peak about half a mile from the APO site, ran nighttime seeing tests with the vacuum tower telescope during 1976 and 1977. The median seeing was 0.8 arcseconds and the seeing was better than 1 arcsecond 56% of the time (Beckers et al. 1979). The APO site should be even better than this, since it typically samples a more nearly undisturbed wind flow than does the Sacramento Peak site.

Figure 3.6

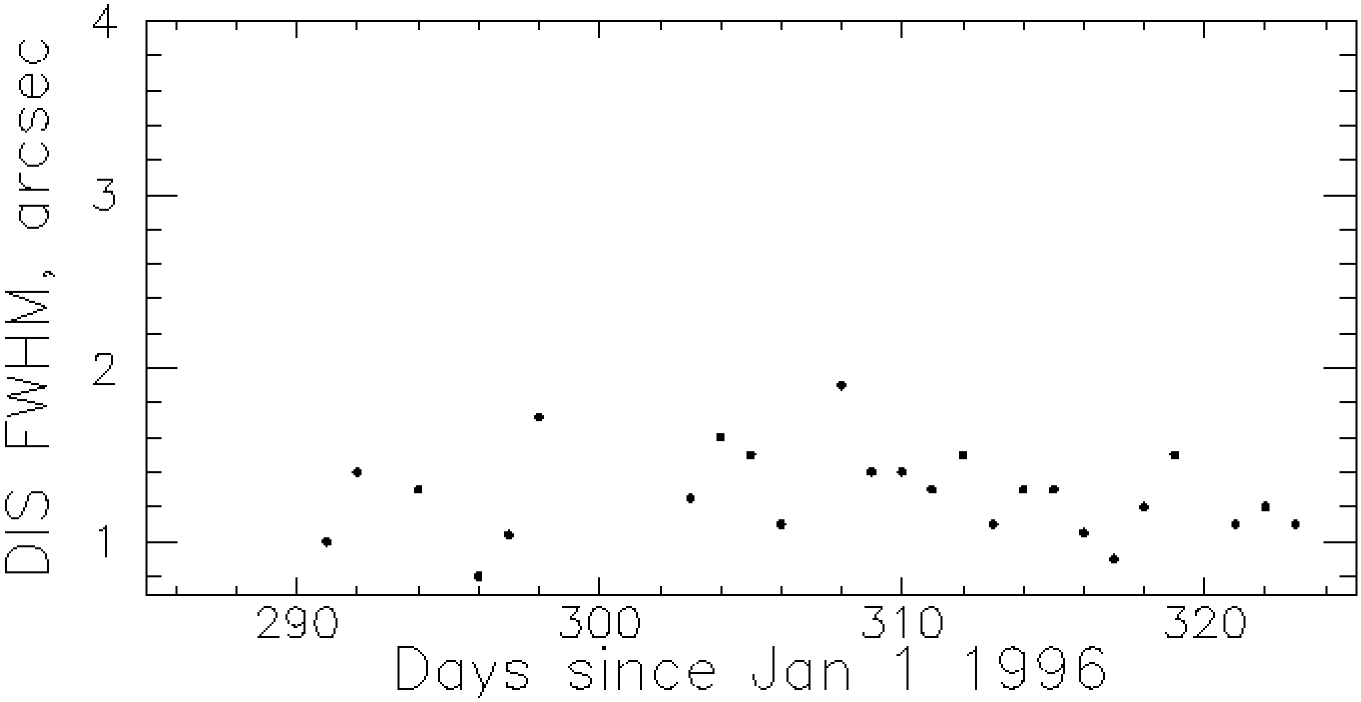


Image size at g measured by DIS between October and November 1996.

A weather tower was installed at APO in 1984 and has been in almost continuous operation since, under the direction of Kurt Anderson (New Mexico State University), the ARC site director. This tower is located on the ridge occupied by the 3.5 m and 2.5 m telescopes, about midway between them. The 23-m tower has microthermal sensors at heights of 9.2, 12.6, 16.8 and 21.4 meters above the ground. Each sensor consists of two coils of 40-micron nickel wire, separated by one vertical meter and connected in a bridge circuit. The resistance is temperature sensitive, and the measured temperature differences can be converted to differences in the index of refraction of the air at different levels, using the temperature and pressure of the ambient air, also measured at the tower. The mean-square differences of refractive index can then be converted to the index of refraction *structure coefficient* C_n^2 for Kolmogorov turbulence:

$$C_n^2(h) = \left\langle [n(h+r) - n(h)]^2 \right\rangle r^{-2/3},$$

where h is a vertical vector to a layer of height h and r is a separation vector between two points.

With a vertical exponential scale length for the index variations and an atmospheric model (a plane-parallel, standard atmospheric model was used), the resulting structure coefficient is simply related to the Fried parameter, r_0, (Roddier 1981) which in turn is related to angular image size theta : theta ~ lambda /r_0 .

Eleven years of data are now available from this tower. Seeing effects at 1000 meters, measured by airplane flights in 1977 and 1978, can be combined with the tower measurements to yield the scale height of the index variations (Anderson and Morales 1991). The results can be summarized as follows: The seeing measured in this experiment has a median near 0.7 arcseconds, and is below 0.5 arcsecond 10-20% of the time, depending on the season. The microthermal tests show that the seeing is best in the winter months and worst in the late summer. Typical measurements for a month in November 1989 and another in June 1990 are shown in Figure 3.4. The interpretation of these data is not quite straightforward; one could simply average the values at different levels, or one could take the highest level as most representative of the atmosphere above the tower; the actual contribution to the integral of the values near the ground is small. In any case, the results are somewhat better, on average, than the direct measurements of Beckers et al. (1979), possibly because of his smaller database, possibly because his site is not quite so good, and certainly because his measuring resolution was only 0.5 arcsecond.

We are not yet able to confirm the expected good seeing performance from direct observations with the ARC 3.5 m telescope, which has been in operation for the past three years; the telescope optical system (particularly the secondary mirror optical figure) has not yet been optimized. The image size measured by the Digital Imaging Spectrograph (DIS), is shown for the period November 1994 - October 1995 in Figure 3.5, and for October - November 1996 in Figure 3.6. The image size measured by DIS in the last year (November 1995 - October 1996) has been below 1.5" 70% of the time, a big improvement.

Measurements of the optical figure of the telescope by C. Stubbs (1996) show that the telescope optical system contributes 0.6" while the large pixel size (0.8") of DIS contributes about 0.3" . The intrinsic seeing corresponding to an image size of 1.2" (Figure 3.6) is then about 0.7" . The median seeing at APO is thus less than 1" at visible wavelengths. Observations with ARC's low-resolution infrared spectrograph and imager (GRIM) yield a median seeing of 0.8" and have directly measured image sizes of < 1" 25% of the time in 1996. The 3.5 m data thus confirm that the seeing at APO is good enough enough of the clear time (25%) to carry out the SDSS.

In summary, the Apache Point site is characterized by mild weather, a very dark sky, a high incidence of clear skies, and topography which ensures that the seeing is excellent a large percentage of the time. Prior development of the site and proximity to the NSO make siting the survey telescope there much less expensive and more convenient than would be the case at any new site. The nature of the project makes it highly desirable that a continental U.S. site be chosen, and we doubt strongly that a better one exists for our purposes.

References

Anderson, K., and Morales, A., 1991, (private communication).

Beckers J.M., Breedlove W.O., DeMastus H.L., DeVegvar P.G.N., Johansen E.E., Gilliam L.B., Mann G.R., Mauter H.A., and Phillips G.L., PASP 91, 857, 1979.

Hull, C., Limmongkol, S., & Siegmund, W. 1994, *Proc. S.P.I.E.*, 2199.

Roddier, F., 1981,*Progress in Optics* **XIX**, 281.

Schneeberger T.J., Worden S.P., and Beckers J.M., PASP 91, 530, 1979.

Smith, B. A. and Salisbury, J. W., 1962, *Air Force Cambridge Laboratories Report* **1064**.

Stubbs, C.W. 1996, (private communication).

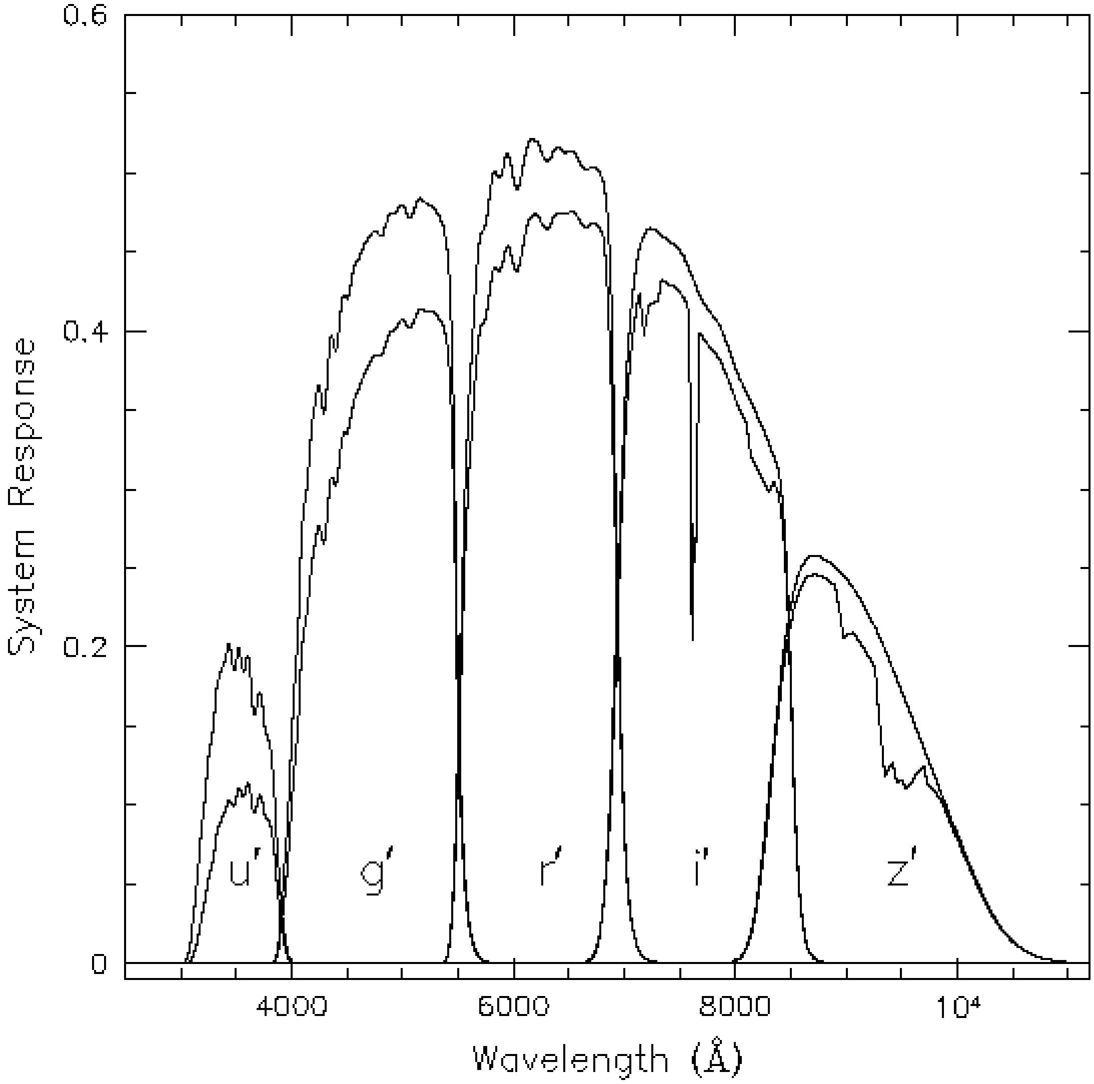
The Photometric Camera and the CCDs

The Photometric System

The photometric camera for the SDSS consists of two TDI scanning CCD arrays, one, using 30 Tektronix/SITe 2048 x 2048 CCDs in a 5 by 6 array for five-color photometry, and the other using 24 2048 x 400 chips for astrometry and focus monitoring. We here discuss these arrays, the filter system, the scheme for focusing the telescope during an imaging scan, and the CCDs and associated electronics for the camera and the spectrograph.

There are several important considerations in selecting the filter system. First, one wants the largest color baseline possible. It is desirable but not necessary to use filters for which there already exist a large amount of data. For example, the Thuan-Gunn (1976), (see also Schneider et al. 1983) uvgriz system, which has been used for much of the work at Palomar on distant clusters, avoids the 5577Å night sky line, and the i band cuts off in the near-infrared just shortward of the major OH bands. The g and r filters more-or-less match the J and F bands of the new Palomar Sky Survey. The uvgriz filters provide coverage of the total wavelength interval and are almost uniformly spaced in log lambda . They are, however, fairly narrow, and all except u could be usefully widened for a system like ours in which efficiency is of paramount importance. We have designed a filter system similar to ugriz which should transform to and from it with little difficulty, in which the g , r and i filters are as wide as practicable consistent with keeping the overlap small. If we wish the u band to be a "good" one which is almost entirely contained between the Balmer jump and the atmospheric cutoff (we can clearly not do anything about that end), it cannot be significantly wider than Thuan-Gunn u . We call the new system u' g' r' i' z' . The response curves and sensitivity data with the coatings and CCDs we will use are shown in Figure 4.1.

Figure 4.1



The SDSS system response curves. The responses are shown without atmospheric extinction (upper curves) and as modified by the extinction at 1.2 airmasses (lower curves). The curves represent expected total quantum efficiencies of the camera plus telescope on the sky.

The characteristics, are, in brief: u' , a band with 3540Å effective wavelength, 570Å full width at half-maximum, and 60% peak transmission (such a filter can be made with available glasses; it has a small redleak problem which can be solved with suitable coatings); g' , a very wide blue-green band centered on 4770Å (full width at half-maximum 1370Å); r' , a red band of the same width centered at 6230Å; i' , a far-red band centered at 7630Å with a width of 1530Å, and z' , a near-infrared band centered at 9130Å with a width of 950Å. These bands share the advantages of the uvgriz system of avoiding the 5577Å [OI] line, and the i' band cutting off before the catastrophic OH brightening and water absorption set in just longward of 9000Å. All the filters except u' are easy to make and have very high central transmission; building the u' filters proved to be difficult, but the results are quite satisfactory.

The filters are all made with one or more Schott colored glass elements in combination with a single-surface interference film which serves different purposes in different filters. The u' filter is made of 1 mm of UG11 and 1 mm of BG38 with an interference layer which both acts as an antireflection coating in the filter passband and an efficient red blocker in a region a few hundred Å wide around 7000 Å; the glass filters have sufficient red rejection outside this region, but the uncoated filter has a peak transmission of about two percent there. The coated filter has a maximum transmission of about 10^{-4} outside the primary band. The g' filter is made of 2 mm of GG400 and 3 mm of BG38, with a film which acts as a short-pass element cutting on at 5500 Å; the BG38 blocks the red efficiently out to the silicon limit. The three longer filters consist of a single colored long-pass element and for the two shorter ones a short-pass interference film, the r' 4 mm RG550 and short-pass cutoff at 7000 Å, the i' 4 mm RG695 and short-pass cutoff at 8500 Å, and the z' 4 mm of RG830 open to the red. Each filter is brought to 5 mm physical apex thickness by means of an (quartz for u' , BK7 for the rest) individual field flattener cemented to it; this element is necessitated by the fact that the chips are not flat, but are convex to the incoming light by about 230 microns measured center to corner. The final filters were all constructed and coated by Asahi Spectro-Optical in Tokyo; the characteristics and uniformity are superb.

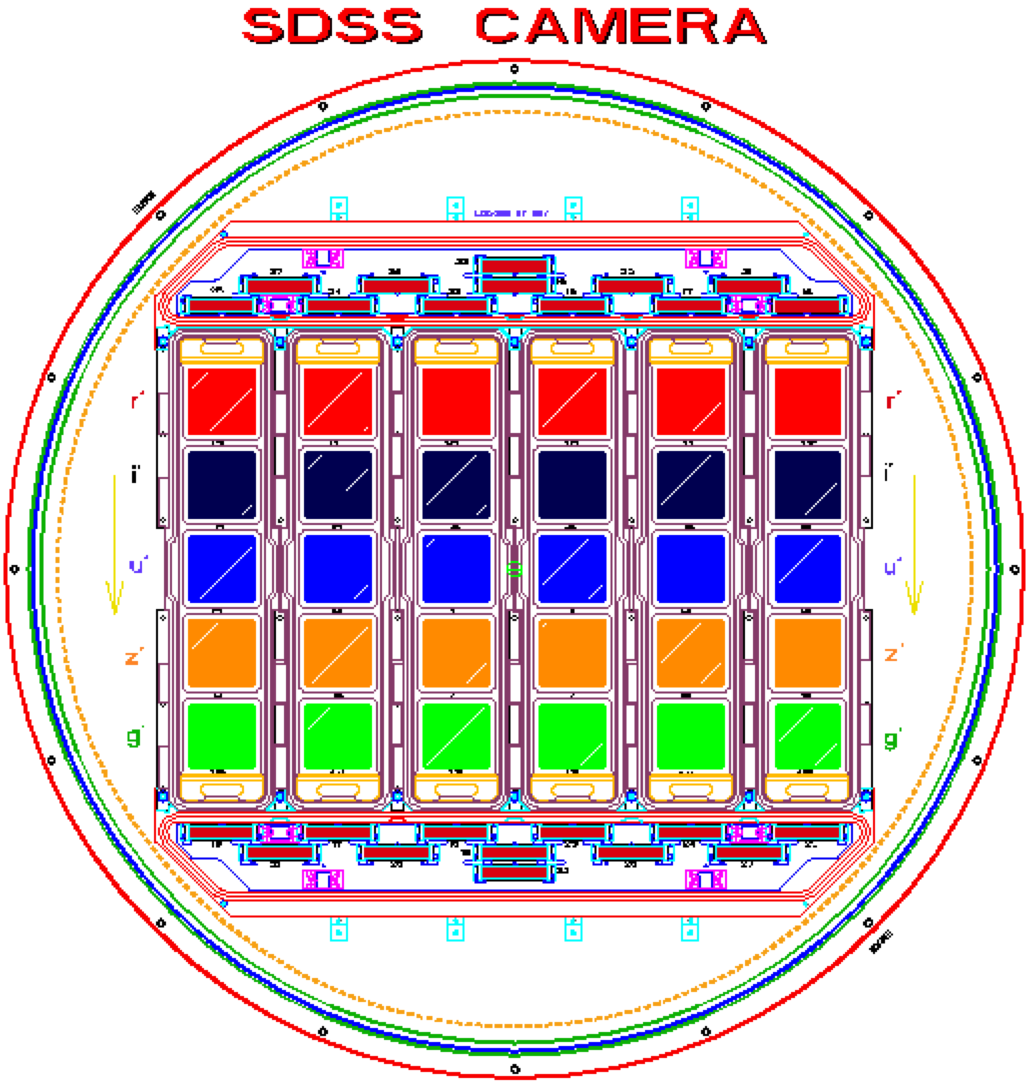
The filter assemblies in turn are cemented to the flat rear face of the final corrector element. The front face of the corrector has been anti-reflection coated in horizontal stripes tailored for each filter passband, resulting in about 0.2% reflectivity over each band. The coating was done by QSP in Los Angeles. Since the filters are cemented assemblies and are themselves cemented to the corrector, there are only two optical surfaces near the CCD, and one of those has very low reflectivity. The other is the filter coating itself, which has low reflectivity in the passband but unfortunately has intrinsically high reflectivity near the long-wave cutoff over a very small wavelength region. It is these "crossover" regions which contribute most of the ghost energy. We have calculated ghost images for the camera, and they will not present a serious difficulty for the photometric reductions.

With our arrangement of CCDs operating in TDI mode at sidereal rate, we obtain near-simultaneous color data in the five bands; the time lapse from the r' data, the first received, to the g' , the last, is about five minutes.

It is worth noting that this color set is *not* any standard one, but this survey will produce two orders of magnitude more photometry than currently exists, and to adopt a system not optimized for the survey seemed folly. The filter set is in any case a good one for CCDs and has some quite nice astrophysical properties; we hope that it will quickly *become* a standard system. Toward that end we are in the process of producing as complete an astrophysical characterization of it (metallicity, gravity, temperature, galaxy type and redshift dependences) as we can; we clearly badly need these data ourselves. We will attempt to ensure that workers wanting sets of filters for this system will have reasonable avenues open to obtaining them.

With these filters, the CCDs with which they will be used, and with the SDSS telescope scanning at sidereal rate, the limiting magnitudes at the detection limit, S/N 5:1, will be about u'=22.3 , g'=23.3 , r'=23.1 , i'= 22.5 and z' = 20.8 . These limiting magnitudes are for stars, note, and those for galaxies are typically between half a magnitude and a magnitude brighter at the same signal to noise ratio. A signal to noise ratio of 50:1 (and hence photometry at the 2% level) is reached for stars at 19.3, 20.6, 20.4, 19.8 and 18.3 in the five bands. Note that these are magnitudes on the AB system, referred to a spectrum with constant f_{nu} (not an A0V star), and represent the fluxes a flat spectrum source with a given V magnitude would have near the effective wavelengths of the filters. It is these zero points we intend to adopt. These sensitivity calculations assume 1 arcsecond FWHM seeing (with a probably somewhat pessimistic non-gaussian PSF) and a sky with a V surface brightness of 21.7 mag/arcsecond^2 (typical of the Apache Point site) with average extinction at the zenith. A darker site helps, with roughly half a magnitude gain in limit per magnitude improvement in the sky (at these faint levels shot noise in the sky is the dominant noise source).

Figure 4.2



Front view of the SDSS camera assembly. This diagram shows the camera as it would be seen with the front cover and shutters removed, showing the 30 photometric and 24 astrometric/focus CCDs and their associated dewars and kinematic supports.

Although the u' band is obviously useful, the other bandpasses provide a maximum wavelength ratio of 2.4, and adding u' only increases the baseline to 3.0. Is there much advantage to using u' versus a slightly longer wavelength band such as Johnson B or Thuan-Gunn v ? The answer is resoundingly yes, on three accounts. First, the u' -band is much more sensitive to population and metallicity effects than any of the longer wavelength bands. For example, for normal spiral galaxies, U-B changes 1.5 times as fast as B-V . In elliptical galaxies, color gradients (which most likely reflect metallicity variations) are 3 times as big in U-B as in B-V ; one must reach from B to the K -band to achieve comparable sensitivity. The u' -band lies shortward of the H and K break in late-type galaxy spectra and provides the best indicator of the presence of recent star-formation activity. Second, at high redshift, galaxies are usually observed in the more accessible and sensitive longer wavelength bands, which register light that, in the galaxy rest frame, is emitted in the ultraviolet. The u' -band is shifted to g' at $z = 0.34$, r' at $z = 0.83$, and i' at $z = 1.3$. Third, the ultraviolet is essential in any color-based QSO survey to separate low-redshift ($z<2$) QSOs from stars, particularly low-metallicity halo stars, and, incidentally, to separate low-metallicity halo stars from high-metallicity stars.

We had originally planned to devote two of the five chips in a given column to the u' band for the sake of efficiency, but new UV anti-reflection coatings and an accompanying surface treatment developed by Tektronix/SITe for the u' chips, together with a higher-than-expected efficiency of the u' filter, have led us to go with the current set of five independent bands, including the new z' filter. The new band is especially useful in the detection of very high-redshift quasars and extremely red stars, and, perhaps most importantly, distinguishing them from each other.

The Photometric Array-Layout and Performance

The photometric camera is designed to use thirty 2048^2 sensors in a six-wide by five-high array, so arranged that two transit swaths (strips) cover completely a stripe 2.53° wide in the five independent filters, with some overlap (about 1 arcminute) between sensors. About 8.3% of the area of a given stripe is imaged twice. The focal plane of the camera array is illustrated in Figure 4.2. Also shown there are the 22 smaller (2048 x 400) chips used to provide astrometric calibration, and the two 2048 x 400 focus monitors. The scanning is top-to-bottom in Figure 4.2, along the five closely-spaced CCD columns, and it is handy in discussing the camera to think about the geometry in this orientation. (Thus "top" of the camera will always refer to the leading edge.)

SDSS Filter Set and Photometric Sensitivity

Table 4.1a: Filter Parameters

name	lambda	FWHM	q_t	$q_t$dl/l
u'	3543	567	$1.24\text{x } 10^{-1}$	$1.85\text{x } 10^{-2}$
g'	4770	1387	$4.51\text{x } 10^{-1}$	$1.18\text{x } 10^{-1}$
r'	6231	1373	$5.63\text{x } 10^{-1}$	$1.17\text{x } 10^{-1}$
i'	7625	1526	$5.02\text{x } 10^{-1}$	$8.74\text{x } 10^{-2}$
z'	9134	950	$1.29\text{x } 10^{-1}$	$2.23\text{x } 10^{-2}$

Table 4.1b: Photometric Saturation and Sky Background

Filters	u'	g'	r'	i'	z'
Star saturates at AB	12.1	14.1	14.1	13.8	12.3
Eff Sky, mag/sec^2	22.1	21.8	21.2	20.3	18.6
Sky + bkg count/pxl	45	401	690	1190	1120

Table 4.1c: Photometric Sensitivity

Filter	u'		g'		r'		i'		z'	
AB	count	S/N	count	S/N	count	S/N	count	S/N	count	S/N
17.0	32489	173.2	207754	442.2	206173	432.3	153488	352.8	32840	127.3
17.5	20499	134.6	131084	345.4	130086	334.4	96845	266.0	20721	88.8
18.0	12934	103.5	82708	267.4	82079	255.3	61105	196.4	13074	60.5
18.5	8161	78.4	52185	204.5	51788	191.6	38555	141.4	8249	40.3
19.0	5149	58.2	32927	153.7	32676	140.7	24326	99.3	5205	26.4
19.5	3249	42.2	20775	113.1	20617	100.7	15349	67.9	3284	17.1
20.0	2050	29.8	13108	81.2	13009	70.2	9684	45.4	2072	11.0
20.5	1293	20.5	8271	56.8	8208	47.8	6110	29.9	1307	7.0
21.0	816	13.8	5219	38.7	5179	31.8	3855	19.4	825	4.4
21.5	515	9.1	3293	25.8	3268	20.9	2433	12.5	520	2.8

22.0	325	5.9	2078	16.9	2062	13.5	1535	7.9	328	1.8
22.5	205	3.8	1311	11.0	1301	8.7	968	5.1	207	1.1
23.0	129	2.4	827	7.0	821	5.5	611	3.2	131	0.7
23.5	82	1.6	522	4.5	518	3.5	386	2.0	82	0.4
24.0	51	1.0	329	2.9	327	2.2	243	1.3	52	0.3
24.5	32	0.6	208	1.8	206	1.4	153	0.8	33	0.2
25.0	20	0.4	131	1.1	130	0.9	97	0.5	21	0.1

The expected photometric performance of the camera is outlined in Table 4.1, which presents the assumed filter parameters, derived from the curves in Figure 4.1. The quantity q_t is the peak system quantum efficiency in the system, and $q_t$dl/l is the integral of the system efficiency times d(ln lambda) ; it is this quantity which relates the monochromatic flux averaged over the filter passband to the resulting signal:

$$N_{el} = 1.96x10^{11}t(q_t dl/l)10^{-0.4AB_{nu}} .$$

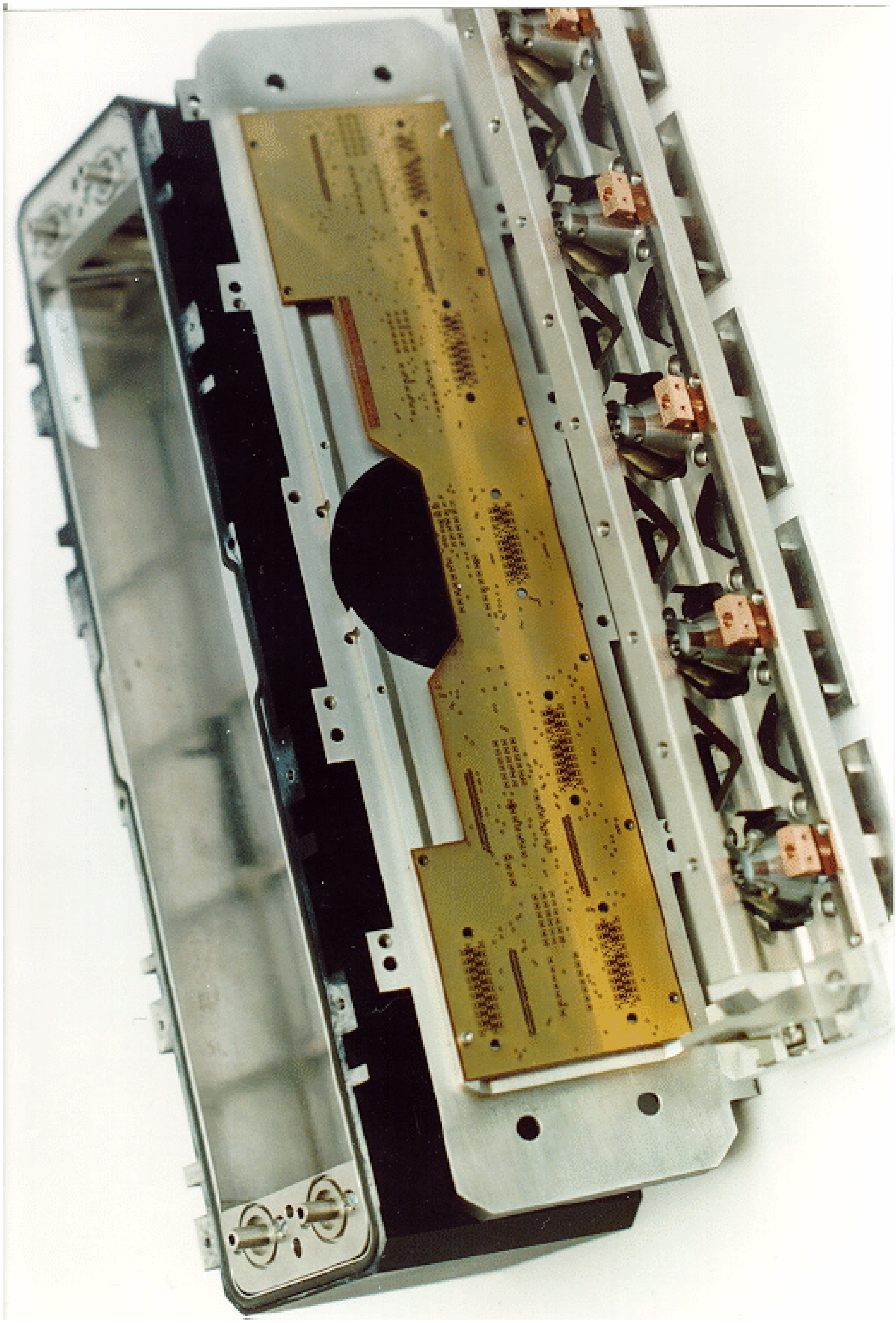
The counts and signal-to-noise ratios for an assumed sky with a V brightness of 21.7 mag per square arcsecond (about 1.2 airmasses for the typical sky brightness at Apache Point) and a Palomar spectrum (Turnrose, 1974) are given in Table 4.1c for stellar objects with 1 arcsecond FWHM images with a reasonably realistic seeing-dominated PSF (see Chapter 9). The assumed exposure time is 55 seconds, which corresponds to a TDI scanning rate near sidereal; this rate is near optimal from a variety of considerations. The assumed full well is 3x 10^5 electrons and the noise is a bit worse than typical at 7 e^-. The assumed CCD quantum efficiencies are averages of the measured ones for the chips that will actually be in the camera; the dispersion is quite small. Typical galaxy images reach a given S/N half a magnitude to a magnitude brighter at the faint end, depending on their surface brightness.

Mechanical Design of the Photometric Array

Since the telescope has been designed carefully to accommodate the CCDs, the camera design is simply a matter of housing the chips stably, controlling their temperature, and tending to their various electronic needs. The only complication in this is brought about by the scale of the project and the attendant necessity of putting a lot of stuff, particularly circuitry, into a relatively small space. The only moving parts on the camera are associated with a simple set of pneumatically actuated shutters which cover each dewar and are intended mostly to protect the CCDs from saturation during the day and when the enclosure lights are on, though the shutter can also be used to take non-critically timed `snapshot' exposures. The camera will be installed and removed from the telescope with a special cart/handling fixture which also carries the "saddle" upon which are mounted a pair of 10-liter liquid nitrogen dewars and the power supplies for the camera. The camera is mounted to the telescope with a trefoil kinematic mount which is also shared with the plug-plate cartridges. The camera side of this mount is carried by a steel ring girder which provides the basic structure for the camera and a stable attachment point for the cell for the final corrector, which is the optical substrate upon which the sensors and filters are mounted, as will be described in greater detail below.

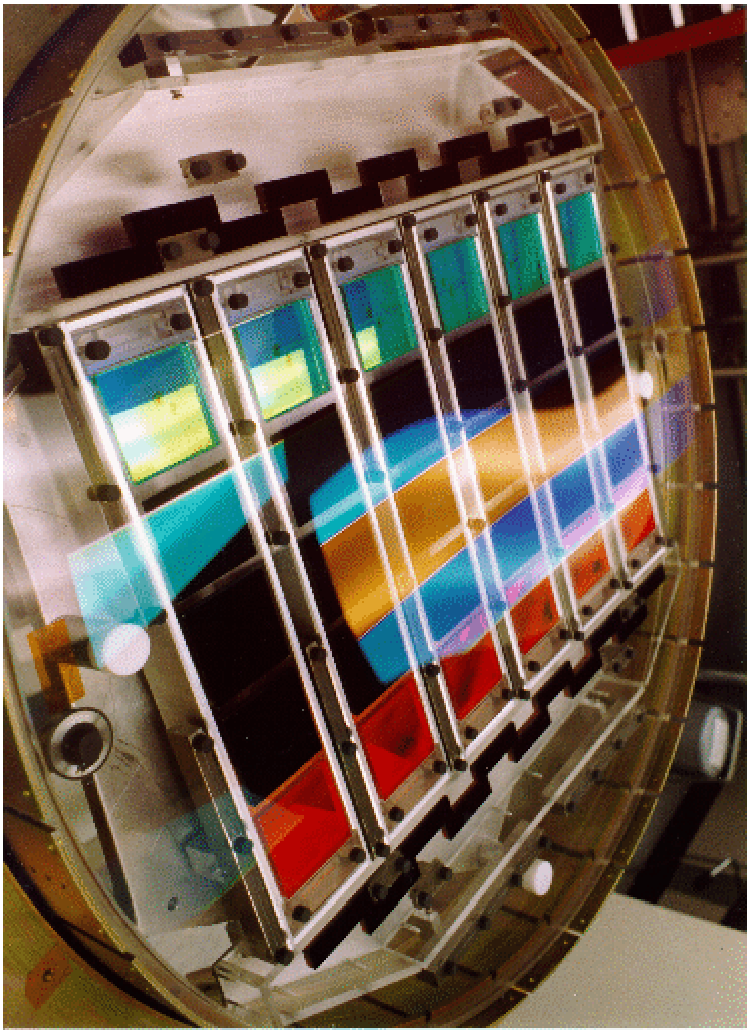
The CCDs for the photometric array will be housed in 6 long thin dewars (Figures 4.2 and 4.3) machined from aluminum blocks, each containing the 5 chips in one column. The CCDs are quiet enough to be run at -80°C, and will be kept at that temperature by an auto-fill liquid-nitrogen system which will be described in the next section. The optical system is fast enough and the focal plane big enough that mounting the chips and maintaining dimensional stability is potentially a very difficult problem. We seek to perform astrometry at the 30-milliarcsecond-per-coordinate level, which corresponds to 2 microns in a focal plane 600 mm in diameter. We have elected to solve this problem in a rather unusual but, we think, very satisfactory manner. The final corrector in the optical system is a quite thick piece of fused quartz with a flat rear face, 45 mm thick in the center and some 8 mm thicker at the rim, and we use this element as the mechanical substrate to which all the CCDs are registered and all the dewars attached. The corrector is thus both a mounting substrate and window for the camera dewars. Quartz is strong, reasonably stiff, has excellent dimensional stability and very small thermal expansion coefficient, and the small mechanical deflections associated with loading it with the camera have completely negligible impact on its optical performance. Figure 4.2 shows the front view of the whole assembly as it would look with the shutters removed, through the final corrector. Figure 4.4 shows the assembled corrector plate, mounted in its support, with the filters and the dewar mounting rails.

Figure 4.3



A disassembled photometric dewar. The CCDs are mounted on the invar optical bench (foreground).

Figure 4.4

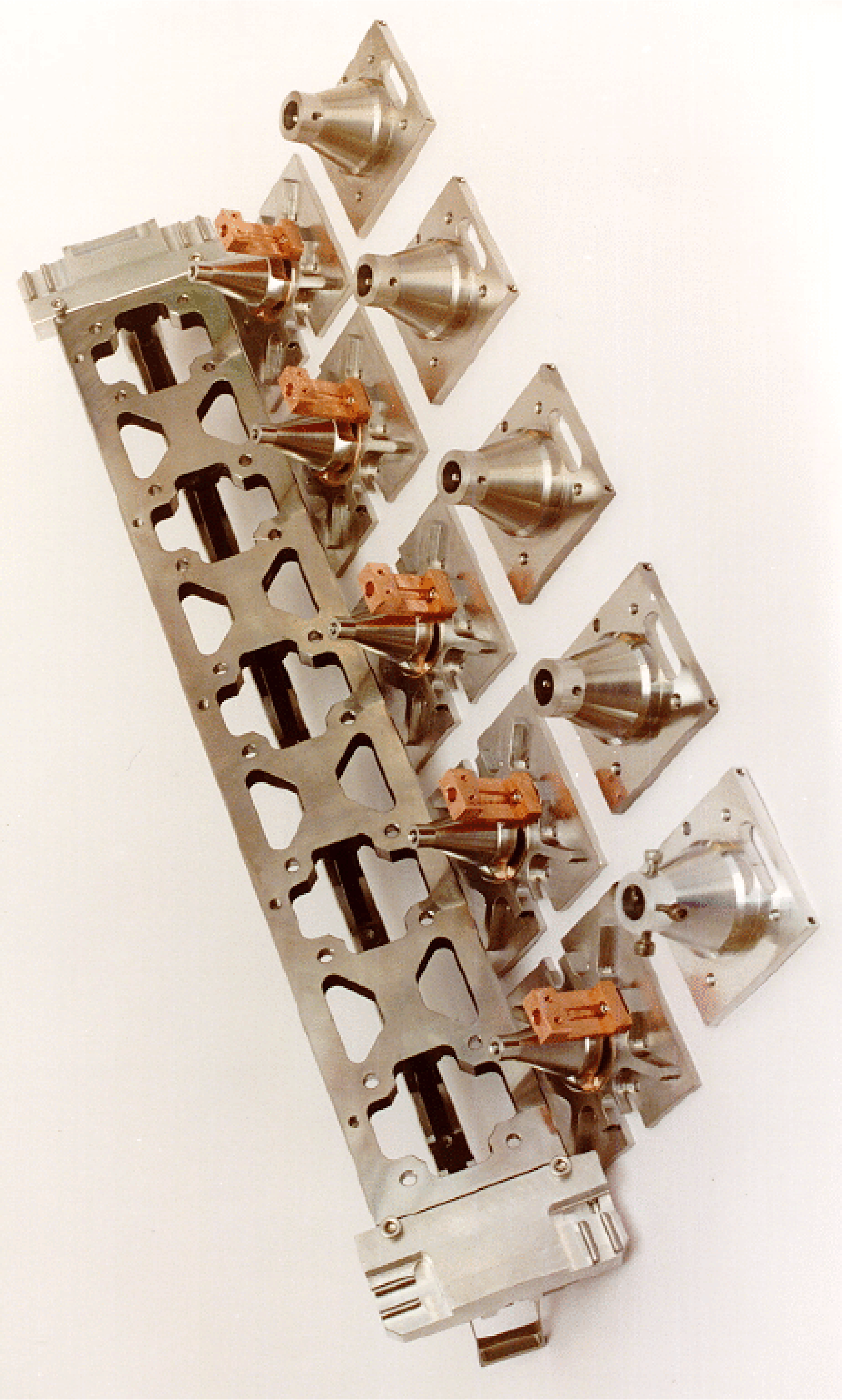


The face plate (photometric corrector) of the photometric camera. The plate has been mounted in its steel cell. The filters from top to bottom are g' , z' , u' , i' , and r' . The astrometric r' filters are above and below the photometric array, though what is seen is the neutral density filters associated with those. The quartz kinematic mounts for the photometrics and invar ones for the astrometrics can be seen, as can the invar rails for mounting the dewar bodies and the screw anchors in the quartz for both. The boxes occupying the spaces the dewars normally reside in are just protective covers for handling.

The CCDs in the column are mounted 65 mm center-to center, which leaves 1.5 mm gaps between the 63 mm square Kovar packages in which the devices are mounted. There is a similar gap between the edges of the packages and the sidewalls of the dewars. The CCDs are individually cemented to the ball element of an invar-36 ball-and-socket mount whose center of curvature is at the surface of the chip (Figures 4.3, 4.5, and 4.7). These mounts provide tilt adjustment via a set of three push-push screws on each. These assemblies are in turn mounted on a `T'-shaped invar-36 optical bench (Figure 4.5) in a manner which allows small independent rotation of the chips and shimming for piston. The optical benches, one per CCD column, are mounted to the quartz corrector by a kinematic mount which consists of a quartz column bonded and screwed to the corrector, and a set of four ball-and-double-rod pads. On one end, the arrangement consists of two balls on sets of parallel ways, one parallel to the long axis of the bench and the other perpendicular, which locates that end in both dimensions to within small rotations. On the other, one ball rests in a set of parallel ways parallel to the bench, which fixes the rotation but is free to move along the bench, and the other in sets of *mutually perpendicular* ways, which is completely free to *slide* in the plane. The bench is sufficiently flexible in torsion that the four vertical constraints can be mated independently with quite reasonable dimensional tolerances and forces (50N) on the balls, and in fact *needs* the four-point support for torsional stability. The 3/16" balls are made of titanium, which is tough, combines reasonably good Young's modulus and reasonably low thermal conductivity, and the 3/32" rods of hard stainless steel. The conductive losses for each ball joint are about 0.5 watt, with the 100°C temperature drop shared roughly equally between the ball joint and the quartz pillar. We were quite amazed that the measured conductivity of the ball-rod joint was within about twenty percent of the calculated value for a wide variety of materials we tried; the scheme appears to work very well. The one disadvantage we know about for this scheme is that the stresses on the balls and rods are very high owing to the tiny contact area. There is no danger of failure at the static stress levels, but dynamic loading associated with handling could easily permanently deform either member. To avoid this, each dewar uses a set of four electroformed nickel bellows which are pressurized at 40 PSI to bring the optical benches into contact with the kinematic mounts for observing. When the camera is being mounted or otherwise moved, this pressure is relieved and springs retract the optical benches about 1 mm and latch them away from the kinematic mounts.

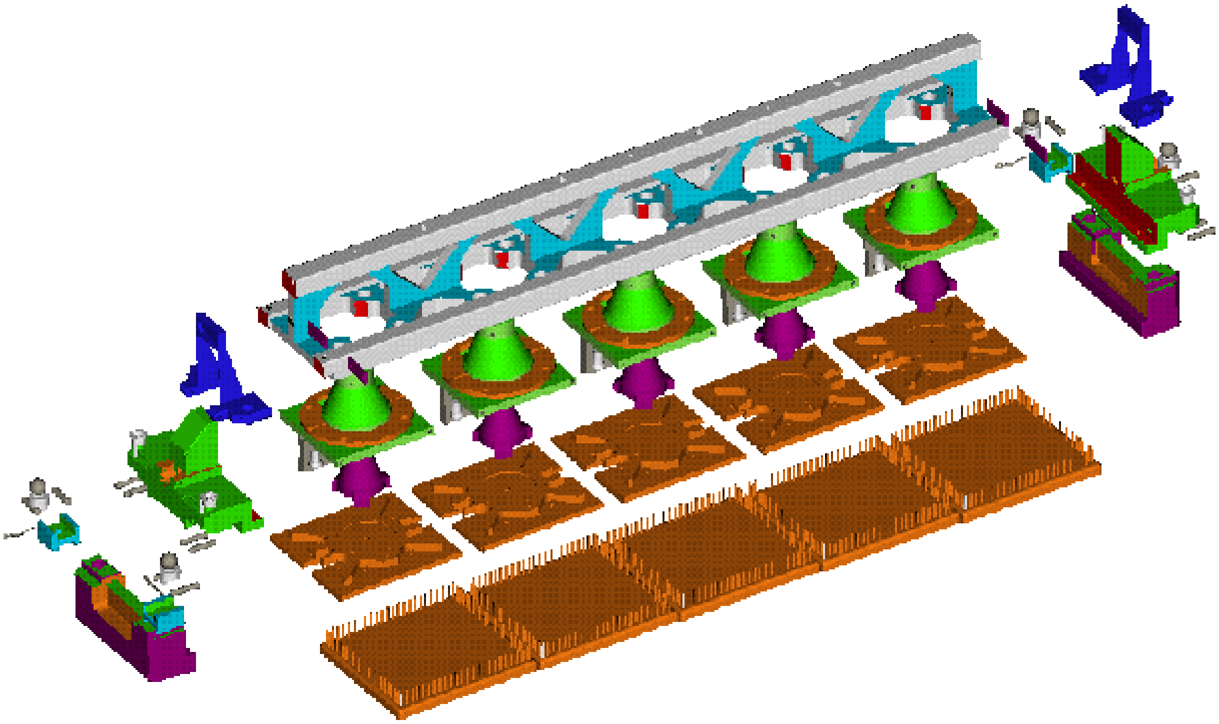
The optical benches (Figure 4.5) are quite stiff in bending. Loaded with their own mass and that of the CCDs with their ball-and-socket mounts their mass is about 2.7 kg and deflect about 1.3 μ in the focus direction and 1.0 μ in the focal plane under worst-case gravity loading. The kinematic mounts deflect a similar amount. Since the loading changes are very slow, the contribution to the astrometric error from these deflections should be negligible. The overall deflection of the corrector in the focus direction (which is the *only* direction there are appreciable deflections) is about 2 microns neglecting the stiffening by the dewar bodies. At worst focus changes induce centroid motion a factor of 20 smaller, so there is no appreciable error from this source. Since the dewars in this design are simply vacuum enclosures and, except for the preload bellows pushrods, do not even contact the optical benches, the load paths are very direct from the kinematic mounts to the telescope structure. The only tricky part of the design is the fact that there was a fair amount of (simple) machining to do on the quartz corrector. There are about 100 holes for screw anchors for the kinematic columns and the dewar mounting rails. The screw anchors consist of knurled brass inserts epoxied into these holes. The whole process went without mishap, though there was a fair amount of anxiety, as there was in applying the striped coatings.

Figure 4.5



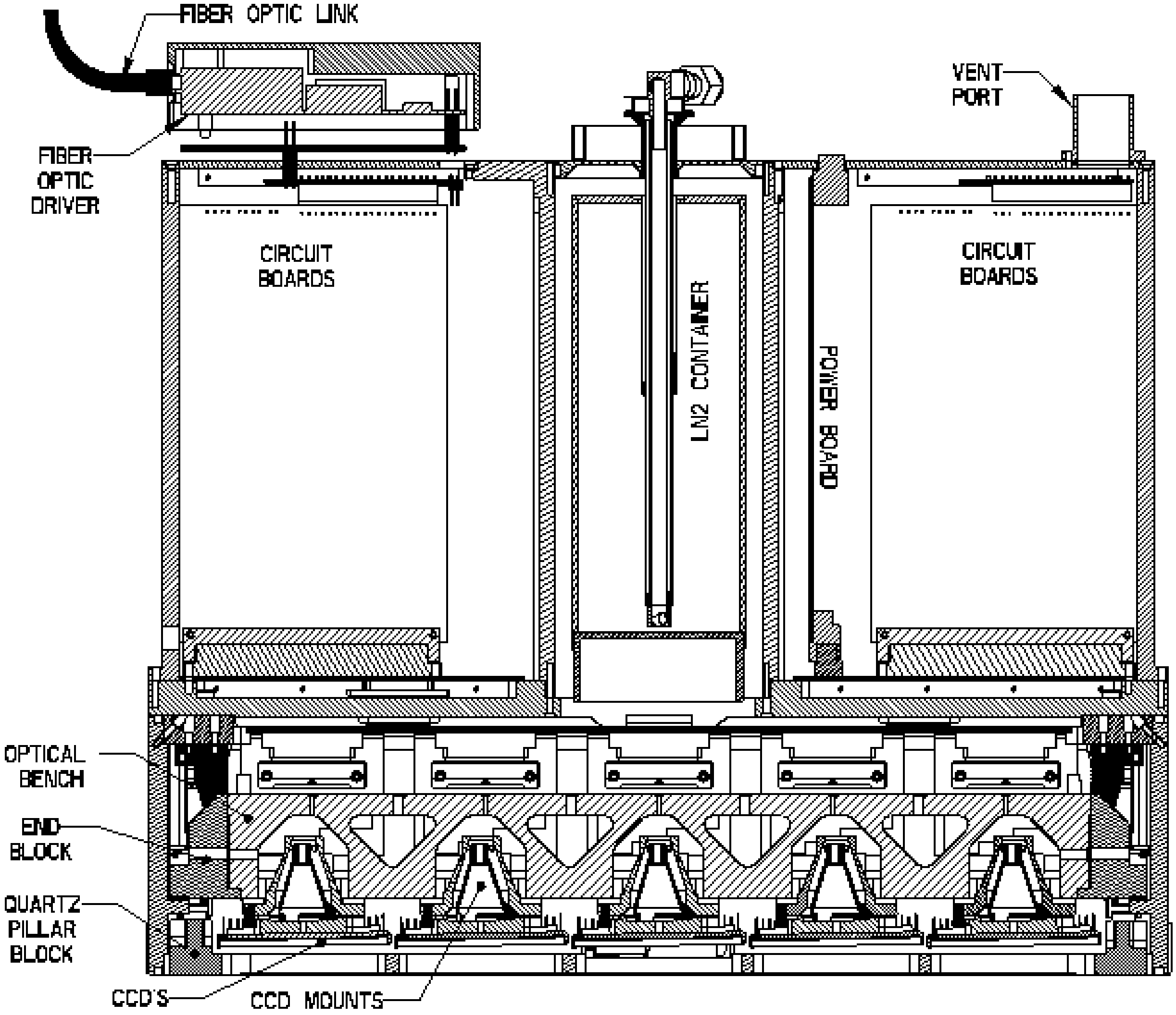
Mounting the CCDs. The picture shows an invar optical bench, CCD mounts and the ball-and socket joints for tilt, rotation and focus of the CCDs. The copper contacts remove the heat load to the LN_2 reservoirs via silver conducting straps. Note the setup of the kinematic mounting pins; the set in the upper left is the floating set, which mates with a pair at right angles on the quartz post.

Figure 4.6



A CADD-generated exploded view of the optical bench assembly. The picture shows the bench, the kinematic mounts, and the CCD mountings. The ball mount which is cemented to the CCD package is itself a vacuum-brazed structure consisting of the stiffening plate (gold) and the ball (magenta).

Figure 4.7



A central section side view of the photometric dewar. The locations of the liquid nitrogen reservoir and the electronics are shown.

It is, of course, necessary to adjust the tilt, rotation, and focus of each CCD fairly exquisitely; we allow 25 microns tilt error, 5 microns rotation error, and 25 microns total piston error. The tolerances on the absolute x,y location of the CCD are not severe, since there is no really significant software advantage to having objects centered on the same pixel on successive chips in the column and there is a 150-pixel overlap between successive TDI strips, but in fact the CCD assemblies and the machining and mounting of the optical benches has been done with sufficient precision that we expect the chips to be located within about 10 microns of their nominal positions. (We will not know for sure until the first TDI data on the sky is in hand). The tilt adjustment will be done on initial assembly of the ball and socket joint to theoretical optical design values. It can be changed on the basis of later tests but only with some difficulty; we feel, however, that it is the most reliable of the adjustments to predict. The rotation is much the most critical and will be done by the installation of a rigid lever on the socket part of the assembled ball and socket which extends some distance perpendicular to the optical bench. The optical bench will then be mounted accurately parallel to the ways of a precision linear slide and several reference points manufactured into the CCD which are exactly aligned with the CCD columns will then be examined with an inspection microscope as the optical bench assembly is moved along the slide. The rotational correction is then calculated and imposed by moving the lever with a micrometer screw. The lever arm translates the 5 micron tolerance over the 5 cm length of the CCD column to about 50 microns, which is easily resolved by the micrometer actuator. The socket must fit into the receiving hole in the optical bench to quite high precision, but the tolerances are not in fact very difficult with modern techniques, and were easily met in manufacture. Piston will be originally set with a set of shims to match the design focal plane. We expect to have to trim both the rotation and focus at the mountain, hopefully only once.

The dewar bodies themselves are machined of aluminum, and have O-ring seals to the quartz in front and to a fitted lid which carries the cooling system and electronics in back. They are about 75 mm tall and 330 mm long, so the atmospheric pressure on the sidewalls results in about 2500 N when they are evacuated. This is taken up by a lip on the lid of the dewar in back and in front by a frame machined integrally into the piece, which consists of horizontal bars between the filters. Thus the forces do not act on any dimensionally critical element. A similar force in the focus direction acts to seal the quartz to the dewar body, and demands that the face of dewar body be quite accurately flat in order that it not distort the rear surface of the corrector. This would have no optical consequence, but would change the effective shape of the focal plane. This strong bond with the dewars results in considerable stiffening of the corrector by the dewars, since the Young's moduli of aluminum alloys and quartz are similar; though the dewar walls are relatively thin (about 1 cm) they are twice as deep as the corrector is thick, and most of the stiffness is in the dewars. The great disparity of their thermal expansion coefficients, however, demands that the joint be a low-friction one; thin fluoropolymer gaskets appear adequate to the task. When the dewar is removed, fixtures attached to it bring the optical bench away with it for ease of maintenance. The optical bench can then be removed from the dewar through the top, carrying the in-dewar circuit boards (see Section 4.8 below) with it.

Cooling the Photometric Array

The dark currents for the Tek/SITe CCDs range from 30 to about 200 pA/cm² at 20°C. At -80 °C, this is reduced by about a factor of $3\text{x}10^5$, which for a 24 μ pixel becomes 0.01 electrons/second for 100 pA at room temperature. The sidereal-rate exposure time of 55 seconds thus yields a dark signal of about 5 electrons, which does not contribute significantly to the read noise. The total thermal losses for the dewar design we are using come to about 6 to 7 watts per dewar, three watts from radiation (mostly from the active detector area itself), 2 watts through the kinematic mounting, and another two from miscellaneous sources such as radiation to the nitrogen container, conductive losses through the flexible printed circuits (FPCs) connecting the CCDs to their support circuitry, conductive losses to the force actuators on the kinematic mounts, etc.

We will use an autofill LN2 cooling system patterned after the one used on the 4-shooter camera at Palomar (Gunn et al. 1987). Each dewar (Figure 4.7) will have a small LN2 container which holds about 200 ml of liquid, which will keep the detectors cool for about an hour. Under normal operating conditions all the LN2 dewars vent through a common vacuum-jacketed fill line and are kept under moderate pressure. There is a temperature sensor on the LN2 dewar body which is monitored continuously, and the dewar is filled when its temperature rises, but normally the dewars are filled on a schedule kept by the executive microprocessor (see below) which ensures that they never go empty. The fill is accomplished by opening all the solenoid valves on the individual vent lines and allowing liquid under pressure to enter; these are closed one by one as the dewars fill (as indicated by thermistor-based liquid sensors in the vent lines).

The heat is conducted from the chips through the kovar ball mounts cemented to them, and thence to copper posts which are connected via silver straps (silver is only slightly more conductive than copper but is nearly a factor of two more flexible at a given cross-section) to a cold-finger on the base of the LN2 container, which will also house a quantity of Zeolite molecular sieve getter. The very low heat conductivity of invar makes it difficult to keep the optical bench isothermal, but we will keep the chips at a uniform constant temperature with small make-up heaters associated with each CCD and mounted to the copper posts, as seen in Figure 4.5, which attach directly to the CCD mounts. The thermal expansion coefficient of invar is so small that small deviations from isothermality have little effect on the dimensional relationship in the bench. The temperatures will be measured with small platinum resistors fed with resistive dividers which result in reasonably accurately linear response over the relevant temperature range. Sensors of the same type are used to monitor the temperature of the LN2 container.

On the saddle which carries the power supplies (see below) are mounted two 10-liter intermediate supply dewars which supply the autofill system, each one supplying LN2 for one astrometric and three photometric dewars. These in turn are kept filled from a 160-liter dewar on the rotating floor of the enclosure, using an autofill system of rather different design.

The Astrometric/Focus Array: General

It is necessary to do pretty good relative astrometry simply to place the fibers; with the other problems such as differential refraction with wavelength and position, hole placement uncertainties, etc, it seems necessary to find positions to accuracies of the order of 200 mas (12 microns in the focal plane) or so. This is difficult to do with the imaging array, because it saturates at about 14th magnitude in the bands useful for astrometry (g' and r' , with r' preferred because of the smaller refraction corrections), and there are almost no astrometric standards at this brightness level. It would be in principle possible to calibrate the camera astrometrically, but one would have to depend on its stability and the stability of the telescope drives for relatively long periods to make use of the calibrations without suffering intolerable overheads. It would also be very nice from an astronomical point of view to do substantially better, and the astrometric CCDs illustrated in Figure 4.2 facilitate this.

They should allow tying the astrometric system of the survey eventually to the Hipparcos reference frame with an accuracy of order 30 mas locally, and if the survey could be redone in 10 years' time or so, would provide a wealth of proper motion data for faint sources. Even if there is no repeat, the utility provided by tying a faint QSO system to the Hipparcos one would be felt throughout astronomy. The astrometry is discussed more fully in Chapter 6 with respect to performance, Chapter 1 with respect to the survey strategy; here we address the hardware aspects.

It is a simple job to scale the existing 2048-square mask set to a nonsquare device with 2048 columns, so the astrometric CCDs are the same width as the photometric chips, but with fewer rows. A mask set for a 2048x 400 device was designed and produced by SITe, and two foundry runs of frontside devices made with it. We need a less tall device simply to fit into the available space above and below the photometric array, but there are other compelling reasons to favor such a geometry. We needed a number of these devices (24) comparable to the number of photometric chips, and it helped substantially that they were very much cheaper than the photometric CCDs; a device which is 2048 x 400 , which is the size we decided on, can be made several to a wafer, and is correspondingly cheaper than the one-to-a-wafer square device. Also, as we shall see, we require quite dense neutral filters to be able to observe available astrometric standards, and the smaller integration times associated with less tall chips allow us to use lower densities. The same consideration leads us to use less sensitive (by a factor of about 2) and much less expensive frontside illuminated CCDs. The disadvantage of the shorter columns is that the shorter integration times lead to larger position errors because of seeing-related image wander.

We have from these runs about thirty good devices. They have been mounted on precisely machined invar-36 headers of our design which allow the them to be mounted quite close together in the column (short) direction, as shown in Figure 4.2; the minimum distance is in fact determined by the necessary oversize of the filters to allow for the f/5 beam. The electrical signals to and from the chips are carried by flexible printed circuits of Kapton with very thin and narrow copper conductors. These FPCs are mounted permanently on the CCD headers and the chips are bonded out to pads on the FPCs.

For a chip 400 pixels high the integration time at the sidereal rate is about 11 seconds, and we argue in Chapter 6 that the one-dimensional positional accuracy achievable in that time is about 30-40 mas with the seeing conditions which must prevail during the imaging survey. These chips can be run somewhat warmer than the photometric imaging devices because we are not interested in low signal-to-noise objects and the integration time is short. Cooling to -60 °C is sufficient, which yields a background of about 10-20 electrons per pixel in the 11 second integration time for these devices. The centroid of a star which deposits about 2000 electrons can be measured to about 30 mas (shot and readout noise errors alone) in 1 arcsecond seeing with this background. A star which saturates in the central pixel ($4 \times 10^5 e^-$ for these devices) has a total signal of about 4×10^6 electrons, so the dynamic range for 30 mas accuracy, considering shot and read noise alone, is about 8.3 magnitudes.

We will use the astrometric sensors in the r' band, which seems optimum. With no further attenuation, they will saturate at about 11.3, and one can do 30 mas astrometry to about 19.5. This is not particularly useful, since the Hipparcos net and the AGK3 have very few objects as faint as 11.3. We will use 3.0 magnitude neutral filters, for a useful range of 8.3 to 16.5 This filtration results in an overlap of about 2.5 magnitudes between saturation of the imaging array (about 14.0), and the 30 mas accuracy limit for the astrometric array, and there are more than 200 stars per square degree in that magnitude interval near the pole. Thus the frame defined by the astrometric chips and the one defined by the red imaging chips (and with only a little more difficulty, that defined by any band in the imaging array) can be tied together very accurately indeed. Most of the Hipparcos stars are fainter than 7th, so more dynamic range would not really be useful.

We will use a total of 24 of these devices, one at each end of each column of photometric CCDs, 5 across the top and bottom of the array to tie the columns together, and two outside these to provide focus information. The sensors at the bottom provide a check on the tracking rate and direction; stars cross the bottom set 7.5 minutes after they cross the top, and they go through the red photometric sensors at 1.4 minutes. Thus drive errors with frequencies lower than 2 x 10^{-3} Hz can be corrected for, and we can monitor higher frequencies statistically. The astrometric and red sensors define a continuous frame which will drift slowly with respect to any initially defined absolute frame because we are basically measuring rates. We will use the astrometric standards only to pin this instrumental frame to the sky at intervals (see Chapter 6).

There are about 1.5 Hipparcos stars per square degree in our magnitude interval, and the array will encounter one every 70 seconds on average. If the drives can be held to the accuracy we would like (25 mas rms stochastic component in the frequency interval 2x 10^{-3} to 3x 10^{-1} Hz), as seems very likely from measured errors on the manufactured parts, tying the survey to the Hipparcos net to 30 mas should be possible; in any case, we should be limited to the errors imposed by seeing, which is our goal.

Mechanical Design of the Astrometric/Focus Array

The linear dimension associated with 30 mas is about 2 microns, and achieving dimensional stability to this order over a focal plane as large as ours is not an easy task. Invar-36 has a thermal coefficient of expansion of about 1.5x 10^{-6} /°C, and a one-degree temperature change induces a dimensional change over the 455 mm width of the array of 0.7 μ, so one-degree temperature control is adequate. The coefficient for silicon is similar, so changes on the scale of one chip with reasonable temperature control are negligible. Our approach to the astrometric array is thus very similar to that for the photometric; the chips are mounted on invar-36 headers (which are machined so that the curvature in the long dimension is similar to the curvature of the square photometric devices) with a thin compliant conductive epoxy film. These headers are mounted with screws onto an invar optical bench (which now, of course, goes *across* the array) on invar shims which are machined with the tilts necessary to fit the field curvature. The shorter vertical dimension means that rotation is not so critical, and shimming for piston will be the only adjustment normally performed. The optical bench is located to the corrector with kinematic mounts of the same sort as used in the photometric dewars, but here the pillars are much smaller, shorter, and are made of invar instead of quartz. (see Figure 4.2). The optical bench is housed in an aluminum dewar attached to the corrector in the same fashion as the photometric ones, again with an O-ring seal against the quartz. Reinforcing the dewar against transverse atmospheric pressure is trickier in this case than in the photometric case because there is no clear place to put a spreader bar close to the focal plane with the chips overlapping as they do, but spreaders behind the main horizontal web of the optical bench through holes in the vertical web serve here. This complicates the assembly somewhat, but not disastrously so.

The dimensional relationship between front and rear ranks of the astrometrics is also crucial to control, since that determines the accuracy with which errors in the drive rate and direction can be measured. This is greatly facilitated by our use of the quartz corrector as the metering substrate, and the dimensions should not be much less accurately maintained than those within one of the optical benches, particularly when one considers that the deflections associated with the kinematic mounts are the same for the front and rear benches (and furthermore quite similar to those for the photometric benches).

At an operating temperature of -60 °C the total heat load to the astrometric optical bench is about 6 watts, roughly the same as the photometric dewars. We have decided for simplicity and economy to use a cooling system essentially identical to the one adopted for the photometric dewars, and in fact one which shares many parts with its photometric counterpart. In this case we must endeavor to keep the bench as isothermal as possible.

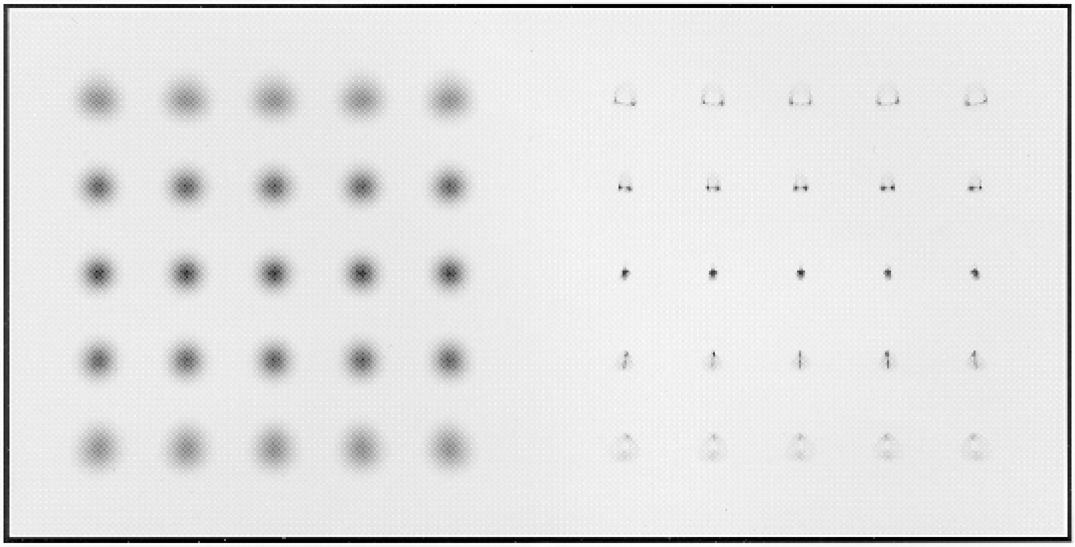
The bench has a cross-section of about 7 cm^2, and the conductivity of invar is about 0.2 watt/°C-cm, so a flux of 1 watt in the bench is associated with a temperature gradient of about 0.7° per cm. Temperature *differences* across the bench of 1 °C induce a bow in the bench with an amplitude of about 1μ. We will distribute the heat load from the bench to a heavy cold bar attached to the LN2 container's coldfinger at four points along the optical bench, again using thin silver straps. This will result in temperature inhomogeneities of the order of 1 degree, but in a pattern which should remain quite stable and hence innocuous astrometrically.

Focus

It is clear that one of the major contributions to image degradation in normal observing circumstances is the inability to keep up with focus changes brought about by flexure and temperature changes, and we will address this problem at the outset. Our requirements are especially severe because of the astrometry, but it appears that a simple scheme will suffice to provide excellent control.

There are two astrometric-type sensors housed in the astrometric dewars which will be used as focus sensors; though they are quite far from the center, the image quality is quite good at their location. They will have only an r' filter with no neutral density; this filter is cut into three parts, and the optical thickness associated with the neutral filter on the other sensors is taken up by three clear glass spacers of varying thickness. When the the rest of the array is in focus, the center of the focus chip is also, but the ends are, respectively, 300 microns inside and outside of focus; this defocus, which results in image degradation from defocus comparable to the expected 1 arcsecond seeing, is optimal from a focus-determination signal-to-noise point of view. Comparison of the images in the two outer thirds will provide a sensitive differential measure of the focus, which we will adjust dynamically; the resolution of the secondary motion is such that that should cause us no difficulty. Figure 4.8 shows images at five positions on the focus sensor through focus with 150 μ focus steps, as produced by the design optical system and convolved with 0.8-arcsecond Gaussian seeing.

Figure 4.8



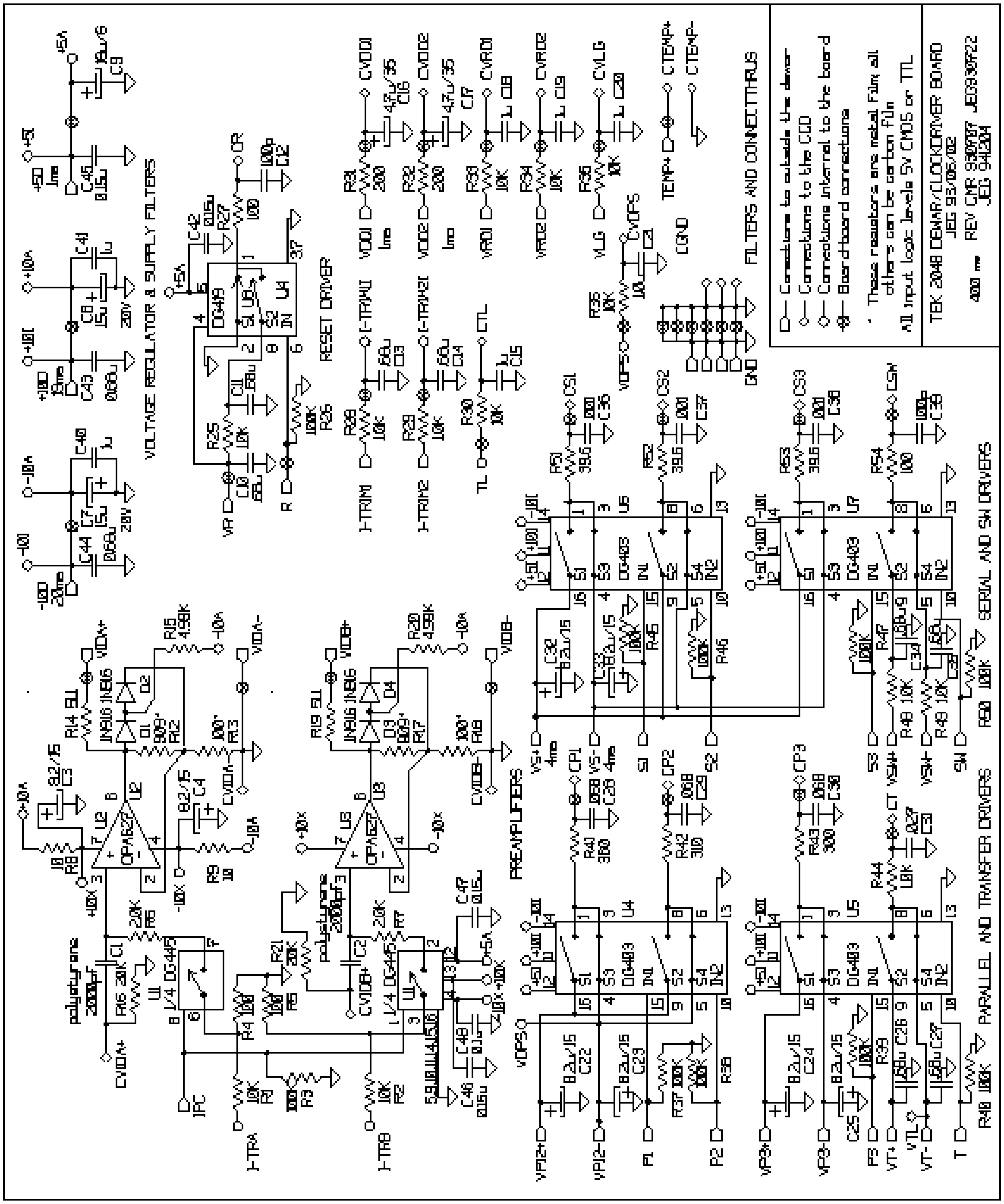
Focusing the camera. Images in the focus CCDs are shown at at five positions across the chip and at focal positions -300 μ, -150 μ, 0, +150 μ, and +300 μ referred to best focus. The left panel are the images as produced by the telescope, the right convolved with 0.8 arcsecond Gaussian seeing.

We need to control the focus very accurately to maintain astrometric accuracy. If we want 2 μ positional accuracy in the focal plane, we must control the focus to about 35 μ, since the maximum angle which the central ray makes to the focal plane is about 0.055 radian. Focus errors this large contribute only 7 μ RMS to the image diameters and are negligible, but we would like the focus errors to be negligible for the astrometric determinations as well. The factors which contribute to focus errors are residual aberrations across the field of the focus chips, photon statistics, and, of course, seeing. The focus errors from aberrations and photon statistics (the actual errors will almost certainly be dominated by *seeing*) are about 3 microns for any star brighter than about r' =15.5, the limit set by psf variations across the focus chip. For fainter objects, photon noise becomes important and the errors are about 6 microns at 17.0, 9 at 18.0, and 20 at 19.0. At the galactic pole, the star counts are about 250, 500, 820, and 1300 per square degree brighter than 15.5, 17, 18, and 19, respectively, and fainter than the saturation limit at r' =11.5. The three focus zones are each about 4 arcminutes wide, so the camera will see about 0.0167 square degrees per minute of time in each of the zones. Thus it will see 4 stars brighter than 15.5, 8 brighter than 17, 13 brighter than 18, and 21 brighter than 19 per minute. The timescale for focus changes is not known at this point, but the thermal time constant for the telescope is of the order of an hour, so it seems not unreasonable to expect that one can effectively average over a hundred to a few hundred stars to generate a focus signal, and the statistical errors will be completely negligible.

The Electronics

We will pack the clock drivers and the preamps in the dewars as we have always done in the Palomar systems to minimize noise (Gunn and Westphal 1981). There are monolithic FET-input operational amplifiers now (Burr-Brown OPA627s) which will serve as preamplifiers with noise performance almost equivalent to the more complex discrete ones we have used in the past, but with much better overall speed and settling behavior. The optical design has sufficiently good distortion characteristics, as we discussed in the last section, to allow clocking all the chips in the array synchronously, so the driver and digital control circuitry is also very simple. These two facts keep the in-dewar circuitry simple and compact, though the sheer number of sensors makes the system and particularly the interconnections rather complex. We have designed, produced, and tested the entire complement of surface-mount printed circuit cards for the dewar circuitry, which includes two small (43 mm x 53mm) piggy-back polyimide printed circuit cards, one for the preamp proper and the other for the clock drivers. The latter are CMOS analog switches with RC pulse-shaping circuitry on the inputs to the CCD gates (Figure 4.9). The CCDs are electrically connected to this assembly via a 25-conductor kapton flexible printed circuit with very thin copper conductors. This FPC terminates in a connector on the preamp board and is soldered to a polyimide PC card which acts as a socket for the CCD on the other end.

Figure 4.9

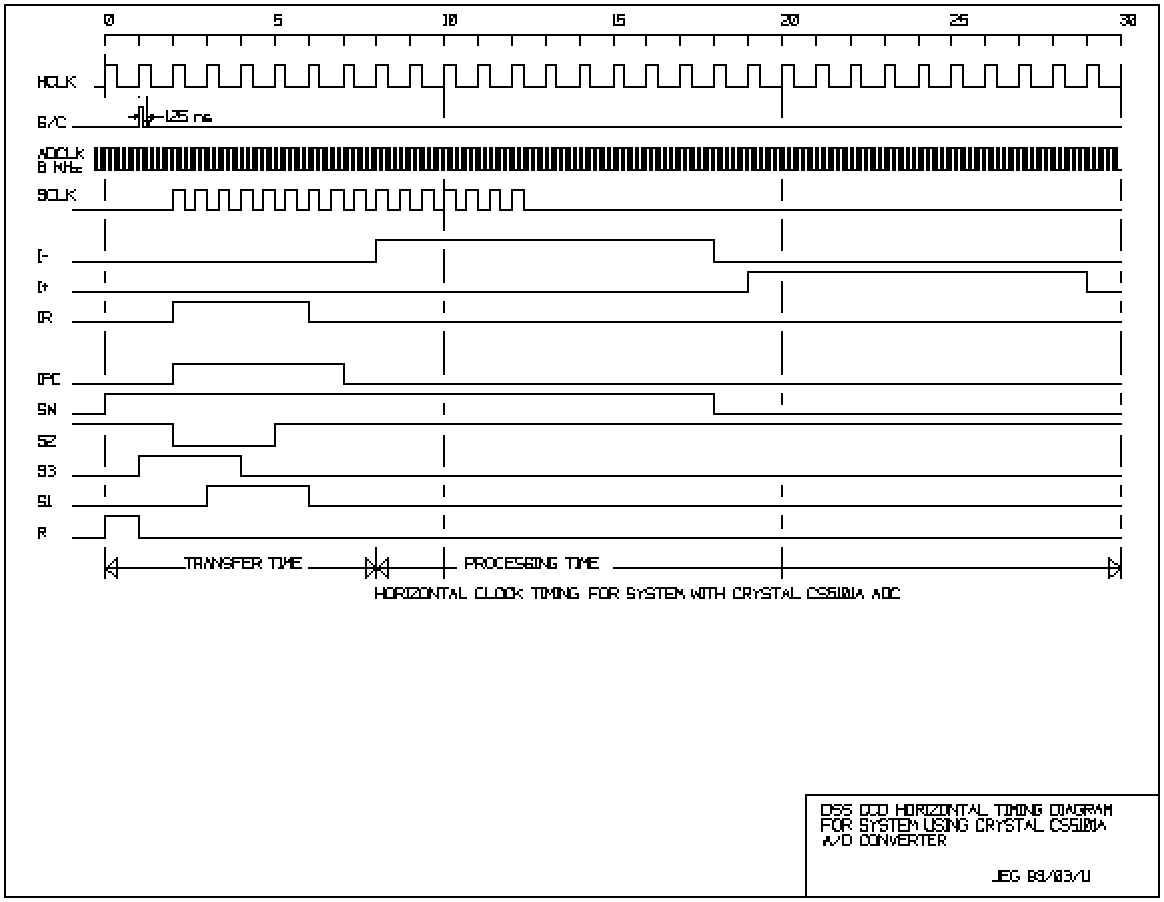


Schematic of the in-dewar electronics. Included are the dual preamplifier and the CCD clock drivers. This is implemented as two small piggybacked surface-mount cards which are mounted on the dewar wall.

Each CCD is associated with either one or two analog signal chain channels; the Tek/SITe chips have split serial registers and we originally hoped that we would be able to obtain all our devices with two good amplifiers on at least one of the serial registers. This would have allowed us to use rather slow readout, with correspondingly good noise performance. As it turned out, it was expedient to accept a small number of devices (6) with only one good amplifier, which we have to run twice as fast. The noise penalty is only about a factor of 1.3, and in most cases is not a serious problem. We strongly considered running *all* of the devices with one amplifier, but unfortunately not all the devices will run fast enough. Thus the electronics is not quite synchronous, there being two serial clocking schemes, one running exactly twice as fast as the other but both generated from a single master clock and locked in phase.

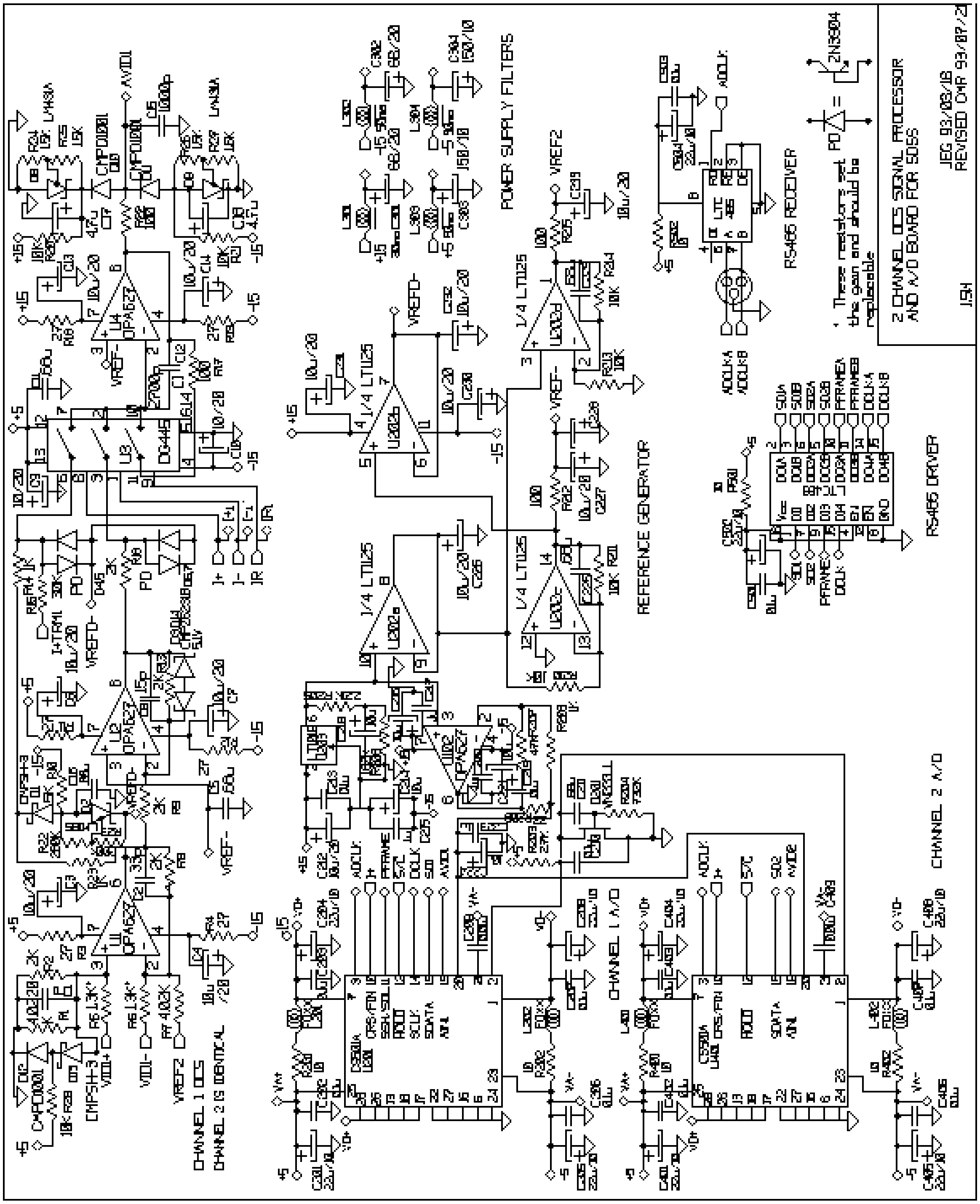
The architecture of the serial registers on the devices is such that there are 20 `extended' pixels between the edge of the imaging array and the on-chip source follower amplifier, and we will read another 20 overscan pixels in the two-amplifier chips, so each half-row will consist of 1024 + 20 + 20 = 1064 pixels. The single-amplifier chips will have the 20 leading and trailing extended pixels at either end of the data pixels and will in addition have 40 overscan pixels at the end. At sidereal scanning rate, the scale of 3.643mm/arcminute corresponds to 38.05 lines/sec, or 26.28 ms/line. The vertical transfers require about 700 μs, leaving 25.58 ms per line. At 24 μs per pixel, the time for 1064 pixels is 25.54 ms. (The fast channels run at 12 μs per pixel, clearly.) We need about 32 master clock cycles per pixel for all the functions in a classical dual-slope double-correlated sampling system, so a 1.3333 MHz (750 ns) clock yields 24-microsecond pixels. This will be derived from an 8 MHz master clock, which is needed by the A/D converters and the data transmitters.

Figure 4.11



Timing and CCD control. The figure shows the timing diagram for the horizontal charge transfer, double-correlated sampling control, and analog-to-digital conversion for one channel of a double-amplifier chip. The pixel clock is 1.3333 MHz, derived from the 8 MHz master clock which also runs the internal conversion circuitry on the Crystal CS5101A converter.

Figure 4.10



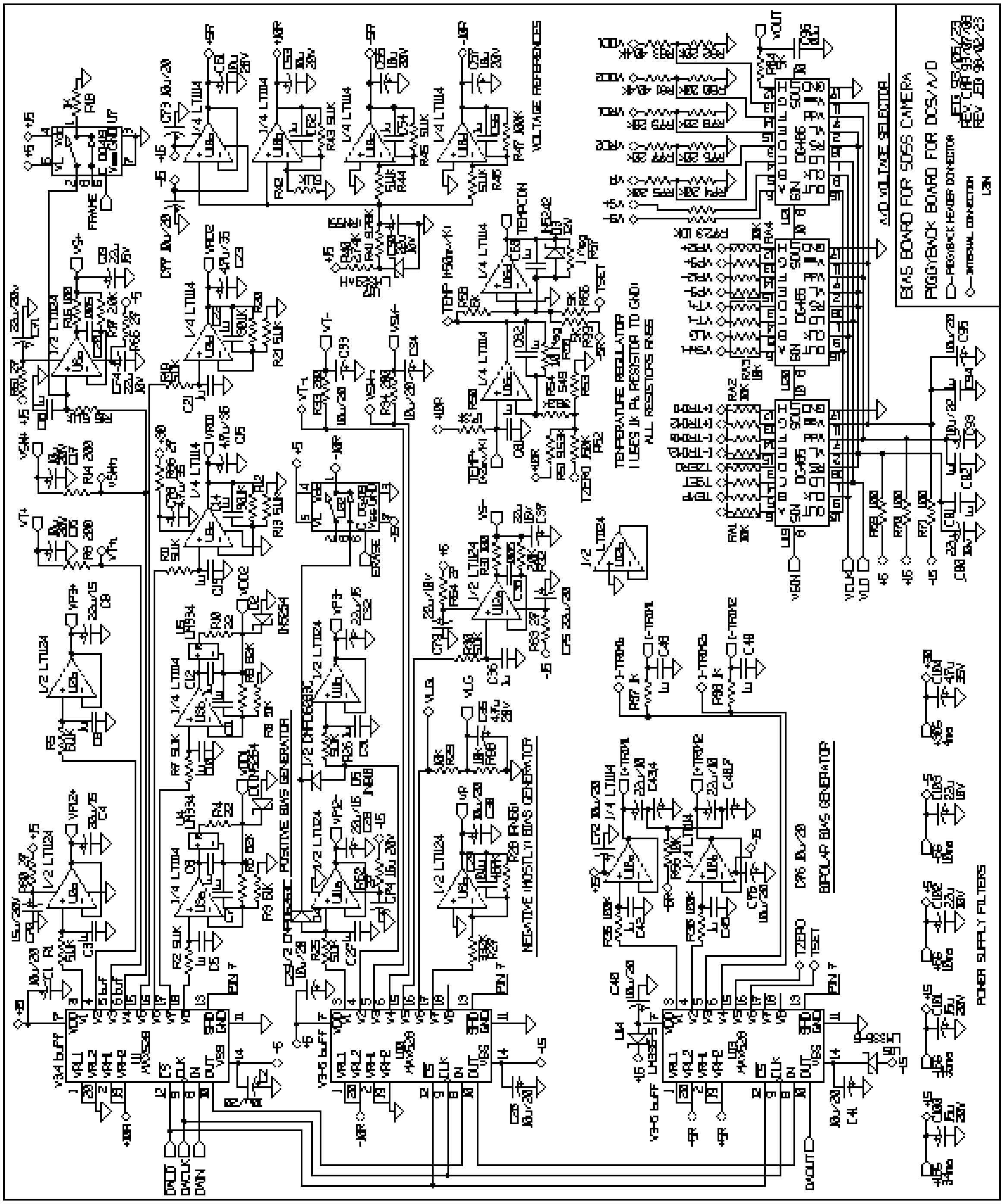
A schematic of the dual signal chain and A/D board.

The preamplifiers are AC coupled to the output of the CCDs through a more-or-less classical `back-porch' clamp which has no time-constant effects during the critical sampling period. The coupling through the rest of the system is DC. The output of the preamplifiers goes to a simple dual-slope integrator which is also used as a hold amplifier for the input to the A/Ds. The latter are Crystal Semiconductor CS5101As, which convert to 16 bits in 8 μs with an 8 MHz input clock (fast enough for both the slow and fast channels) and which have internal calibration circuitry which keeps the internal accuracy at ±¼ LSB. The circuitry for this is illustrated in Figure 4.10, and the timing diagram in Figure 4.11.

There is some question about whether 16 bits is enough (probably academic, since at the moment more accurate converters in this speed range are not readily available), but the problem is not severe. In the g' , r' , and i' bands the sky levels are such that digitizing at roughly 5 electrons/(AD unit) results in a step which is small compared to the standard deviation in the background, and accommodates with 16 bits the full 300,000 electron full well of the device. In the u' chips the situation is not quite so good. The sky in each chip is only of order 50 electrons, and the noise with 5 electrons read noise per chip is 9. If we demand that we digitize at at least two levels per unit sigma, this will result in a digital full well of slightly less than the physical full well of the CCD. This is not a serious problem, however, since the fluxes and efficiencies are such that no object which does not saturate in one of the other bands is likely to overflow the u' A/D.

The CCD voltages and the voltages required to trim the dual-slope circuitry are generated by a set of three octal 8-bit DACs, Maxim 528s, for each chip. These converters have serial input, and all the converters in the system are daisy-chained to be set up with one very long serial word. The output voltages are sampled and placed on an output bus for measurement with a set of serial-input 8-position analog switches. The circuit for this voltage generator and monitor is shown in Figure 4.12. The signal chain electronics and this circuit are implemented in surface-mount technology on opposite sides of one small 8-layer Eurocard, all of which again have been constructed and tested. There is one of these cards per chip on the photometric camera. Some of these cards, associated with the single-amplifier chips, use only one signal channel, but all the cards are identical except for the input resistors which are used to trim the gains to match the individual output amplifier gains for each CCD, and the integrator capacitor, which is a different value for the fast single-amplifier circuitry. Thus each photometric dewar contains 5 of these cards, plus one card which is basically a bus receiver for the clock signals described in the next paragraph; the astrometrics employ 6 plus a bus receiver; *all* of the astrometric chips run with only one amplifier to save cards and space, and so we run two chips with each signal-chain-bias card -- the performance penalty is negligible.

Figure 4.12



A schematic for the voltage generator board for the SDSS cameras. All the voltages are generated by DACs and the board also has provision for placing any voltage on a measurement bus for monitoring.

The clock signals are generated from programs stored in EPROMs and controlled by a Forth microcontroller using a Hitachi H8/532 CPU from Triangle Data Systems. One controller runs the entire camera; all CCDs are clocked synchronously with identical clocking waveforms (though there are two sets of serial waveforms for the `fast' and `slow' chips.) The voltage rails for all voltages can be optimized for each chip. The controller has two such CPUs, one of which is busy continuously generating the CCD signals, and the other of which (the `executive') is keeping track of other housekeeping chores, such as control of the shutters, initial DAC voltage setup, voltage and temperature monitoring (the micro has an on-chip 10-bit A/D), etc. We monitor all the operating voltages (the CCD temperatures are part of this set) on a round-robin basis, measuring one voltage per line time. There are 22 voltages for each of 54 CCDs, so the process requires about 32 seconds for one cycle. The ADC setting requires a word a little more than 10000 bits long, which will have a leading checkbyte which is received back at the transmitter after making the entire rounds. The clock will be slow, 10kHz, and so the setup time is 1 second. All signals are sent from the controller by balanced RS485 transmitters and received at each dewar by complementary receivers. All signals except the 8 MHz master clock are carried on a 100-conductor ribbon cable bus. The control for the camera comes over a single `RS232' fiber pair to the executive micro. There is considerable flexibility for controlling the chips with this scheme; we implement on-chip binning both horizontally and vertically (independently) and allow some control of the sampling interval. There is, in particular, a "quick" mode in which the CCD clocking is identical to that usually employed but which cuts the sampling time by a factor of five. This allows the chips to be read at twice the usual speed with a noise penalty of a factor of 2.3, which is useful for calibration exposures for the spectrographs (this mode cannot be used for the camera, because the single- channel chips cannot be run this quickly owing to the limitations of the ADCs). We will also have the capability to bin very heavily vertically, which will be used in spectrophotometric calibration for the spectrographs.

The Crystal A/Ds have a serial output, which is fed (again over RS485) to a pair of small cards mounted piggyback on the dewars which carry RS485 receivers, circuitry to format the data for the single-channel chips to resemble that for the dual-channel chips (in the time domain), a programmable logic chip which acts as a serial-to-parallel converter and multiplexer, and a fiber-optic FOXI transmitter to send the data to the DA system in the operations building. Each of these cards is capable of handing 12 slow data channels or 6 fast ones or any admixture. Each photometric dewar has one and each astrometric dewar 2, for a total of 10 data fibers for the 10 DA computers (see Chapter 10).

Each voltage-generator/signal chain card dissipates 2.8 W, the bus receiver and a voltage regulator-voltage distribution board about a watt each, so the photometric dewars dissipate about 16 watts each and the astrometrics about 20 watts. This waste heat is handled by the cooling system for the telescope. The camera enclosure acts as a semi-sealed chamber in which air is both generally circulated and blown specifically into the electronic enclosures associated with each dewar; the heat exchanged into the air is removed with two liquid-cooled heat exchangers on the fans. Water-glycol is provided to these exchangers at about two degrees below ambient, the goal being to keep the camera enclosure at ambient temperature. A similar scheme is used in the camera power supplies.

The power supplies are not mounted directly with the camera but are carried on a separate "saddle" which is mounted with, travels with, and is dismounted with the camera but is mounted to both the cart and the telescope mechanically and thermally independently. The saddle also carries two ten-liter intermediate "holding" dewars which supply LN2 to the small-capacity LN2 containers associated with each CCD dewar. Each dewar has its own power supply, each of which is implemented using small encapsulated linear supplies on a single 20x25 cm printed-circuit card. Each supply carries a fairly sophisticated monitoring and shutdown system mounted on a separate smaller PC card. The supplies are packaged in two identical chassis, each set serving one astrometric and three photometric dewars. There is in each chassis also an "auxiliary" supply to run the control micros, the LN2 system, the fans, solenoid valves, etc.

Thus the camera has a power cable which is made up of ten identical dewar power cables, an RS232 fiber pair, and ten data fibers as electrical connections to the outside world. It has in addition cooling water/glycol, a single vacuum-jacketed LN2 supply, pressurized N2 for the shutters, and pressurized N2 and vacuum for the dewar kinematic mount force actuators.

Identical analog electronics are used on the spectrographs; there there will be a controller per spectrograph, so each spectrograph has a data fiber, and since those CCDs are not used in TDI mode, there is no need for an executive micro. The spectrograph dewars are copies of the Palomar 4-shooter dewars and will employ the same autofill system as the photometric camera.

The CCDs

This whole project hinged quite visibly on the availability of many (42, including spares) 2048 x 2048 CCD sensors, at least 8 of which need excellent UV sensitivity and very low readout noise (< 5 electrons.) At the inception of the work more than ten years ago, it was by no means obvious that the chips could be obtainable. The project is really not viable in anything like its present conception without them, and though quite powerful surveys could be pursued with smaller detectors, we would either have given up a great deal which we believe is vital, have been stuck with an impossibly complex system (would you *really* like to align a focal plane with 120 1024 x 1024 CCDs and tend to their electronics?) or extend the survey time-to-completion unconscionably.

Table 4.2: The SDSS CCD Complement: Requirements

The Photometric Imagers: 2048x2048, 24 micron pixels

1. z' band (9200Å) unthinned frontside devices, no coating
   7 devices, noise < 20 electrons, qe >10% at 9200 Å.
2. i' band (7700Å) thinned backside devices, normal AR coat
   7 devices, noise < 20 electrons, qe >40% at 7700 Å.
3. r' band (6300Å) thinned backside devices, normal AR coat
   7 devices, noise < 15 electrons, qe >60% at 6300 Å.
4. g' band (4800Å) thinned backside devices, normal AR coat
   7 devices, noise < 10 electrons, qe >60% at 4800 Å.
5. u' band (3500Å) thinned backside devices, UV coat
   7 devices, noise < 5 electrons, qe >45% at 3500 Å.

The Spectroscopic Detectors: 2048x2048, 24 micron pixels

5 thinned backside devices, normal AR coated,
noise < 5 electrons,
qe>40 % at 4000 Å, >60 % at 6000 Å, >35 % at 9000 Å.

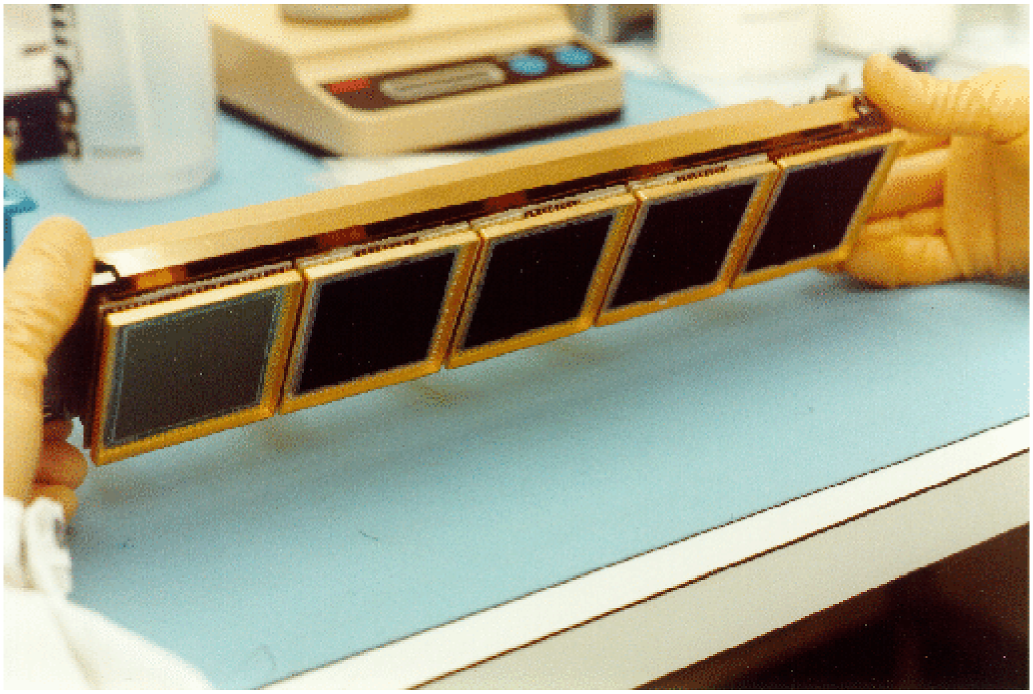
The Monitor Telescope Imager: 2048x2048, 24 micron pixels

1 thinned backside UV coated device,
noise < 15 electrons, like the u' imagers.

The Astrometric Imagers: 2048x400, 24 micron pixels

28 unthinned frontside devices, noise < 15 electrons.

Figure 4.13

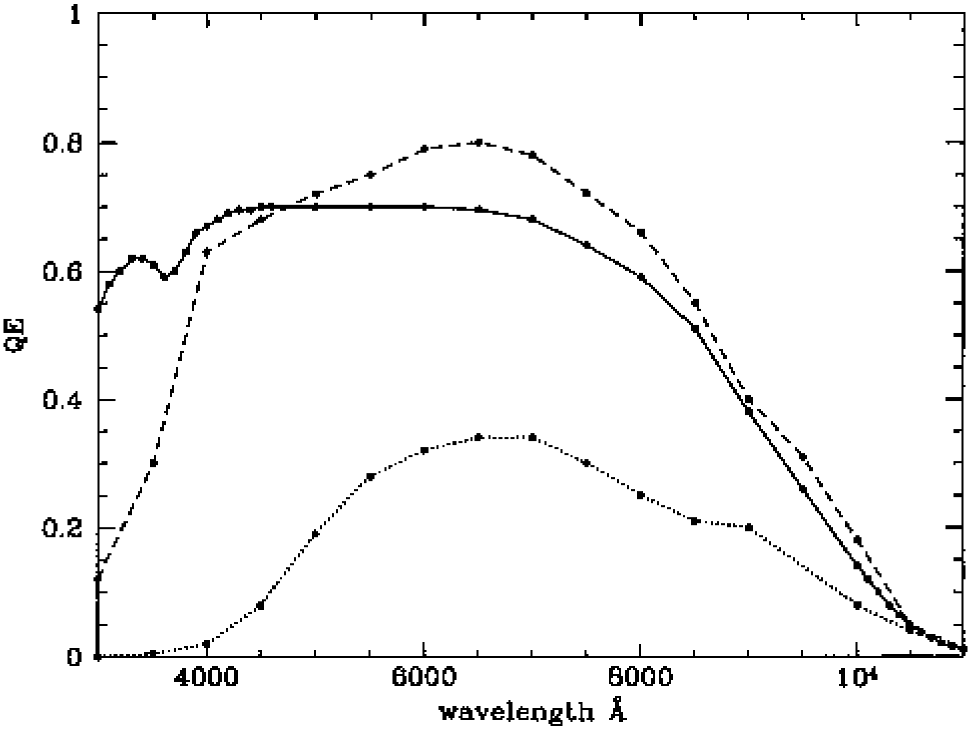


Five CCDs on an optical bench, with the gold plated heat shield in place.

The situation improved very markedly in the intervening years. Tektronix produced large numbers of 2048 x 2048 21-micron pixel devices as part of NASA's STIS program for the second-generation Hubble Space Telescope instrumentation, and went on to produced for our project a foundry run of excellent 24-micron devices, from which we were delivered 5 working and two excellent chips. Subsequent to that, we negotiated a contract for all the chips for the survey, photometric and astrometric/focus, with SITe, the now-independent Tektronix CCD division. Our cosmetic requirements are quite different from most of SITe's customers, since TDI gets rid of a whole suite of defects which would mar performance in normal imaging mode, and they agreed to deliver chips with special grading for us at a price somewhat below that for their normal Grade 1 devices. This arrangement was spectacularly successful, and we now have in hand and have fully characterized all the chips for the project (including the camera, the Monitor Telescope, and the spectrographs) including spares. Five CCDs, mounted on their optical bench, are shown in Figure 4.13. This picture shows the CCD complement for one of the six photometric dewars.

The charge-transfer problems which plagued the early large devices were solved and considerable progress made toward effective antireflection coatings and backside treatment for high stable quantum efficiencies in the blue and near ultraviolet. The engineers at SITe pursued two paths toward ultraviolet sensitivity; one of implantation techniques to keep the backside properly charged, and the other of coatings to reduce the severe reflection losses in the blue and near UV. They have consistently produced chips with better than 50% QE at 3500 Å. The quantum efficiency for thinned CCDs with the "standard" and the delivered UV AR coating are shown in Figure 4.14. One disappointment we have had is that this high UV QE declines substantially at low (operating) temperatures, but still is in the vicinity of 40% for all our devices at -80 °C at 3500 Å.

Figure 4.14



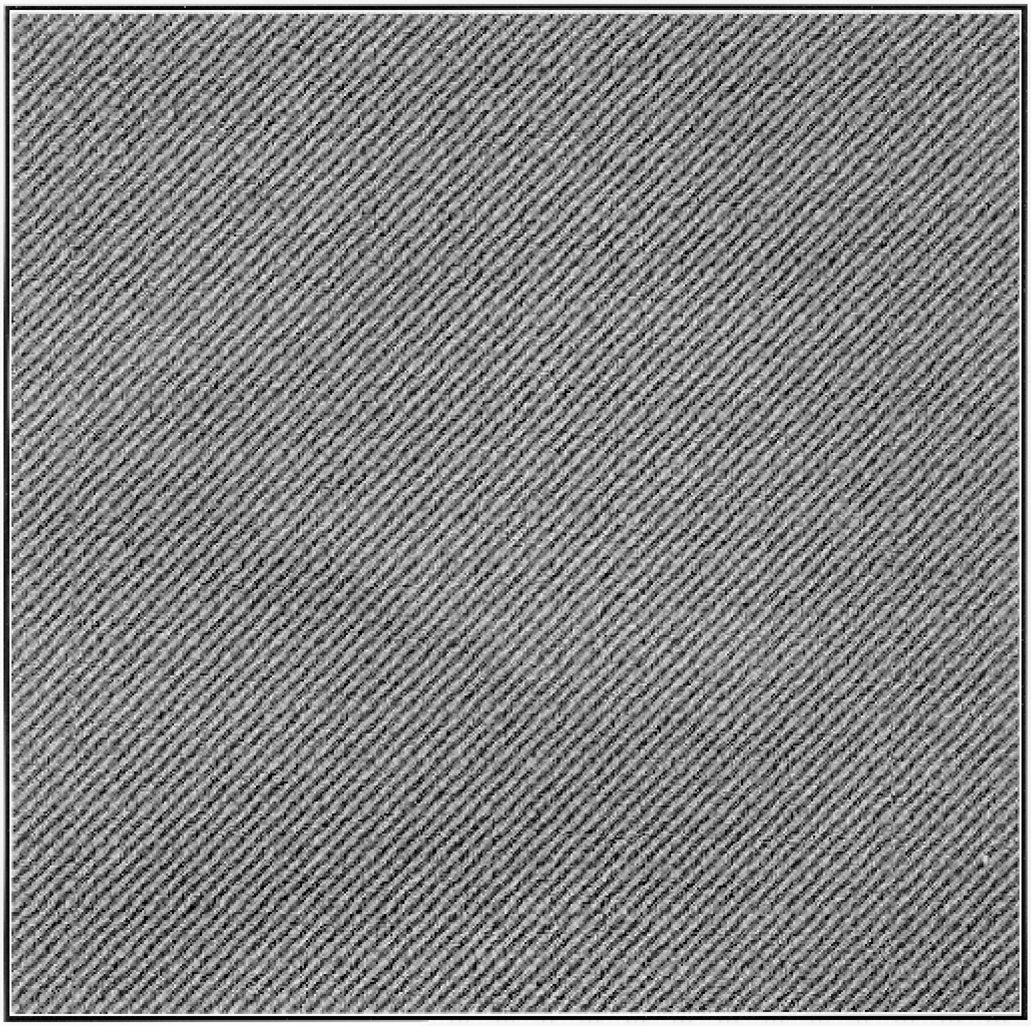
Quantum efficiency curves. Data are shown for the standard AR coating (dashed) and for the new UV AR coating (solid) on thinned SITE CCDs. The QE for the unthinned chips (dotted) is shown also for comparison.

The new chips have many of the latest `wrinkles'; there is a mini-channel in the columns for very good low-level charge transfer efficiency, the chips can be run in MPP mode (not, unfortunately, in the scanning array, since they are clocking all the time), and have amplifiers with very small geometry for the lowest possible noise.

The chips are designed and bonded out so that the four quadrants can be read independently, which for many applications results in a factor of 4 improvement in readout time; for us, it is the split in the serial direction which is useful, since we must integrate over the full chip vertically. It would be well to have all four amplifiers working for the spectroscopic chips, but that is sadly not the case for the chips in hand. The readout time for a two-amplifier chip is about 1 minute, or 30 seconds in the higher-noise `quick' mode, which can be used for spectroscopic flat fields and wavelength calibrations.

The devices were first characterized roughly at SITe and sent to Princeton for final evaluation. We have a cold test station which is controlled by a 486 PC which has electronics similar to the survey electronics. The system noise is about 1 electron, and we have the capability to measure CTE, quantum efficiency as a function of wavelength, uniformity, and also to test the vertical CTE in TDI mode using a special parallel-bar target and a flashlamp. The technique involves running the chip for some time with a uniform "sky" background of appropriate level, and then exposing the bar target, which consists of about 200 thin bright lines parallel to the rows of the CCD, with a flashlamp to impose a low-level signal. This frame of data is then captured as the chip continues to scan in TDI mode, and the bars are superposed to simulate a single such bar traveling along the chip following the charge packets. Doing this is important, because some low-level parallel traps which show up in single frames are satiated by the sky in TDI mode and disappear; other stronger traps permanently damage the CTE in the affected columns.

Figure 4.15



A subimage taken from a typical survey-grade CCD. The target is a 45° Ronchi ruling with a half-pitch of 3 pixels perpendicular to the rulings. The image has 30 electrons in the white bars, less than 3 in the black ones.

Our specifications allowed for single bad columns, the notion being that we are well-enough sampled that we can effectively interpolate across one bad column, but two or more contiguous ones disqualify a device from our consideration. Interpolation does, of course, affect the statistical accuracy of the photometry of objects affected by the column, but that is easy to keep track of and account for in the data reduction. The spectroscopic chips will be arranged so that the spectra are along columns, because the defects in CCDs are almost always along columns; thus even catastrophic column defects simply wipe out the spectra from a few fibers which we know about and can simply avoid using, and we have so *many*. Here, unlike the situation with the imaging array, we *prefer* for a given number of defects for them to be grouped, though here again the loss of a single column in a spectrum is not necessarily a disaster. This is but one example of our special requirements.

Partial blocked columns, for example, show up in TDI as columns with stably depressed quantum efficiency, and if the fraction of the column blocked is not too large has almost no effect. Likewise single hot column defects are integrated over and can be simply subtracted out if the noise contribution is not too great or interpolated over if it is.

Our chips all have good to excellent electrical performance, with amplifier noise varying between 2.5 and 10 electrons for all devices accepted for use in the three bluest bands, and a few with noise as high as 20 for the i' and z' bands, where the sky background will be about 1000 electrons per pixel. The median noise is near 5 electrons, and we measure CTEs of around 0.99998 horizontally and 0.99994 vertically at illumination levels of 30 electrons. This improves to about 0.999993 horizontally and 0.99998 vertically at 300 electrons, which is typical of all the expected backgrounds except for the ultraviolet. At the lower level, which is somewhat less than the ultraviolet sky level, the net transfer efficiency is 86 percent from the upper center of the chip (the pixel with the largest number of transfers). This represents a shift of 0.02 pixel horizontally and 0.07 pixel vertically (averaging over the column as TDI does) and completely negligible PSF degradation. For the higher-background chips these numbers are more like 0.007 pixel horizontally and .02 pixel vertically. The overall cosmetic uniformity is excellent as well, with RMS large-scale QE variations of about 7% in the blue and 4% in the red and infrared. The main defects we have seen are parallel traps of various strengths, many of which are strong enough to cause serious CTE degradation in the vertical direction with the backgrounds we are using, though again they are not fatal if they involve only one column. A subimage in the corner farthest from the amplifier taken with a quite typical device of a ronchi target with 30 electrons signal in the bright bars is shown in Figure 4.15, in which the excellent cosmetics and low-level CTE is seen.

We elected to take chips mounted in their standard kovar packages even though this led to significant mechanical difficulty in their mounting and cooling; demanding better packaging would have precluded culling our devices from a commercial production stream and would have prohibitively increased the cost. The problems incurred are fairly serious, however. The expansion coefficient of kovar matches silicon (and the substrate for the thinned devices, which *does* match silicon well) so poorly that the overall curvature of the devices, already serious at room temperature because of problems in high-temperature processing, is much worse at operating temperature. As mentioned above, the chips are convex toward the incoming light by about 230 microns center to corner, and that value roughly doubles cooling to -80 C. We have dealt with this both for the spectroscopic and photometric chips by cementing a heavy kovar stiffener to the back; in the case of the photometric chips, this is part of the ball tilt mounting. It was also necessary to build a precision measuring microscope to aid in the gluing of the photometric chips, since we wished to position the chips to pixel-or-subpixel accuracy and the chips are not mounted very accurately in their headers. We thus used reference points on the die itself to reference the CCD to its ball mount, and succeeded in doing so to an accuracy of about 3 microns RMS.

Our requirements for spectral sensitivity and noise vary a great deal across the array, though for the u' chips (7 with spares; we have spared the equivalent of a complete 5-chip dewar) we certainly require both good UV sensitivity *and* very low noise. The original requirements we imposed are summarized in Table 4.2. The chips we have received and will use are better than these specs in most respects. The operating temperature quantum efficiencies are well exceeded in all devices except those for the u' band; the mean qe for the z' chips is 14 %; for the i' , 60%; for the r' , 78%; for the g' , 70%; and for the u' , 38%; the dispersions in these numbers are small and probably dominated by measuring error. We have built in to the dewars UV floodlamps and can use them to enhance the UV quantum efficiencies to approximately the spec 45%, but whether we will use them or not is a complex operational issue, since the effects of UV flooding are not permanent and there is a significant dark current penalty. The noise specs are also all met except for the u' , where the worst figure is 6.4 electrons.

It goes almost without saying that the development of these devices for the survey has had tangible benefits for all of optical astronomy. Though the major thrust of wafer-scale large CCDs has now gone toward devices with more but smaller pixels resulting in slightly smaller devices overall, the production of these chips with high yield depends heavily on techniques learned here.

References

Gunn, J. E. and Westphal, J. A. 1981, *Proc SPIE* **290**, 16.

Gunn, J. E., Carr, M., Danielson, G. E., Lorenz, E. O., Lucinio, R., Nenow, V. E., Smith, J. D., and Westphal, J. A. 1987, *Optical Engineering*, **26**, 779.

Gunn, J. E. 1989, in *Clusters of Galaxies*, W. Oegerle, M. Fitchett, and L. Danly eds., proceedings of the third Space Telescope Symposium, Cambridge, 341.

Schneider D. P., Schmidt, M., and Gunn, J. E., 1989, *A. J.* **98**, 1951.

Turnrose, B. 1974, *P. A. S. P.* **86**, 545.

Photometric Calibration and the Monitor Telescope

Motivation for High Precision

The scientific returns of the imaging survey depend to a large extent on the accuracy of its photometric calibration. A 10-Terabyte multiband data set that covers pi steradians of the sky, in which there are 10^8 sources bright enough to be well measured, can be used to investigate exceedingly subtle phenomena, provided the photometry is adequate. The fundamental issue of the isotropy of the galaxy distribution over large angles depends in an essential way on how well the photometric zero point has been transferred across the sky, and on the proper evaluation of Galactic absorption. For example, Burstein and Heiles (1978, 1982) estimate the errors in their reddening maps to be at the level of delta E(B-V)=0.03 ; this variation alone would result in an error in the galaxy surface density of about 14%.

The Required Photometric Precision

The photometric accuracy is driven by three science requirements:

1. QSO Photometry: Experience with using existing color data to separate stars and QSOs is that systematic photometric errors in the colors should be no larger than about 5%.

2. Large-scale structure studies: Imagine scanning a single strip for the entire night(2.5° by 120°). The most demanding analysis that one can do is to divide the strip into two halves and measure the difference in galaxy counts between the two. For the spectroscopic sample (magnitude limit about 18 in r'), the rms scatter in this difference due to the known large scale clustering is expected to be at least 8%. The rms systematic photometric error should therefore be no larger than half of this, or 4% in galaxy counts; using a value of 0.5 for d log N/dm , this corresponds to a 3% rms error in photometry in the band used to select the galaxies. If we repeat the exercise for ALL galaxies identified by the imaging survey (which reaches much fainter than 20^m), the fluctuations due to clustering will be much weaker, perhaps less than 1%. Thus, we could make use of data with systematic errors smaller than 0.5% rms.

3. Interstellar reddening: The ultimate limit in analyzing galaxy counts is set by fluctuations in the interstellar reddening and extinction. The primary technique used to find the extinction is to measure the shifts in color of some class of objects and to use standard conversions from selective to total extinction. From published Gunn-Thuan photometry (Thuan and Gunn 1976; Schneider, Gunn and Hoessel 1983), we find that $A_r = 2.1\ E_{g-r}$, so that the colors must be measured twice as accurately as the desired accuracy of the extinction itself. Calibration of the Galactic extinction thus requires that color errors be less than about 1.5% rms for the analysis of the spectroscopic sample and 0.25% for the full imaging sample.

Regardless of the science one is trying to do, the limiting accuracy to which reddening and extinction can be measured is ultimately set by irreducible systematic errors in the technique used to measure the reddening. If we did nothing beyond using the Burstein-Heiles extinction map we would not be able to extract the full amount of information available in the imaging survey. We expect to investigate several techniques to improve on the existing reddening maps but the limiting accuracy is not knowable in advance. For example, one promising technique is to measure the g' - r' color distribution of stars at around magnitude 18; the halo stars form a strongly peaked distribution whose width is about 0.3 magnitudes and whose centroid can be measured to perhaps 0.03 magnitudes. The position of the peak will depend on both reddening and population variations (metallicity and age). If the latter effects can be calibrated, we will have a powerful technique for measuring the reddening. To do so with sufficient accuracy requires that the stellar colors should have systematic photometric errors of order 0.015 magnitudes.

After considering the above factors, we have set the requirements for the photometric calibration accuracy of the imaging survey to be no worse than 2% rms in magnitudes and colors averaged over degree sized areas, with a goal to achieve twice that accuracy.

Limits to Achieving Photometric Precision

Our requirement of 2% rms error ought to be easily achievable based on our and others experience with carrying out precision photometry. The limits to photometric accuracy are usually set by the accumulation of several small sources of error, each of comparable magnitude. Some of these are:

1. Atmospheric extinction: Accurate calibration will require the frequent observation of standard stars. Extinction variations on the order of a few percent can occur on hour time scales. The stability of the extinction at APO is not yet known.
2. There is an inherent irreproducibility for broadband flux measurements of sources with unknown spectral energy distributions f_{nu} obtained at different airmass. This arises because the quantity T_{nu} f_{nu} A_{nu} will integrate to different values, where $T_{nu}(t)$ is the instrumental sensitivity and $A_{nu}(t,X)$ is the atmospheric transmission at time t and airmass X . In practice this problem is minimized by using the colors themselves to deduce the general form of f_{nu}, iteratively if necessary, but there will still be a residual. The residual will be larger for galaxy photometry than for stellar photometry because of the greater variety of f_{nu} for redshifted galaxies. Synthetic models of star and galaxy spectra can be computed and folded through the SDSS filters to obtain at least a handle on the magnitude of this residual.
3. Instrumental calibration errors associated with flatfielding problems, etc.

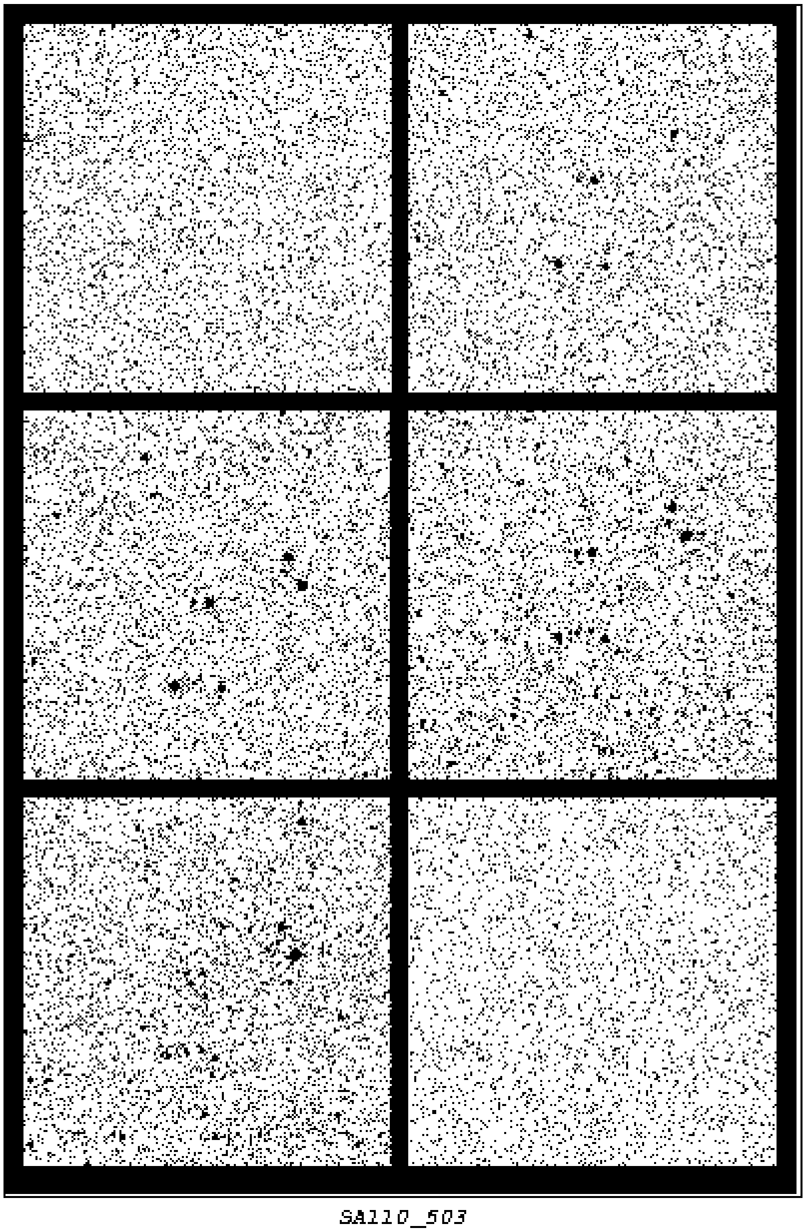
The Monitor Telescope

We will establish the primary photometric calibration for the survey by using a small robotic telescope at the 2.5-m survey telescope site. This instrument will observe continuously and automatically a set of photometric standards to calibrate the main survey.

The primary standard stars for the SDSS system are a grid of 20-30 bright, isolated stars in each of the northern and southern skies. All are photometric standards in current use, and a few have absolute calibrations. The energy distributions of these standard stars have been transferred to the SDSS filter system (Fukugita et al. 1996). A subset of the primary standards will be observed nightly by the Monitor Telescope, efficiently placing the survey telescope imaging data onto an absolute photometric system.

The telescope itself is a 0.61 m automatic telescope purchased commercially from the Autoscope Corporation. It is an f/10 Ritchey-Chrétien design with a Gascoigne corrector and field flattener to provide 0.6" images over a 40' diameter field of view. The telescope control system was supplied by Autoscope but it is being reworked for the survey. The data acquisition system hardware is provided by Fermilab. The camera is a slightly modified version of the spectroscopic camera electronics. The detector is a thinned Tektronix 2048 x 2048 CCD with 0.8" pixels and a 27' square field of view. The telescope and its dome will be capable of autonomous operation from a cold start for extended periods of time. The telescope and all instrumentation are installed at the site and have been undergoing commissioning tests for several months. Monitor telescope images of a standard star field are shown in Figure 5.1.

Figure 5.1



CCD images of the Landolt standard star field SAO110503. Five images were obtained with the monitor telescope in the SDSS u'g'r'i'z' filters. The sixth panel is intentionally left blank.

The Monitor Telescope will be used for several months in advance of the main Survey to observe extensively the primary standard star network and define the SDSS photometric system. During the survey, it will operate in parallel with the 2.5 m telescope, observing both primary standard stars and a set of secondary transfer fields that will be be scanned by the 2.5 m telescope. These data can be analyzed to provide the three necessary calibration coefficients needed to place the survey imaging data onto the SDSS photometric system.

The first coefficient is the atmospheric extinction. A minimum of 6 fundamental standard stars at a variety of airmasses will be observed each hour during the night in order to determine the extinction coefficients on an hourly basis.

The second coefficient is the photometric zeropoint for each 2.5 m photometric CCD. The Monitor Telescope will observe a grid of transfer star fields distributed throughout the SDSS survey area. We have selected a grid of 2392 fields, chosen to maximize the distance away from bright stars (see Figure 1.2). The Monitor Telescope observations calibrate the transfer fields so that when the main survey scans across them, we can use the known magnitudes in the fields to determine the zeropoints of the main survey photometry. Each transfer field must contain about 10 stars that do not saturate the SDSS camera but are measured with better than 1% accuracy rms by the Monitor Telescope. The zeropoints of the main survey may change with time because the field rotates with respect to the telescope pupil, so we will provide transfer fields for each scanline hourly. Each of the six scanlines will need its own transfer field. The Monitor Telescope will observe the transfer star field as nearly simultaneously with the main survey scan as feasible.

Last, there may be mismatches of the system efficiency as a function of wavelength between the main telescope and the Monitor Telescope because a single thinned Tektronix chip is used to calibrate all five bands (although the MT optics are coated to minimize this problem). This introduces color terms into the solutions. Experience has shown that such color terms change slowly with time (we will measure them once a week). The Monitor Telescope will observe secondary standard star patches known to contain a good distribution of star colors. Then, when convenient (for example, when there is some moonlight), all the CCDs in the SDSS camera will scan one of the color patches.

Second order coefficients (which account for the slight dependence of the atmospheric extinction on spectral energy distribution of a star or galaxy) will be included in the photometric solutions as well. We will try to measure them as well, although they are small enough that they can likely be computed in advance with sufficient accuracy.

A quick photometric solution will be computed in real time during the night. and provided to the observers conducting the main survey. This will allow them to monitor the quality of the night.

There are several special cases which need to be discussed. The u' band observations require special treatment as the stars that are bright enough to serve as standards in u' are saturated in the g'r'i' bands. We address this by allowing the random error to be higher in the u' band (2% instead of 1% ; this allows us to use fainter stars as standards) and by doubling the Monitor Telescope integration time on the the u' standards.

A second special case is the rapid variability in the near infrared sky, especially the water absorption bands. This will affect our ability to calibrate the i' and z' bands properly. We may choose to monitor the infrared sky on a nightly basis with two filters centered on the A band and an infrared H_2O band. Observations through these filters will allow us to apply color terms due to the changing atmosphere to the photometric solution.

A third special case is the spectrophotometric calibration of the spectroscopic data. This will be accomplished by the use of 10 narrow-band filters, 8 spanning the wavelength range of the spectrograph, the other two to measure the A band and a suitable infrared water band, probably the same filters described in the last paragraph. The primary standard stars with absolute photometric calibrations will be observed with this filter set, and the system transferred to a set of secondary spectrophotometric standard stars, approximately three per spectroscopic field. These secondary stars in turn will be observed by the main survey spectrograph to provide the system response. The calibration of the secondary stars is an onerous task, but probably can be short-circuited. There are some tens of F subdwarfs per spectroscopic field which are sufficiently bright that high-quality spectra can be obtained. These stars are spectrophotometrically very similar, and the 5-color photometric data may be enough to determine the energy distribution to sufficient accuracy for flux calibration. We then need only to obtain 10 band photometry of a grid of such subdwarfs, which is a much smaller number than the total number of secondary spectroscopic standards.

The Rate of Operation

The Monitor Telescope must be able to observe 6 bright standard stars and 6 SDSS transfer star fields per hour in order to keep up with the imaging survey and provide enough stars to calibrate the atmospheric extinction and monitor the atmospheric transmission quality. Because the field of view of the monitor telescope is large, it is possible to observe transfer fields for two adjacent scanlines of the SDSS with one Monitor Telescope field, doubling the number of stars per scanline and cutting the number of Monitor Telescope fields in half. The monitor camera CCD will have 2 corner readouts with digitization times of 15 μsec per pixel. The time budget is then as follows:

(1) Integration time of bright standard star field plus 5 sec for filter positioning and overhead - 15 sec/frame; 6 fields x 5 filters; 7.5 minutes total per hour.

(2) Readout of bright standard star field - (central 1024^2 only) - 32 sec/frame; 6 fields x 5 filters; 16 min total per hour.

(3) Integration time of transfer fields plus 5 sec for filter positioning and miscellaneous overhead - 60 sec per frame (g'r'i'z'), 115 sec per frame (u'); 3 fields x 5 filters; 17.8 min total per hour.

(4) Readout of transfer fields - 32 sec/frame; 3 fields x 5 filters; 8.0 min total per hour.

(5) Slew time: 1 minute/field; 9 fields; 9 minutes total per hour.

Total: 58 minutes per hour. This leaves us little margin for unexpected difficulties, but some of the timings above are pessimistic and all are well known with the exception of u' band exposure times.

IR Sky Monitor and Weather Station

There will be two adjunct systems to the Monitor Telescope. Both provide data on the quality of the night to the survey observers. The first is an IR all-sky scanner. This 10 μ linear array camera complements the Monitor Telescope by providing continuous all-sky images that are very effective at detecting clouds. The scanner has been built, tested, and is now in routine use at the 3.5 m telescope. It performs extremely well. The second system is a small weather station that monitors temperature, wind, pressure, and humidity. We will use it in making qualitative assessments of the SDSS data quality (e.g., does bad seeing correlate with wind direction, wind speed, rapid temperature variations, etc.).

References

Burstein, D., and Heiles, C.E., 1978, ApJ 225, 40.

Burstein, D., and Heiles, C.E. 1982, AJ 87, 1165.

Fukugita, M., Ichikawa, T., Gunn, J.E., Doi, M., Schneider, D.P., and Shimasaku, K, 1996, AJ 111, 1748.

Schneider, D.P., Gunn, J.E., and Hoessel, J.G. 1983, ApJ 264, 337.

Thuan, T.X., and Gunn, J.E. 1976, PASP 88, 543.

Astrometry

The most crucial requirement for astrometry is the need to drill holes accurately for the spectroscopy. The camera and telescope will be made much more accurately and stably than are required for this, however, and we hope that the survey will, when finally calibrated, yield very good positions. The galaxy positions for any given spectroscopic plate will ultimately come from 4 to 6 independent imaging strips that needed to be joined together with high accuracy. This pretty much requires that the galaxy positions from each strip be calibrated absolutely on the sky.

Requirements for Spectroscopy

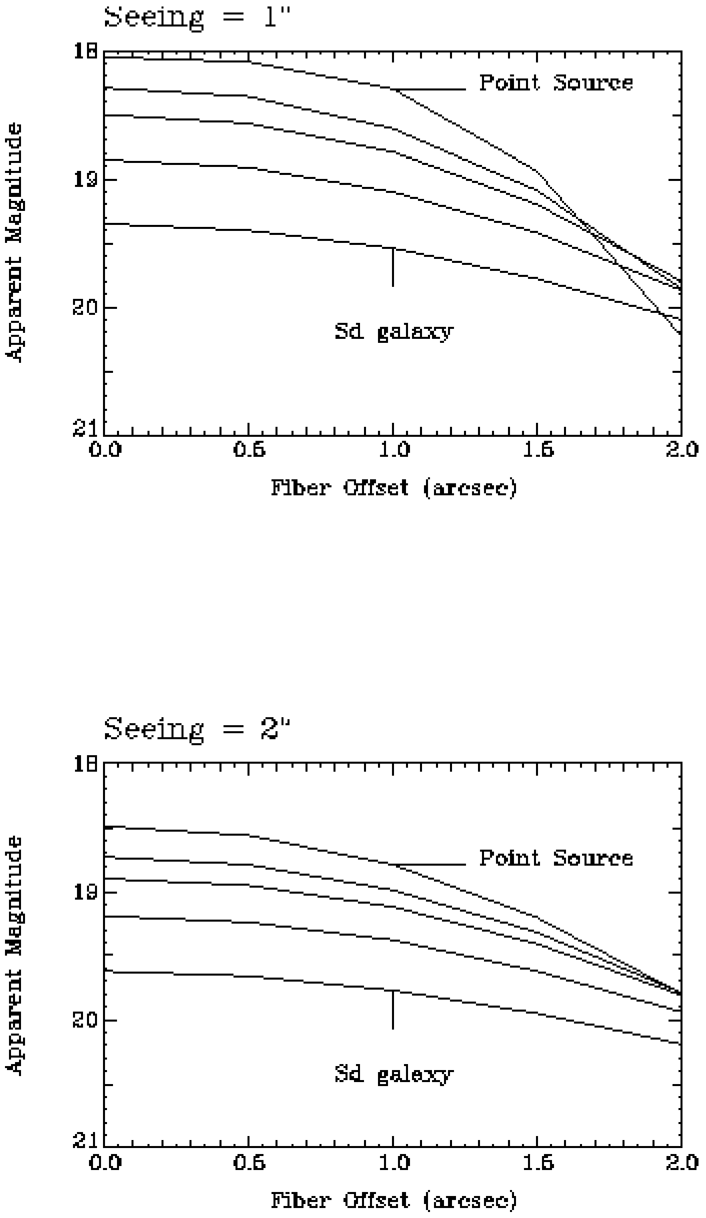
The maximum allowable de-centering error is determined by the requirement that the amount of light falling on a fiber not be seriously decreased for a worst-case galaxy or QSO. The amount of light that a fiber sees is a function of the total magnitude of an object, its concentration factor, the seeing, clouds, and fiber de-centering. Kent (1985) has measured the light concentration factors for a large number of galaxies of all types on this system. Based on this information, we have simulated observations of galaxies with properties spanning those of typical galaxies including the effects of seeing and fiber de-centering; the results are shown in Figure 6.1. As might be expected, point sources suffer more from de-centering errors than fuzzy extended objects, but the absolute amount of light that is collected is still higher for point sources until the de-centering error becomes quite sizable. Therefore, if exposure times are governed by the amount of light that comes from the low surface brightness galaxies, we can tolerate de-centering errors of order 1" . For the QSO spectroscopy, the tolerances are tighter (all QSOs being point sources) since we are trying to observe to a limit 1 mag fainter than that for the galaxies. Somewhat arbitrarily, we set the criterion for the maximum tolerable pointing error to be that it does not cause us to lose more than 20% of the light - this translates to a maximum error in radius of 0.75" . If the distribution function of errors is uniform over a circle of radius 0.75" , then the rms error in radius is 530 milliarcsec (mas) and the error in either the x or y coordinate is 375 mas.

The total de-centering error is the sum of contributions from astrometry, plate drilling, focal plane transformation, differential refraction, collimation, and guiding during observations. The collimation errors arise from the fact that the plate and glass correctors rotate relative to the primary and secondary mirrors, which introduces a rotating distortion pattern if the telescope optics are not perfectly collimated. In order that astrometric errors not dominate the error budget, we demand that the rms astrometric error be less than half the total allowed. Hence we require that the astrometry have errors of no worse than 187 mas in any coordinate, or 265 mas in radius. Table 6.1 gives our estimates of the allowed contributions from all sources. The differential refraction error is a worst-case situation for observing extreme southern or northern fields at high air mass.

Table 6.1: Contributions to De-Centering Errors

Source	RMS (milliarcsec) in Radius
Drilling	270
Astrometry	265
Differential Refraction	260
Collimation	170
Guiding	40
Focal plane mapping	20
Combined	**490**

Figure 6.1



The effect of decentering. The apparent magnitude of light passing through a 3" diameter fiber as a function of de-centering offset is shown. In each panel, the different curves correspond to galaxies with differing light concentration factors, ranging from point sources (top curve) to values typical of Sd galaxies (bottom curve). All galaxies are assumed to be viewed face-on. The top and bottom panels are for 1" and 2" seeing respectively. All galaxies have a total apparent magnitude of 18.

Technique

There are several possible approaches to the problem; one might, for example, scan a field with a high density of astrometric standards (such as the fields established for calibrating the instruments -- see the next section) enough times a night to determine the scale on each chip, its orientation, and its location with respect to some fiducial point in the array. Another approach, and the one we favor, is to use astrometric standards we encounter in the course of the imaging to update continuously a coordinate system constructed from a large number of fainter stars whose celestial coordinates we do not know. To do this, it is necessary to increase the dynamic range of the imaging array quite substantially and at the same time provide some way to determine the separations of the six CCD columns.

We will accomplish both of these objectives by the use of an augmentation to the imaging array which consists of 22 2048 x 400 astrometric CCDs, as illustrated in Figure 2.2 and discussed from the point of view of mechanics and electronics in Chapter 4 and data handling in Chapter 10. These chips have a shorter effective exposure time than do the photometric chips because they have fewer rows. In addition, we will use a neutral filter with a density of about 1.2 (3.0 magnitudes) in front of these chips. These two effects, plus differences in quantum efficiency between the two types of chip, will allow us to reach an r' magnitude of 8.5 before saturation. The high density astrometric catalog that we will use goes substantially fainter than this bright limit. We will be able to do quite accurate astrometry with this sensitivity down to about r'=17.0 , an overlap with the imaging array of $3^{m}.0$, in which range there are more than 300 stars (hereinafter referred to as the `overlap stars') per square degree everywhere in the survey region. We encounter at least 2 overlap stars per second, and at any given time there are about 3 such stars on a given astrometric chip and 15 on a given imaging chip. The five astrometric sensors across the top (in the first row) overlap the six sensors in the second row (and thus the photometric CCD columns) by an arcminute on each end and serve to tie the columns together. A star suitable for this purpose is encountered in each of these ten overlap regions about every 50 seconds.

For arrays 400 pixels high, the integration time for the astrometric sensors is about 11 seconds. As was discussed in Chapter 4, we would like to use sensors which are taller, in order to get more integration time and thus better average the seeing wander, but both geometric and financial constraints limit us to these chips.

The best choice for an astrometric catalog for the primary standard stars is clearly the Tycho catalog, due to be released in May, 1997. This catalog has a high density (about 10^{6} stars total over the celestial sphere, an average density of 25 stars per square degree with 10 stars per square degree at the North Galactic Pole) and high astrometric accuracy (estimated to be .030" in each coordinate) (Høg *et al*., 1992).

The only shortcoming of the Tycho catalog is its lack of accurate proper motions. However, partly in anticipation of this problem, the U.S. Naval Observatory (USNO) has prepared an astrometric catalog based upon plates taken with their 8-inch Twin Astrograph (the mean epoch of the catalog is about 1981). This Twin Astrograph Catalog (TAC) contains virtually all of the Tycho catalog stars and has, on average, over 20 stars per square degree to a magnitude of about V=12 . The stellar density at the North Galactic Pole is about 10 stars per square degree. TAC is supplemented by accurate proper motion determinations based upon a new solution of the old Astrographic Catalog (AC) Zone plates. TAC will be accurate (in each coordinate) to about 0.080" at epoch, with an uncertainty in the proper motion of about .003"/ year yielding positional accuracies of 0.10" at J2000. USNO assigned top priority to reducing those zones contained within the SDSS survey area and those reductions are now complete. As soon as Tycho is released USNO will provide SDSS with proper motions for the catalog stars. The resulting positions should be accurate to about .040" per coordinate at the epoch of SDSS observations.

The Calibration

Calibration of the astrometric array is done in two separate processes. The first process is to determine the positions of the individual CCDs relative to one another as they are mounted on the rigid camera base. This will be done by scanning the telescope through a portion of sky with a dense grid of stars with very accurate astrometry. The USNO has prepared several such astrometric calibration fields using observations obtained especially for this purpose with the 8-inch Transit Telescope at Flagstaff. Sixteen fields, each 3.2 degrees in width and 1/2 hour (7.5 degrees) in length, have been established all around the equator. The limiting magnitude is typically 17.1, stellar densities vary from a few hundred to a few thousand per square degree (depending upon Galactic coordinates) and one-dimensional internal positional errors for stars brighter than about $15.^{m}5$ are < 0.025" , falling to .050" at $17.^{m}0$. Comparison with preliminary Tycho data indicates that the (external) astrometric accuracies are better than .050" .

The second process is to determine the actual scanning path through the sky during a particular drift scan and the orientation of the array relative to that path. For the discussion which follows, consider the calibration of the array of CCDs as a single large CCD.

The telescope nominally tracks a great circle. Introduce (μ, nu) as great circle coordinates. The longitude μ increases along the great circle and the latitude nu increases perpendicular to the great circle; we have nu =0 along the great circle itself. For the moment we ignore the complications of precession, refraction, etc.

The CCD device has coordinates (X,Y) . The units are device pixels. X increases perpendicular to the scan direction (to the North). Y increases along the scan direction (to the East).

The CCD image has coordinates (x,y) . The units in the x direction are ordinary device pixels (also called x -pixels below when it is necessary to distinguish the two); the units in the y direction are y -pixels. x increases perpendicular to the scan direction (to the North). y increases along the scan direction (to the East). The distinction between device and image coordinates is necessary for drift scan images because the two do not necessarily coincide exactly. For example, the y axis is always aligned precisely with the drift direction of the telescope even if the Y axis is rotated slightly with respect the the drift direction. Also, a y pixel may not necessarily map to the same number of arcseconds as an x pixel; it will do so only if the CCD clocking rate perfectly matches the drift rate of a star across the CCD.

Introduce the following parameters:

- p: The size of a pixel in mm
- f: The focal length of the telescope in mm
- s: The pixel scale in arcsec/ x -pixel; s=206265p/f .
- C: The CCD clock parallel clock rate in tsec/ y -pixel.
- T: The telescope tracking rate in arcsec/tsec.
- (μ_0,nu$_0$): The coordinates of the pixel x=0, y=0 .

If the telescope tracking and rotator alignment are not perfect, we have two extra parameters to describe the misalignments:

- theta: The tilt angle between the Y axis and the drift direction (= tilt angle between either the X or x axis and a vector in the latitude direction). (Positive values correspond to the x axis being rotated east of north).
- phi: The tilt angle between the actual telescope tracking path and the intended great circle of nu =0 . (Positive values correspond to the telescope tracking north of the intended path).

To good accuracy, we have

$$\begin{aligned} \mu &= \mu_0 + CT(y + [\theta - \phi]x) && (10.1a) \\ \nu &= \nu_0 + s(x + \phi y). && (10.1b) \end{aligned}$$

Thus there are six parameters (counting CT as one parameter) that must be calibrated for each scan (or section thereof). To first order, we simply scan the array across a number of astrometric standard stars and solve for the six parameters using standard least-squares.

Next, we consider the complications introduced by the fact that we have a mosaic of CCDs. Select one CCD as a reference. Then each other CCD has positional offsets D$_x$, D$_y$ in mm relative to the first and a rotation theta$_b$ with respect to the coordinates defined by the reference CCD. (We shall also add an extra parameter to allow for a difference in scale factor). For N CCDs in the array, we need to calibrate N-1 times 3 (or 4) parameters. To do so we will scan across the high density astrometric standard fields.

In addition to the information provided by scanning the astrometric standards, we get additional information from multiple crossings of the front and rear CCDs in the array by the same overlap stars. This provides us with a near-continuous monitoring of two numbers: the front-back x differential and the front-back y differential. If we have *a priori* calibrations of D$_x$, D$_y$, and f for the front and back CCDs, then we reduce the original six free parameters in Eq. (6.1) to essentially four (the angle theta$_b$ adds another parameter, but it can be determined from the front-back crossings quite accurately). The usefulness of the extra 2 measurements will depend on being able to measure D$_x$ and D$_y$ accurately *and* on the time behavior of the calibration parameters. It may be possible to use the r' (and even the g' if we are careful about differential refraction) photometric CCD to add additional crossing measurements (this requires that the photometric CCDs be mounted with the same rigidity and thermal behavior as the astrometrics).

Next, we consider the time dependence of the parameters. We require that parameters be stable on a time scale long enough that we can scan over enough standard stars to get a satisfactory calibration. We use the lowest density of astrometric reference stars (10 per square degree, found at the North Galactic Pole) in what follows. The camera images a swath which is 2.3 degrees wide and if we scan at a sidereal rate we will cover 2.3 square degrees every 4 minutes. Hence, we will be scanning over at least 6 Tycho stars per minute, each star yielding 2 measurements (x and y). In 10 minutes we thus have 60 pairs of measurements to constrain 6 parameters. Therefore we have some freedom to introduce a small number of extra degrees of freedom per parameter for each 10 minutes of integration in a scan, and our error estimates later will make this assumption.

Finally, we consider the rotations and distortions introduced by such effects as precession, nutation, and aberration. There are seven effects to worry about. Precession, nutation, polar motion, and UT1-UTC offsets correspond to rotations of the sky without distortion. These will be accounted for already by the telescope tracking algorithms; any residual corrections will be absorbed in the 6 constants of the calibration process. Aberration and refraction correspond to a distortion of the (x,y) system relative to the (μ, nu) system, and refraction is large enough that second order corrections will be needed. They can likely be calculated accurately enough that there is no need to introduce additional unknown parameters. Low order imperfections in the telescope drive surfaces will be removed by the telescope control computer. High order imperfections will appear as time variations in the tracking sensitive parameters.

It is appropriate to describe here the tracking algorithm to be used by the telescope. The telescope control computer keeps track of a specific point in the telescope focal plane that is called the boresight. The boresight is not fixed in the array but is at one of two places for the two strips that compose a stripe (see Figure 2.2). It is the boresight that will track great circles. Technically, the boresight tracks a path that is a great circle in J2000 coordinates as viewed from the solar system barycenter. Aberration and refraction distort the path so that it is no longer a great circle as viewed from the ground; however, the distortion is small enough that it does not introduce any significant curvature in the paths of stars in the drift scans. The Y axis of the array is oriented along the apparent equator of the great circle as viewed from the ground. Refraction and aberration also distort the scale of the great circle such that a step of 1 arcsec in J2000 coordinates does not correspond to 1 arcsec as viewed from the ground. The tracking rate of the telescope (which is computed using J2000 coordinates) will be adjusted so that the apparent rate of stars drifting across the focal plane is strictly constant.

Once the positions of the overlap stars are calibrated, one can imagine transferring the coordinate system so established to the photometric array in any one of a large number of ways, the most straightforward of which is simply to use the overlap stars as secondary astrometric standards. The photometric CCDs will certainly wander with respect to the astrometric array in response to thermal and flexural changes, but on a timescale which is very long compared to that required to tie to the astrometric system with very high accuracy. In particular, the output of the r' chips can be treated very much as the output of the astrometric chips is treated, and provides an interpolative point between the leading and trailing chips to monitor drive irregularities in both coordinates at frequencies corresponding to the residence time of a star on the array. These frequencies will be very high compared to any relative drifts between the astrometric and photometric arrays.

Measurement Errors

Will we be able to achieve the desired accuracy of 260 mas rms per coordinate? Here we look at all the possible error sources.

Table 6.2 gives a summary of all errors, statistical and systematic, in calibrating the secondary standard stars; such a set would then be used to calibrate a single frame from the photometric array. The following assumptions go into the table:

1. We scan the telescope for 10 minutes and solve for the 6 coefficients in Eq. (6.1).
2. Each secondary astrometric standard is scanned 4 times (once per strip by both the leading and trailing arrays and twice again when the second strip of the stripe is scanned).
3. We use the Tycho catalog for astrometric primary standards.
4. We scan through the North Galactic Pole and encounter 6 Tycho stars per minute, or 60 for the 10-minute scan.
5. If a single primary standard star has a random error of sigma in one coordinate, then the rms 1-d error in the pointing solution after we scan N randomly distributed stars is sigma $(7/N)^{1/2}$.
6. The 1-D seeing error for a single measurement is 60 mas.
7. The rate of crossing primary standards is low enough that seeing errors are uncorrelated.
8. The secondary standard stars are spaced closely enough in time that seeing errors amongst them are perfectly correlated. This is certainly wrong, but produces a worst-case error estimate.

Table 6.2: Contributions to Astrometric Errors

Source	RMS Radius Error (mas)
Error in primary standards	24
Seeing in primary standards	25
Seeing in secondary standards	42
Shot noise in secondary standards	20
Uncorrected tracking errors	55
Combined	**80**

This is well below our target of 265 mas. There will be additional errors in going from the secondary stars to the QSO and galaxy positions, but they will be small compared with the tie between the primary and secondary stars.

Here we discuss the errors individually.

Errors in Primary Standards

We assume that the Tycho positions have (conservatively) an rms error of 50 mas per coordinate at epoch J2000 and that we encounter a minimum of 6 catalog stars per minute. The error quoted is for a single scan of the telescope; repeated scanning, obviously, will not reduce this contribution.

Seeing

There is some discussion of seeing at APO in Chapter 3.5; here we concentrate upon theoretical aspects.

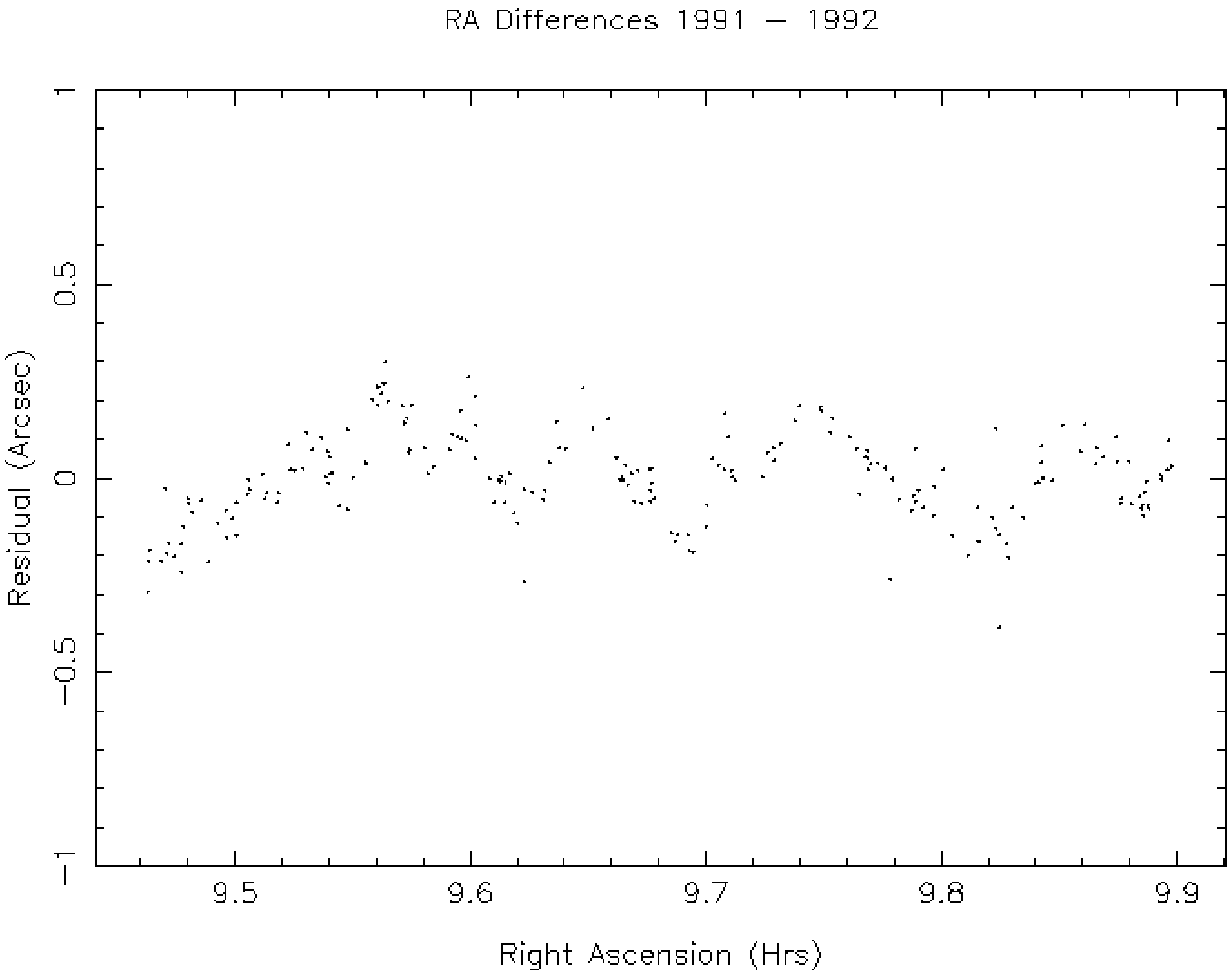
The evaluation of the contribution which seeing makes to the position error for a telescope in TDI mode is similar to the considerations used for ordinary imaging. Under the assumptions that (1) the seeing is described by Kolmogorov turbulence with some outer scale r_{max} and whose amplitude is characterized by a Fried parameter r_0 with which the long-integration-time seeing FWHM w scales inversely (w ~ lambda /r_0), (2) the outer scale is larger than the telescope aperture, and (3) the integration time is longer than the time it takes the outer scale in the seeing screen to traverse it can be shown that, at a given wavelength, (see, e.g., the developments in Roddier, 1981)

$$\sigma \quad \propto \quad w^{5/6} r_{\mathrm{max}}{}^{-1/6} \left(\frac{Vt}{r_{\mathrm{max}}}\right)^{-1/2},$$

where sigma is the one-dimensional standard deviation in the position, t is the integration time, and V is the characteristic wind velocity at the height where the seeing originates. With these assumptions the error is independent of the telescope aperture. There is not much data on the size of the outer scale, and it is probably variable, but it is usually assumed to be of the order of 10 meters.

We have analyzed a data set obtained with the four-shooter camera on the Hale telescope at Palomar by opening the shutter for a long period with the telescope parked with the instrument aligned so that the trajectories of stars are accurately along the CCD columns. The wander in those trails is entirely the consequence of seeing, and the behavior of the autocovariance function of that wander gives one amplitudes and time dependence. It appears from this data set that the characteristic time associated with the outer scale, r_{max}/V , is about 0.5 seconds; this is determined by the integration time at which one enters the $t^{-1/2}$ behavior above. For a wind speed of 20 m s^{-1} , which is reasonable, the outer scale is about 10 meters. The seeing at the time the trails were taken was 2.0 arcsec FWHM, and the image wander averaged over 11 seconds was 56 mas. Scaling to 1.0 arcsec seeing, we would expect about 32 mas. We have also analyzed driftscan data from the Mt. Hopkins 1.2 meter telescope. TDI mode driftscans of the same portion of sky were obtained in two separate years; the instrument was a Ford Aerospace 2048 CCD with 0.66" pixels and an integration time of 50 seconds; the seeing was about 1.5 arcsec. Bright stars in the two scans were matched up and positions compared. A plot of the right ascension residuals (after removing an offset and drift) is shown in Figure 6.2. The data span about 40 minutes. The rms difference between the two sets is 120 mas, or 85 mas per single observation, or 60 mas for 1" seeing. The residuals show considerable correlation over time scales of order a few minutes with an amplitude rather larger than seeing theory predicts. All told then, we can expect seeing errors to be no better than 32 mas, but no worse than 60 mas.

Figure 6.2



Comparison of star positions from drift scan data taken in separate years. Only the residuals in right ascension are shown (the declination residuals have a similar behavior). An offset and slope have been removed from the data. The rms difference is 0.12" . The data span about 30^{m} of time. For both sets of data, the telescope was stopped and the drift scanning was done at sidereal rate.

Shot Noise

Shot noise can be a significant source of error for the measurement of the position of an individual faint star (it is almost entirely negligible -- on the order of a couple of mas -- for bright stars with many thousands of electrons in the image). Computer simulations have been carried out to estimate the size of this effect using realistic values for dark current, read noise, quantum efficiency and sky background for the astrometric chips. At r' = 17.0 , the faint end of the `overlap' star range and, hence the worst (though most frequent) case, simulations show that the astrometric error contribution from shot noise is about 50 mas per coordinate per star at this level. The shot noise contribution to the astrometric error for a frame depends upon the number of such stars on the frame. No fewer than 10 such stars should show up per frame, and not all will be at the faint limit. The adopted value, 20 mas in radius, should be conservative enough.

Tracking Errors

Chapter 2 has reviewed the expected tracking performance of the telescope. We note that the calibration procedure automatically absorbs any constant offsets and linear drifts in the tracking, so only deviations with respect to a linear drift in the tracking are of concern. The error estimates given here are based on acceptance test measurements of the assembled 2.5-M telescope major components at Apache Point at the maximum tracking rates we expect to encounter.

Ultimate Accuracy

We have shown that we can achieve astrometric accuracy which is plenty good enough to place the fibers. But how well might we be able to do in a finally calibrated version of our catalog? We have a couple of schemes in mind to improve the telescope tracking performance and if we implement them successfully we should become seeing-limited.

Table 6.2 assumes a rather simplistic approach to the determination of the transformation parameters, in which the fact that one has very many more secondary standards which determine *some* of the parameters to very high accuracy is not used to advantage. In particular, the large contribution from seeing in the secondary standards can likely be reduced to levels below that for the primaries if one is careful; in any case, the error is seriously overestimated because of the assumption of perfect correlation among the secondaries.

Thus the astrometric errors can probably be brought down to of the order of 50 mas in radius, or 35 mas in each coordinate. This accuracy would be of enormous utility for statistical parallaxes, tying radio source maps to their optical counterparts, and the eventual measuring of proper motions for the many faint main sequence stars the survey will find. If we do this well, repeating the northern survey to measure proper motions and long-term variability becomes an exciting possibility.

References

Kent, S.M., ApJSuppl 59, 115, 1985.

Høg, E., Bastian, U., Hansen, P.C., van Leeuwen, F., Lindegren, L, Pedersen, H., Saust, A.B., Schwekendiek, P., and Wagner, K., AstrAp 258, 201, 1992.

Roddier, F., *Progress in Optics* **XIX**, 281, 1981.

Spectrographs

Desired Performance

The principal scientific goal motivating the spectrograph design is to obtain in a single exposure 600 spectra of galaxies as faint as the spectroscopic limit of r'~ 18.2 over the three-degree field of the telescope. As always, the projected performance is the result of optimizing scientific return within cost and technology constraints. We will not explain the compromises (cf. the discussion in Chapter 1), but will describe the instruments as we are building them. We first outline the requirements driving the design; those that most directly affect it are resolution, wavelength coverage, number of fibers, and fiber diameter.

Resolution (2000)

The spectral resolution is 2000, roughly where absorption lines in galaxies and quasars are resolved. Lower resolution reduces signal-to-noise by filling in the absorption with continuum light, higher resolution needlessly reduces spectral range, given that we can use at most two 2048^2 sensors to record the spectrum. Higher resolution also reduces the signal per pixel so that readout noise dominates for any but long exposures. The resolution we have chosen and the efficiencies we expect will give us shot-noise limited spectra almost everywhere in the bandpass with exposures as short as 20 minutes. A resolving power of 2000 is sufficient to distinguish velocities of 150 km/s in a single data element and is good enough to measure velocities to better than ± 20 km/s from an entire galaxy spectrum.

Wavelength Coverage (3900-9100 Å)

Placing the blue limit at 3900 Å ensures that the H and K lines of CaII are observed even at zero redshift. To observe H alpha to a redshift of $z = 0.2$ or more requires coverage to at least 8000 Å. Detector size suggests that going to 9100 Å is possible and appropriate, and the extra range is valuable for quasars.

Number of Fibers (640)

To finish the survey in five years requires observing about 550 galaxies in each one hour exposure with efficiency good enough to do up to ten exposures on a single long night. Each field will also have between 50 and 100 quasar candidates plus a few fibers for the sky and the stars used for reddening, spectrophotometric calibration, and correcting for atmospheric absorption bands in the near infrared. The design comfortably accommodates 640 spectra per observation, using two spectrographs with sensors that are 2048 pixels wide.

Fiber Diameter (180 μm)

For galaxies near our brightness limit, a 3 arcsec diameter aperture is a good compromise between collecting most of the galaxy light and rejecting the night sky. Fibers that preserve f/ratio and have excellent transmission over our wavelength range can be obtained in this diameter (180 μm). For galaxies near our faint limit, this diameter gives a galaxy flux comparable to the sky.

Technical Constraints

Telescope

The telescope was designed with fiber spectroscopy in mind so it is hardly a constraint. The f/5 focal ratio is a good acceptance beam for optical fibers and the 2.5 m aperture is sufficient to measure faint galaxy redshifts in reasonable time. The aperture and f-ratio set the focal length and hence the physical diameter of the fibers. The required input aperture diameter, three arc seconds, corresponds to 0.18 mm at the focal plane.

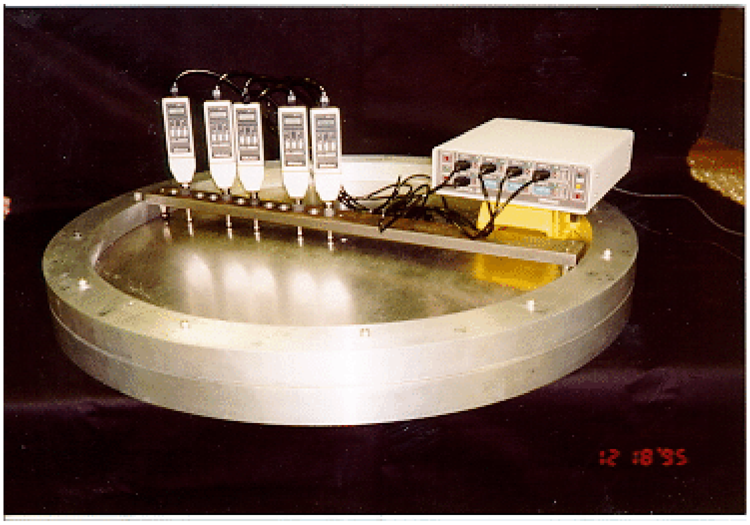
CCD Format

The detectors are Tektronix 2048 x 2048 CCDs with square 24 μm pixels. The large pixels allow a modest demagnification ratio of 2.5 so the challenge is to put the full area of the detectors to good use. This requires a spectrograph camera with an exceptionally large field of view (16.5°) and the largest practical beam diameter (150 mm).

Spectrograph Characteristics

The final design is set after considering the science requirements, telescope, detector format, and operational considerations. The important characteristics are summarized here. Details appear in following sections.

Figure 7.1



Measuring the profile of a fiber plug plate in a drilling template. The template deforms the plate to allow a standard 3-axis milling machine to produce hole axes aligned to the principal ray axis.

Fiber Optics

Fiber optics is the obvious way to do simultaneous spectroscopy of hundreds of faint galaxies. We will use custom drilled aluminum plug-plates to hold the fibers in the telescope focal plane (see Figure 7.1). These plug-plates will be installed in a fixture called a fiber cartridge and manually stuffed with fibers before nightfall. The fibers are brought to the slithead, which is incorporated in the cartridge and which mates with the spectrograph. During the night, up to ten cartridges will be swapped out to observe different fields.

In addition to the object fibers, which are single strands, there will be ten coherent fiber bundles capable of imaging a few arcseconds of the sky. These will be placed on preselected guide stars and feed a CCD camera mounted on one of the spectrographs. The guide stars will be used to center the telescope on the plug-plate field, adjust the plate scale of the telescope, control focus, and guide the exposure. One larger (30 arcsecond) bundle will be used to measure the sky brightness. Image quality and photometric data from these guide and sky bundles will be used to estimate the exposure time required to complete the observation.

Two Spectrographs

The images of the fiber ends must be adequately sampled by the detector and be spaced sufficiently far to prevent crosstalk. Each fiber end (180 μm diameter) will just span three pixels (3 x 24 μm = 72 μm) with a spectrograph demagnification of 2.5. Placing the individual fibers at 360 μm intervals we might squeeze up to 340 spectra onto a single 2048 x 2048 CCD. In practice, packaging constraints limit us to 320 spectra per spectrograph and the fiber ends are spaced at 390 μm intervals. This is many fewer than the 600 or so needed to complete the survey in five years so we will use two identical instruments fed by a single plug-plate. With 320 fibers in each instrument, we get 640 spectra per exposure.

Two Channels

If a resolution element (the projected image of a fiber end) is three pixels across, then 2048 pixels cover 1700 Å in the visible at R = 2000 . To cover the desired spectral range of 3900-9100 Å at the desired resolution of 2000, we need more than 2048 pixels. Each spectrograph therefore has two detectors, one covering 3900-6100 Å and the other 5900-9100 Å. This gives the desired spectral range while approximating the R = 2000 requirement. The blue-red split is done with a dichroic coating on a beamsplitter, the blue side reflected and the red side transmitted.

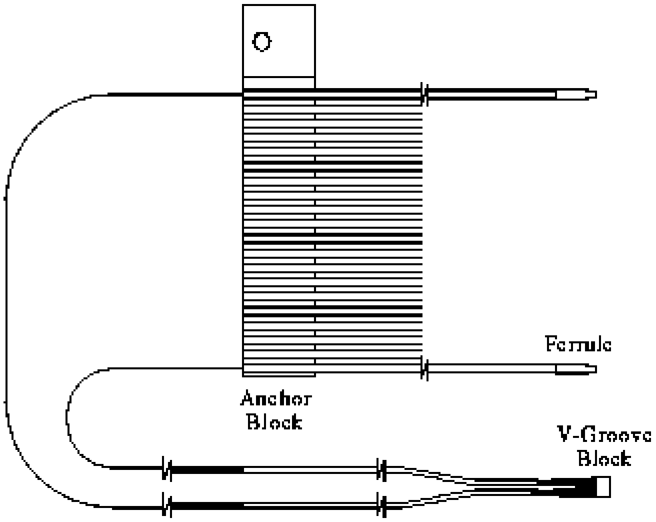
Multiple Fiber Cartridges

To maximize efficiency, ten fiber cartridges, each loaded with a different plug-plate, will be ready to go at the beginning of the night. No nighttime plugging is required.

Spectrograph Mounted on Telescope

The spectrographs will be mounted on the telescope to maintain good fiber performance during and across exposures. Having captive fibers routed directly to a spectrograph on the telescope avoids repeated bending that would occur if the spectrographs were on the floor. The disadvantage is that the spectrographs must not flex as the telescope tracks.

Figure 7.2



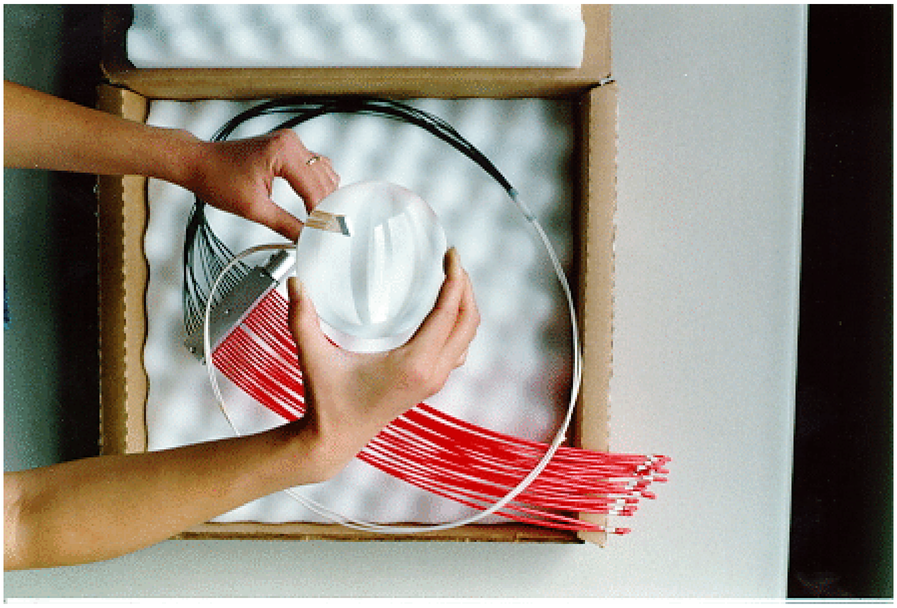
Fiber harness. Twenty fibers are terminated in the v-groove block at one end have ferrules at the other.

Fiber Feed System

Fibers

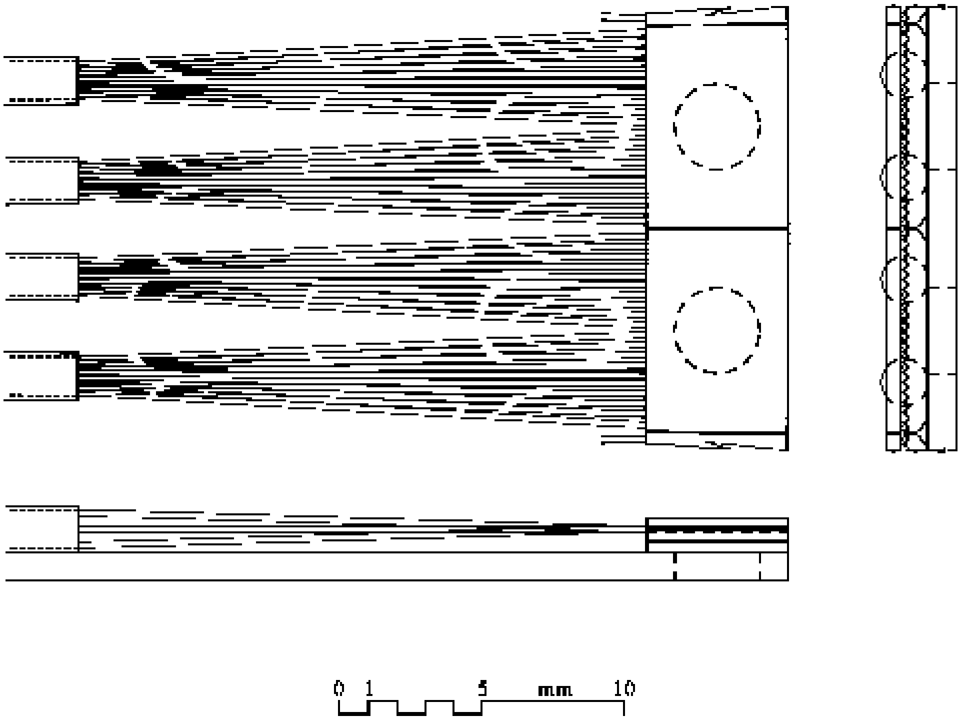
We are using silica UV-enhanced step-index fiber with a core diameter of 180 μm and a polyamide protective buffer. The spectrographs are mounted on the telescope, so each fiber need only be 2 m long. A sample set of fibers is shown in Figure 7.3. Each plug-plate has 640 fibers (plus approximately ten fiber bundles for acquisition and guiding), or 320 per spectrograph. Hence we require 6400 optical fibers plus spares to support ten plug-plate cartridges.

Figure 7.3



Prototype fiber optic cartridge with 20 fibers. The ends, which are plugged and unplugged during operations, are protected by tough nylon tubing. The lens enlarges the v-groove block termination at the slit end.

Figure 7.4



Detail showing v-groove blocks mounted to the slit.

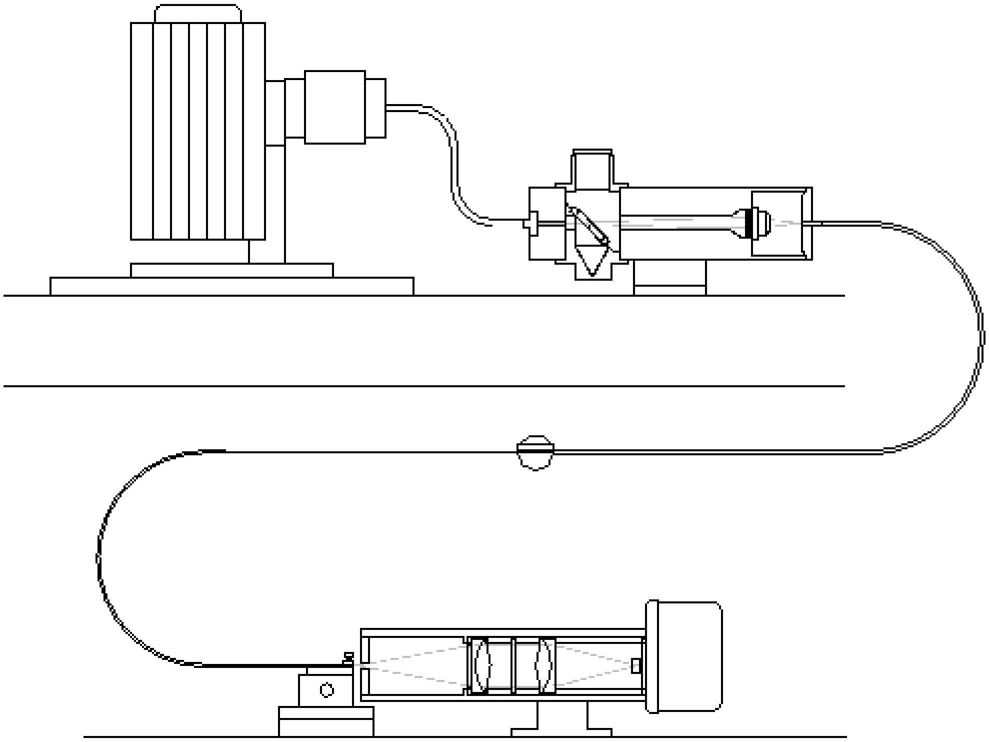
At the slit, the fibers are separated by 390 μm center to center. The fiber outside diameter is only 220 μm, so there is insufficient room to individually terminate the fibers. Our two choices are to mount bare fibers to the slit or terminate the fibers in sets. Bare fibers are fragile and are difficult to polish flat and normal to the fiber axis. Mass-terminated fibers are robust and easily polished flat. We have chosen to terminate the fibers in v-groove blocks of 20 fibers. The resulting v-groove block is large enough to be handled easily, yet has few enough fibers that we can afford to replace a set if a few fibers break. We call a set of 20 fibers a "harness" (Figure 7.2).

At the plug-plate, each fiber is terminated in a stainless-steel ferrule. Jacketing on the fiber applies a torque to this ferrule, providing retention of the ferrule in the plug plate. This approach has been used successfully in many other plug-plate-based fiber spectrographs. Loss of light due to tilt of the ferrules in the holes is 0.3% or less. We achieve this by using high-precision ferrules and holes drilled with high precision spade drill bits held in a custom-made collet.

The fibers are supported below the plug plate by an anchor block. This block absorbs stresses induced by plugging and orients the fibers for maximum retention in the plug plate. Between the ferrule and the anchor block each fiber is encased in a loose-fitting jacket. The jacket protects the fiber from undue bending and applies sufficient torque to the ferrule to retain it in the plug plate.

Between the anchor block and the slit the fibers are not disturbed even during plugging, and so need not be heavily protected. In most of this region the fibers will be individually jacketed. Near the slit, the fibers will be encased in larger tubing, ten fibers to a tube, to assist in routing the fibers along the spectrograph slit.

Figure 7.5



Fiber tester

At the spectrograph slit, the twenty fibers of the harness terminate in one custom-made stainless steel v-groove block (Figure 7.4). The grooves of the block are fanned out slightly in such a way that although the block is polished flat, the beam emitted by each fiber is normal to the slit.

We received and tested prototype harnesses from a number of vendors. These harnesses have adequate throughput and survive lifetime tests. The final vendor selection has been made and two complete sets of 640 fibers have been fabricated and tested.

We will install the fiber optic harnesses ourselves. The harnesses are attached at two points: the anchor block and the slit. The anchor block mounts to the plug-plate cartridge frame with a single bolt. The v-groove block is aligned by pressing the output edge against a curved alignment jig, and attached to the slit with an adhesive. An air-powered fluid dispenser meters the adhesive to prevent contamination of critical surfaces. A hole in the slit under each block allows removal of the block for replacement; a custom-made tool reaches through the hole to push the block off the slit, while simultaneously holding down the two neighboring blocks.

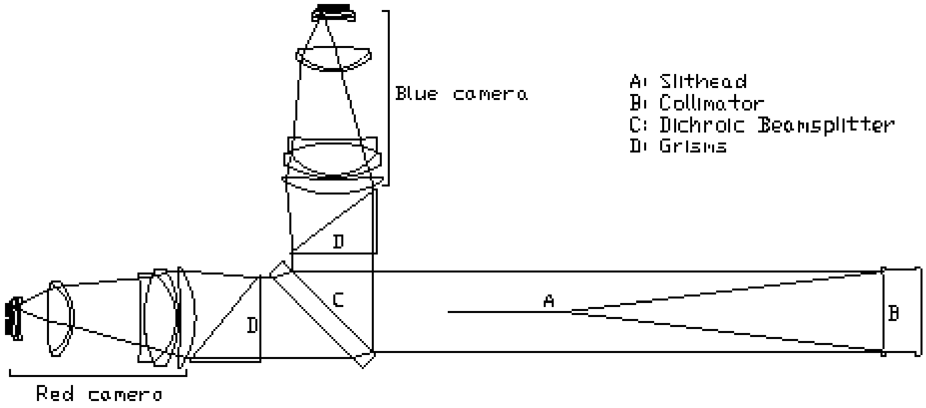
Fiber Tester

During manufacturing, each optical fiber will be tested for adequate throughput using the apparatus shown in Figure 7.5. White light from an intensity-stabilized quartz-halogen lamp is fed to the apparatus via a "source" optical fiber. The end of the source fiber is imaged onto the fiber under test using a microscope objective, which produces a uniform f/5 converging beam. A microscope eyepiece and pellicle beamsplitter give the user a view of the fiber under test.

Light from the output end of the test fiber is collected by a pair of achromatic doublets focussed onto a silicon photodiode. A filter between the doublets flattens the quartz-halogen spectral curve. A calibrated aperture blocks light outside a cone of f/4. A computer-controlled translation stage allows one to accurately locate the appropriate fiber of the v-groove block in front of the aperture. This same light collection system may also be used to measure light from the microscope objective, allowing us to make absolute throughput measurements.

We have constructed two fiber testers. The manufacturer uses one to measure every fiber and reject those found to have inadequate throughput, and we use the other to verify the manufacturer's measurements for some fraction of the fibers.

Figure 7.6



Spectrograph optical layout. The beamsplitter (C) transmits longward of 6000 Å.

Optical Design

The spectrograph optical layout is shown in Figure 7.6. Light from the fiber optics (A) exits in an f/4 beam, expanded somewhat from the f/5 input beam of the telescope due to processes collectively known as focal ratio degradation. The beam encounters the spherical collimator mirror (B) and collimated light returns in a 150 mm diameter beam, passes the slit, and meets the dichroic beamsplitter (C). The blue light (< 6000 Å) is reflected while the red light is transmitted. Past the beamsplitter, the light encounters the dispersing grism (D). The dispersed light exits the grism and meets the wide field f/1.3 240 mm camera that images the spectra onto the Tektronix 2048 x 2048 CCD. Optical details of these components are described here. Mechanical details are in the following section.

Collimator Mirror

The fiber slit is 4.9 inches long, packed with 320 fibers. The obvious collimator design to use with this long slit is a Schmidt. The Schmidt is inexpensive to manufacture, has the required field of view (6°), and uses a mirror, which makes the instrument compact.

Although we explored a classical Schmidt design, we found it possible to eliminate the corrector plate because our imaging requirements did not need the full Schmidt performance and some of the deficit due to the missing correctors could be compensated in the spectrograph cameras. Thus, our final collimator design is a single mirror with a spherical figure. The mirror is rectangular, 7 x 17 inches, with a 49.8 inch radius of curvature. The substrate is a slumped borosilicate gas fusion blank made by Hextek.

We examined alternatives to the Hextek blank including aluminum, eggcrate Zerodur, and monoliths. The gas fusion blank wins easily on weight and cost considerations. The compromise is in the thermal coefficient of expansion, which is non-zero but smaller than aluminum. An aluminum mirror would obviate the need to refocus when the temperature changes, but is expensive and the coatings are fragile. An eggcrate Zerodur mirror is immune to large temperature changes but is expensive and would require the most refocusing. A monolith would work well enough, but would not be significantly less expensive than the lighter and more thermally responsive gas fusion mirror.

An enhanced silver coating will be placed on the mirror for better than 95% reflectivity throughout our wavelength range.

Beamsplitter

The beamsplitter, which divides the collimated beam into red and blue channels, is made of fused silica. This material has excellent transmission and low dispersion in our wavelength range. It is thick, 1.5 inches, because the reflecting surface is a mirror and needs to maintain a flat figure.

The dichroic coating is efficient in reflection (98%), reasonably good in transmission (94%), and has a narrow 200 Å crossover range.

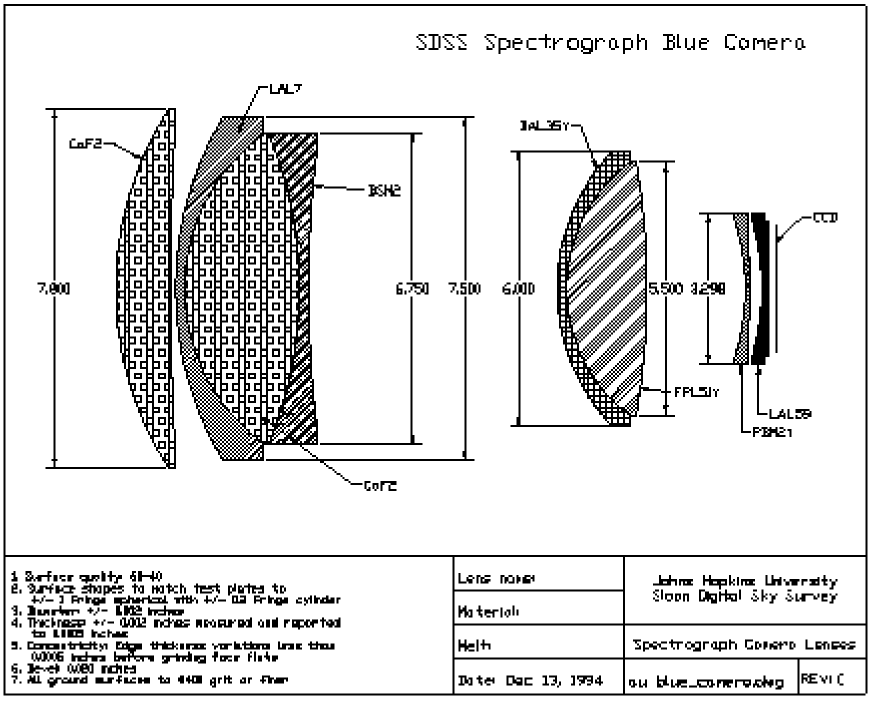
Grism

The dispersing elements are grisms with zero angular deviation at 4960 Å for the blue and 7400 Å for the red. In our case, these are right angle prisms with a transmission grating replicated on the hypotenuse. While a reflection grating might have been used, the grism permits mounting the cameras close to the system pupil, which is about midway on the grating. With a reflection grating, the cameras have to be mounted away from the grating to avoid interference with the incoming beam, making them larger and more difficult to design. A plane transmission grating does not work because the diffracted angle is large, making geometric losses high (the groove facets are foreshortened) and forcing the blaze peak outside the optical band. Our configuration has little groove shadowing or foreshortening and results in high grating efficiency.

The ruling densities are 640 and 440 lines/mm for the blue and red grisms, respectively. Because master rulings of the size and groove angle needed did not exist, new masters were ruled.

Camera

Figure 7.7



Spectrograph camera designed by Harland Epps. Diameters are in inches. The asphere is on the air surface of the FPL51Y lens.

The spectrograph camera is an all-transmission design because the fiber optics generate a filled beam. The usual practice of hiding the detector or a secondary mirror in the center of the beam (as in a Schmidt camera) cannot be used without unacceptable light loss.

The cameras are the biggest technical challenge in the spectrograph. An f/1.3 camera with 240 mm focal length and 16.5° field of view capable of using 24 μm pixels cannot be purchased off the shelf. On the other hand, cameras with similar performance are now being built primarily because of the availability of Tektronix 2048 CCDs with 24 μm pixels and the popularity of fiber spectrographs.

We considered modifying an existing design and examined the KPNO Bench Spectrograph Camera (by George Simmons), the VLT spectrograph cameras (Hans Dekker et al.), and the LRIS and Norris spectrograph cameras (Harland Epps). All are Petzval lenses with two widely spaced, approximately equal powered components. None provided the performance needed.

We chose to contract Harland Epps (UCSC) to design new cameras for our spectrographs, shown in Figure 7.7. The result is similar to the LRIS design but has a smaller diameter and larger field of view. All but one of the surfaces are spherical and the asphere is relatively mild.

Detectors

The CCD detectors are thinned Tektronix 2048 x 2048 with square 24 μm pixels. Readout noise is 5 electrons, full well is 150,000 electrons. Their remarkable characteristic is a high quantum efficiency in the blue spectral region (see Figure 7.9).

Optical Prescription

Table 7.1 shows the spectrograph optical prescription. Some of the surface descriptions can be deciphered from this example: "Doublet,first,back" refers to the doublet component, first lens element, back surface (closest to the CCD). The "Radius" is the radius of curvature in inches, negative implying a concave left surface. "Thickness" is the distance from the current surface to the next; positive to the right. All materials are from Ohara Glass except for CaF2 and the lens couplant, Dow-Corning Q2-3067.

Table 7.1: Spectrograph Optics

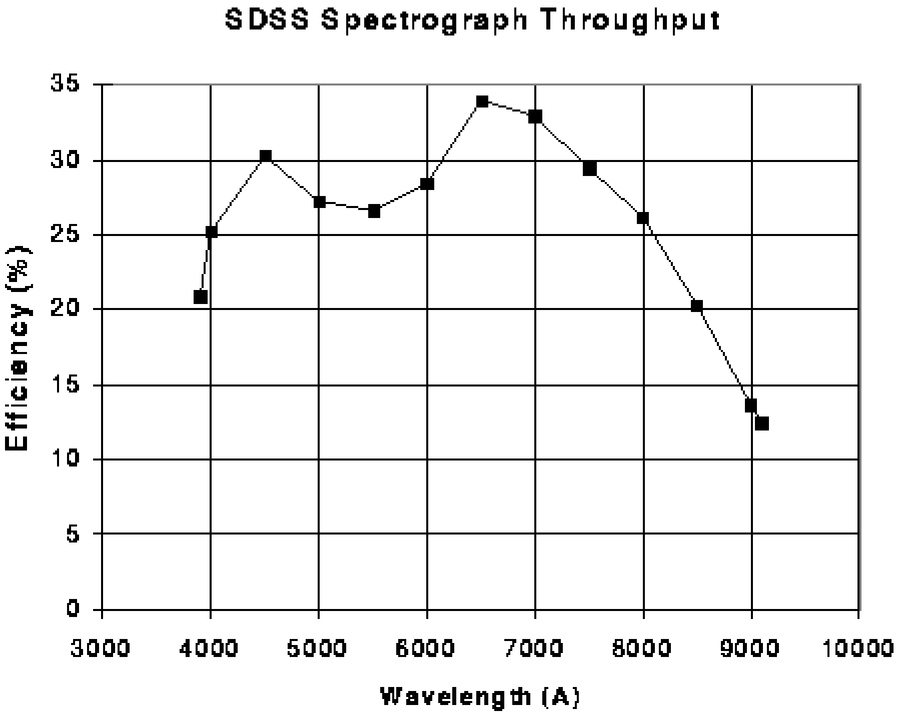
Surface	Radius (inches)	Thickness (inches)	Material
slithead	-25.2	24.803	air
collimator mirror	-49.760	-47.328	air
prism (37°)	flat	2.543	BK7
grating	flat	2.070	air
singlet,front	-7.196	-1.202	CaF2
singlet,back	-71.400	-0.100	air
triplet,first,front	-7.306	-0.203	LAL7
triplet,first,back	-4.168	-0.003	Q2-3067
triplet,second,front	-4.168	-2.501	CaF2
triplet,second,back	8.903	-0.003	Q2-3067
triplet,third,front	8.904	-0.201	BSM2
triplet,third,back	-26.042	-5.391	air
doublet,first,front	-4.350	-0.200	BAL35Y

doublet,first,back	-3.436	-0.003	Q2-3067
doublet,second,front	-3.437	-1.737	FPL51Y
doublet,second,back	16.915 (asphere)	-2.104	air
flattener,first,front	5.323	-0.080	PBM2Y
flattener,first,back	29.111	-0.332	air
flattener,second,front	4.832	-0.15	LAL59
flattener,second,back	flat	-0.144	air
CCD	-91.399	0	

Optical Performance

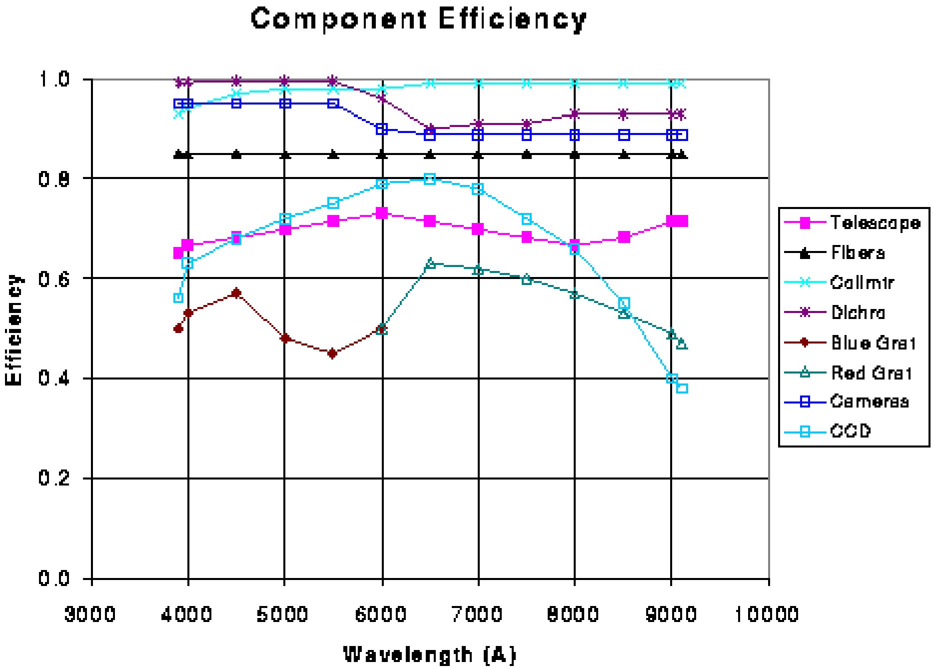
The spectrograph sensitivity is controlled by the grating throughput and the CCD quantum efficiency. We are fortunate to have excellent blue response in the CCDs. Figure 7.8 shows the throughput estimate. This is for photons falling within the 3 arcsec fiber input aperture and does not include losses at the telescope mirror coatings or light that misses the fiber. Figure 7.9 shows the detail efficiencies of the spectrograph components.

Figure 7.8



Spectrograph throughput. This calculation excludes telescope losses.

Figure 7.9



Spectrograph component efficiencies.

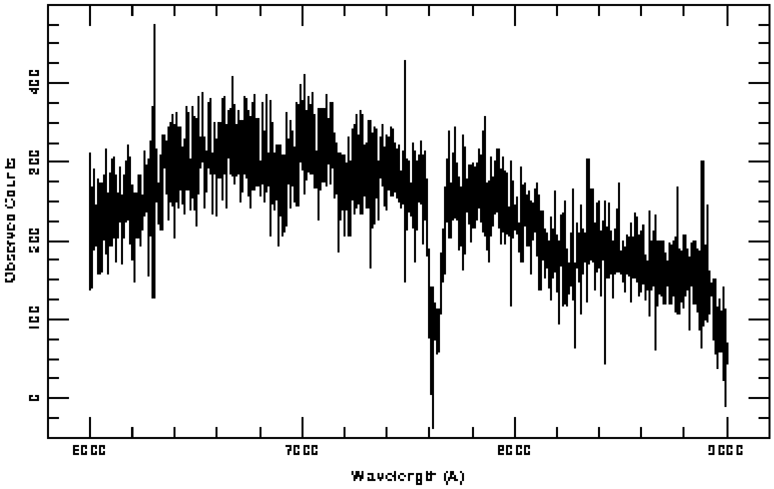
Although the input apertures cover three pixels on the detector, we need better resolution in the optical system to minimize crosstalk and improve signal-to-noise. The spectrograph optical design meets this goal nicely, producing rms spot diameters about the size of a pixel and resulting in nearly zero crosstalk from adjacent fibers. Table 7.2 shows rms spot diameters over the field of the blue CCD. The red performance is similar. The wavelengths (in A) are listed along the top and the slit position, measured in mm from the slit center, along the left. The slits are 4.9 inches high so the 60 mm field position is near the end of the slit, which is imaged along the edge of the chip. The wavelength coverage in the blue channel is 3900-6100 A so the table covers one half the area of the CCD. Performance is excellent over most of the chip. The r.m.s. spot diameters are about the size of a single pixel.

Table 7.2: R.M.S. Spot Diameters

Slit Position (mm)	3900	4000	5000	6000	6100
0	20	17	24	21	22
10	20	17	24	21	22
20	20	17	23	21	22
30	21	17	21	22	23
40	21	18	20	22	23
50	21	18	19	23	24
60	22	20	20	24	27

Scattered light is a notorious problem for straight-through spectrograph designs and we do not anticipate perfection here. The gratings will probably be the primary source of scattered light and unfortunately, there is little we can do to control this. We will minimize the problem by careful attention to baffling and cleanliness; note that the instrument need never be opened to the outside atmosphere during normal operation. The interior will be purged with dry nitrogen, preventing condensation and mineral deposition, and the gas used on all actuators will be exhausted to the exterior.

Figure 7.10

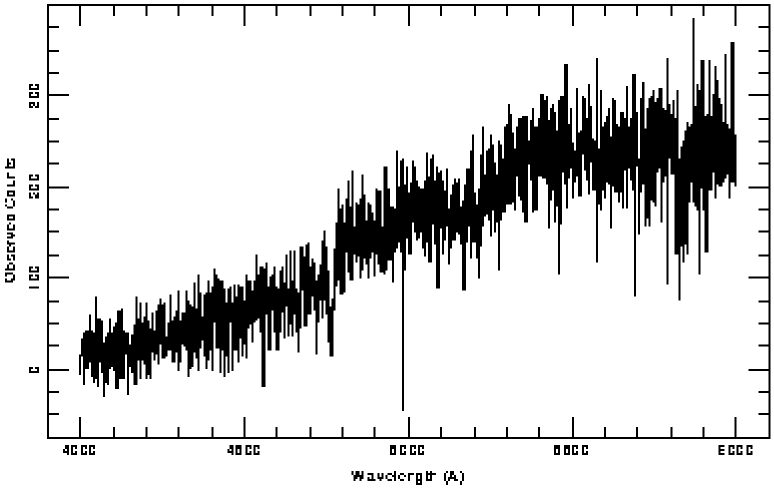


(first part of two-part figure) detected electrons with the averaged sky emission from three fibers subtracted for a g'=19.8 spiral galaxy at a redshift of 0.20.

Mechanical Design

Plug-plates

Figure 7.10



Simulated spectra. Spectra are shown for the red and blue detectors in A simulated sky-subtracted but otherwise raw spectrum of a g'=19.8 spiral galaxy at z=0.2 with this system is shown in Figure [7.10](); the top panel shows the blue half and the bottom the red. The H and K lines of CaII, the G band, the magnesium feature, and the sodium D lines are all clearly seen in absorption in this simulated spectrum of a galaxy considerably fainter than the survey limit.

The telescope optical system is a simple, fast, large field design with a focal surface flat to 2.6 mm but one where the principal ray deviates from the normal to the best-focus surface by up to 37 milliradians. For highest efficiency, the ends of the optical fibers should be positioned on the best-focus surface with their axes aligned with the principal ray. It turns out that plug-plate technology can be made to satisfy these criteria quite nicely.

The plug-plates will be aluminum alloy 2024-T3, 3.2 mm thick and 0.813 m in diameter. By applying bending moments to the edge of the plate (beyond the field of view), finite element calculations show that it can be deformed to match the best-focus surface to an area-weighted 62 microns r.m.s. The greatest departure from the best-focus surface is 200 microns and occurs at the center where the images are the best. Overall, the images are not significantly degraded from the best-focus surface.

As deformed to match the best-focus surface, the hole axes should line up with the principal ray axes. This is straightforward to accomplish if the plug-plate is deformed (in the opposite sense) over a properly curved mandrel, for drilling. If this is done, the drilling can be performed using a three-axis Computer Numerically Controlled (CNC) milling machine, i.e., it is not necessary to tilt the drilling head or the plug-plate.

Drill test results indicate that holes can be drilled with an accuracy of 9 microns r.m.s. in position and 4 microns standard deviation in diameter using short high-precision spade drill bits in a custom-made collet. In the test, four different bits were used to drill 50 holes each. The drilling time was 5.8 sec/hole. No significant degradation in drilling accuracy was observed for a range of slopes in the work-piece surface from 0 to 70 milliradians.

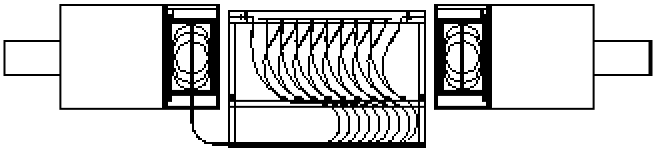
The plug-plates have a mass of only 4.3 kg. The plug-plates are thin enough so that the bending stresses, forces and material costs are reasonable. They are thick enough to provide hole depth adequate to constrain the plug angular alignment with the hole and to prevent significant gravity-induced deflections.

Fiber Cartridges

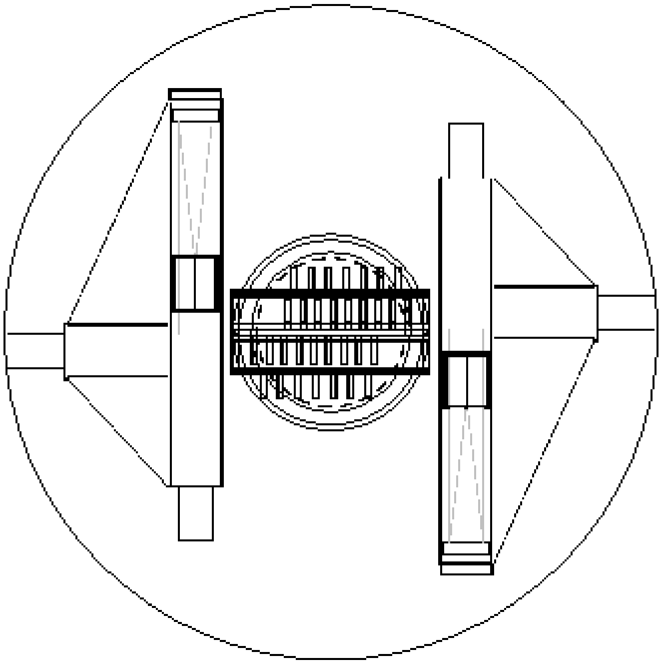
The fiber cartridge consists of a frame that supports the optical fiber harnesses, spectrograph slithead, and plug-plate holder (see Figure 7.11). The plug-plate holder consists of two large rings that warp the plug-plate to match the telescope best-focus surface. The upper ring includes a kinematic interface to the telescope to allow repeatable positioning of the cartridge on the telescope. By assembling the components into a cartridge, they can be handled as a unit, thereby reducing the complexity of the plug-plate changing operation. In particular, this approach addresses the following issues:

- The optical fibers are quite fragile and must be protected during transport to and from the telescope.
- The time available to change plug-plates is limited and must be performed with minimal lighting to avoid affecting neighboring telescopes.
- Cartridge storage space is limited.

Figure 7.11



Side View
(only 9 fiber harnesses shown)



Top View
(fibers not shown)

Fiber Cartridges shown with spectrographs. The large circle is the outline of the instrument rotator.

Ten cartridges are planned, enough so that plugging need not overlap observations except on the longest, darkest and clearest nights. This, in turn, allows plugging to be performed by day-shift personnel, is likely to result in higher reliability and better staff utilization, allows some work-load leveling, and allows time for the cartridges to thermalize completely between plugging and observing.

During the day, for each cartridge that was used successfully the previous night, the plug-plate will be unplugged and removed and an unused plug-plate will be installed and plugged. On average, over the life of the survey, about two cartridges per day will need new plug-plates. Our experiments with a mockup and the experience of other groups with operating systems indicate that this will take about two person-hours per cartridge (Limmongkol et al. 1993). Since a full time equivalent (FTE) employee works only about 4.8 hours per day averaged over all the days in the year, the labor requirement is about 83% of a FTE, and the operating cost is quite acceptable. Unfortunately, the work-load is expected to be very uneven, peaking during dark time in February.

While staffing for this task is likely to evolve, we expect to initially hire two people with primary responsibility for plugging, with secondary responsibilities to survey and observatory operations. Other members of the observatory staff will assist, as needed, in plate changing during peak load periods.

The correspondence of fibers to the holes in the plug-plate is determined by a device called the plug plate mapper, which illuminates each fiber sequentially from the spectrograph ends. The illuminated fiber will appear as a bright point against a dark background. A CCD camera will be used to determine the location of each fiber and to verify its position and throughput. This is a function of the plugging station, occurs unattended, and should require about 5 minutes.

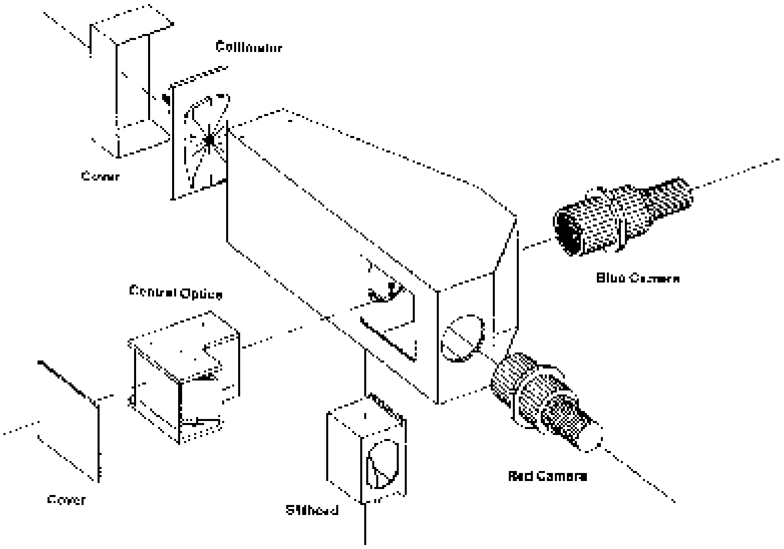
All cartridge operations occur at the same level, i.e., the telescope platform and the adjacent support building. The cartridges are assembled in the plugging room of the support building. They are stored in a space with doors to both the plugging room and the outside. At night, the outside door is opened to allow the cartridges to equilibrate to the temperature of the ambient air. Each cartridge, before use, is moved on a cart to a holding area near the telescope, where further equilibration occurs. The telescope is pointed to the zenith for removal or installation of cartridges. An empty cart is rolled under the cartridge to be removed. A lifting mechanism built into the fork base lifts the cart into contact with the cartridge so that the cart is supporting the weight of the cartridge. The cartridge is detached from the instrument rotator and the cart/cartridge assembly is lowered to the fork base. The procedure is reversed to install the new cartridge. Locating surfaces and sockets guide the cart and cartridge into the proper orientation with respect to the telescope.

As the cartridge is lifted into place and clamped to the telescope, the slitheads are simultaneously inserted into sockets in the spectrographs. The slitheads are attached to the cartridge frame by stiff springs so that they can move slightly with respect to the rest of the cartridge. Once the cartridge has been correctly positioned and clamped to the telescope, the slitheads are loaded against three point kinematic mounts on the spectrographs by pneumatic clamps. Each slithead will be coded and its identification relayed to the observer's workstation when it is inserted. This allows adjustments for each slithead, e.g., image placement on the CCD and focus, to be made automatically.

Optical Bench

The optical bench maintains the optical alignment. Figure 7.12 shows how the major optical components are attached to the optical bench, which is the large boxy piece in the center.

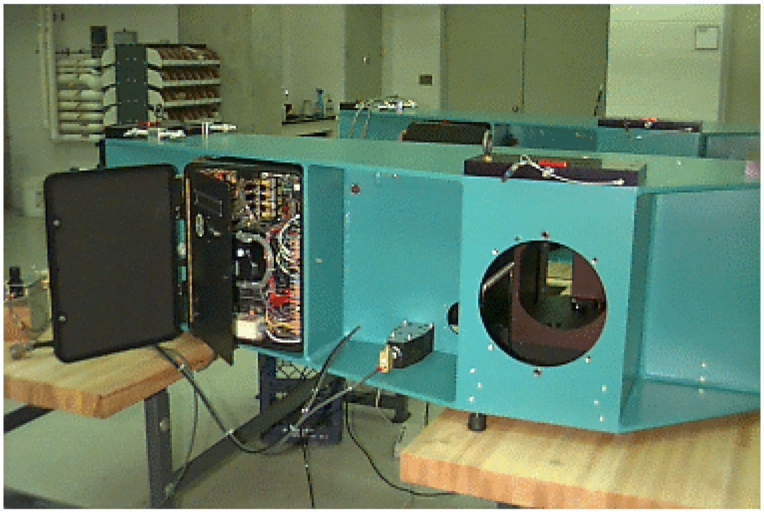
Figure 7.12



Parts of the spectrograph.

The final product is shown in Figure 7.13. The structure was welded, stress relieved, heat treated and painted.

Figure 7.13

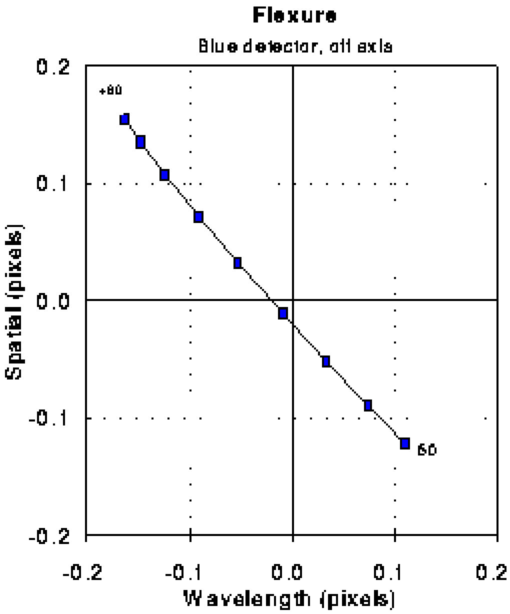


Optical bench. The large hole will hold the blue camera. Electronics are seen on the left and the small box near the bottom center is the slithead pneumatic clamp.

Because the spectrographs are mounted on the telescope, low flexure is important. We will allow up to 1/10 pixel flexure at the CCD due to the optical bench during any one hour exposure.

A finite element engineering model of the optical bench was used to examine the flexure properties and tune the design. Our requirement of 1/10 pixel flexure due to the optical bench in a one hour exposure was met by a box made of 1/4 inch thick aluminum with internal reinforcements and an optimized mounting system. Figure 7.14 shows the flexure performance. In this test, the instrument was rotated about its long axis and the position of a spot in an off-axis corner of the CCD was observed. Figure 7.14 shows the spot position with the spectrograph rotated 60° and turning to -60° at 15° intervals. The worst case movement through 15° (or a one hour exposure) is considerably less than 1/10 pixel.

Figure 7.14



Flexure performance of the optical bench.

Because the telescope focal plane is coupled to the spectrograph through flexible optical fibers, there is no need to fix the spectrograph rigidly to the telescope. Instead, mounting fixtures that accommodate the different coefficients of thermal expansion of the steel used in the telescope and the aluminum of the spectrograph will be used.

Slithead Mount

The slitheads are latched into three point kinematic mounts on the spectrographs and are held in place by pneumatic clamps. Each slithead will be coded and its identification relayed to the observer's workstation when it is inserted. This allows slithead-specific adjustments, image placement on the CCD and focus, to be made.

Hartmann Masks

The spectrograph focus will be determined using the usual Hartmann test. Pneumatically actuated collimator masks take the form of "saloon doors" located immediately in front of the collimator (these are not shown in Figure 7.6).

Focus adjustment is required because of the expansion and contraction of the aluminum optical bench, which moves the slithead relative to the collimator mirror. By moving the collimator to compensate, we keep the slithead in focus. Temperature sensors will allow us to read the focus value from a table after initial measurements are made. Each slithead is expected to have a slightly different zero-point focus because of manufacturing variations, so each will be encoded to allow adjustment for personality.

Collimator Mount

The collimator mount is complex. The collimator must move in and out (piston) to correct for expansion and contraction of the aluminum optical bench due to changing temperature. It must also pitch and yaw with one arcsec precision to allow widening the flat field spectra and positioning the spectra correctly on the CCD. Small variations between slitheads require the ability to position accurately the image on the CCD to avoid, for example, a bad pixel. These motions, which we would like to control to 1/10 pixel, require linear resolution of one or two microns.

Our design uses three DC servo motors at the mirror mounting points driven by an external controller (motors and controller by Physik Instrumente). A small control computer on the spectrograph commands the motors through an RS232 link to the motor controller board.

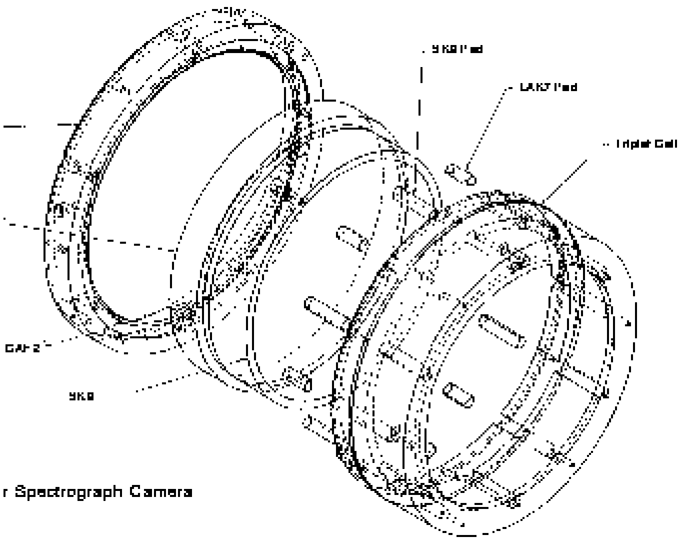
Central Optics

The central optics include the dichroic beamsplitter and the grisms. Each optic is set in a six-point kinematic mount and the cell, in turn, is fixed in a three-point mount within the spectrograph optical bench. Alignment is achieved with tight machining tolerances and custom machined inserts so no screw adjustments will be necessary.

Camera Cells and Housing

The camera lens cells and housing are adapted directly from Michael Carr's similar, but larger, LRIS and Norris spectrograph cameras. The cells follow conventional optical mounting practices except in the radial supports of the large lenses. Here, six precision machined inserts of glass-filled Teflon center the lenses on the optical axis. The more common practice of building a mold of rtv around the lenses has the disadvantage of not being able to disassemble the camera easily. Figure 7.15 shows the concept.

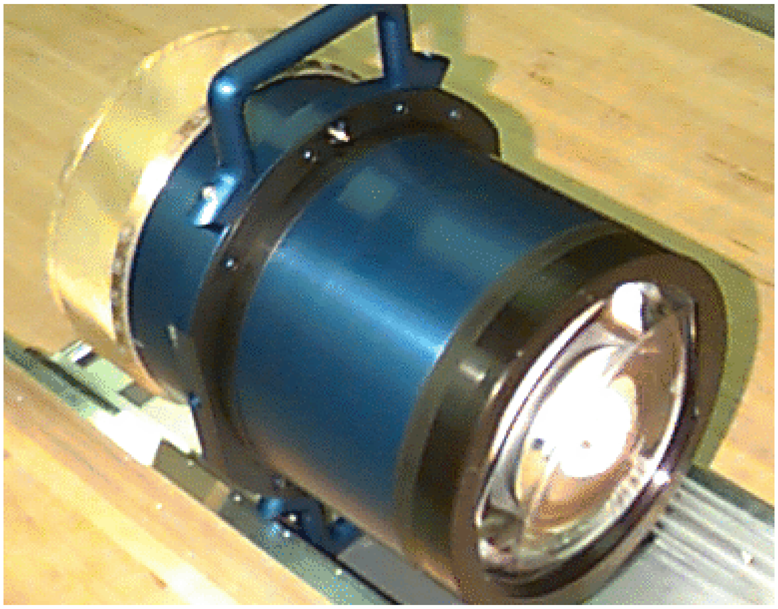
Figure 7.15



Triplet element lens cell.

We have assembled one camera (Figure 7.16). Initial tests show that its optical performance is exactly as predicted by the raytrace.

Figure 7.16



Spectrograph camera

Focus and Shutter

The shutter is located immediately in front of the central optics. Because timing resolution is not critical, we will actuate the shutters pneumatically, opening or closing them within about 300 ms.

A manual focus mechanism moves the dewars (plus field flatteners) with respect to the cameras. This adjustment is suitable for initial camera setup (to correct for red/blue camera differences) and to accommodate focus changes that might be necessary because of detector replacements. Note that because the cameras are designed to be parfocal over our operating temperature range, routine focus of the spectrographs can be done by adjusting the collimator.

Dewars

The CCDs are mounted in custom dewars using the same ball and socket adjustment fixture found in the photometric camera. The dewars can be physically small because they use an autofill liquid nitrogen system and do not need a large capacity.

Controls and Software

Observer's Program

All spectrograph operations are controlled from a single program running on the observer's workstation. This program, the observer's sole interface to the spectrographic system, is a clearinghouse for observing commands that translates observer's requests into the commands required by the independent subsystems used for spectroscopy. These systems include:

- Spectrograph microprocessor
- Telescope and guider
- CCD data acquisition system
- Drilling database
- Plugging station

Microprocessor

The spectrograph mechanical operations will be controlled by Z-World Little Giant Z180 microprocessors, one on each spectrograph. The board was chosen because it is programmable in C and has the right number and type of I/O ports and A/D converters. They communicate with the observer's program through an RS-232A serial port. The processor controls all the mechanical functions on the spectrograph (shutters, Hartmann masks, collimator tip, tilt, and focus) and monitors ambient and instrument temperatures for use in adjusting focus and image positioning.

Telescope

The observer's software communicates fully with the telescope control system via an Ethernet telnet connection. One might imagine that when a new plug-plate is locked on the telescope, the spectrograph relays the plug-plate ID to the observer's software, which looks up the coordinates in the plug-plate database. The observer's software commands the telescope to move to the mid HA position for flat field and wavelength calibration, then to the field for precise positioning, scaling, and focusing. When these operations are finished, the observer's software starts the guider, opens the shutters, and begins observing.

Data Acquisition

The data acquisition hardware for the spectrograph CCDs is a clone of the photometric camera system. The observer's program actuates the system through remote procedure calls (RPC) over the Ethernet. The commands are few and simple, only those required to prep the CCD, expose, and read out the data. Once the data are off the chips and in the data system buffer, the images will be downloaded over a high speed VME link to the observer's workstation where they will be written to disk. At the end of the night, a tape is written and sent to Fermilab.

Drilling Database

The drilling database is the list of objects on a plug-plate. It includes the name and drilling coordinates of each object and the field coordinates (alpha , delta , equinox) of the plug-plate. This information is needed at observing time so the telescope can be pointed and the plug-plate ID can be incorporated into the data file. This database is generated at Fermilab from the photometric imaging data and is delivered on the internet to the observer's workstation at Apache Point well before the observations.

Plugging Database

To keep track of which fiber went into which plug-plate hole, the plugging station will be equipped with a device to map the plug-plate locations to the slithead after plugging is finished. The plugging database information is written directly to the observer's workstation disk at plug time. This information is merged with the CCD data before it is written to disk.

Spectroscopic Observing

Here is a possible observing sequence. We anticipate taking both flat field and wavelength calibrations before every exposure, though the latter may prove superfluous. These will be obtained at the anticipated mid hour angle of the exposure to account for the effects of flexure, both in the spectrograph and the fibers.

For the calibration observations, a white screen is drawn across the top of the telescope. The screen is illuminated by calibration lamps mounted on the telescope structure. Flat field exposures will probably be of two kinds: one is designed to correct for detailed pixel-to-pixel variations on the detectors. These will be taken with the collimator mirror nodding slightly (to move the image a few fiber diameters perpendicular to the dispersion direction) in order to uniformly illuminate the detector. The others, which might even be binned vertically (the dispersion direction) during readout, will not be nodded and are intended to calibrate the fiber throughput as a function of wavelength. Both flat-field and calibration exposures will be read in "quick" mode with reduced sample times (but with the CCD transfer timing the same as for normal readout) to reduce the readout time to 30 seconds.

After the calibration exposures, the telescope is moved to the field. Ten coherent fiber bundles feed images of setup stars to a television camera mounted on the spectrograph. These setup stars are used to center the field on the plug-plate, adjust the scale of the telescope (by simultaneously adjusting the primary and secondary mirrors), and set the telescope focus. Once these operations are automatically done, the guider is started. A series of four short (1 minute) exposures with the telescope offset 1.5 arcseconds in each direction in altitude and azimuth build up a 2 x 2 raster of the plug-plate field. These exposures will be binned heavily in the dispersion direction on readout; the data will be used to check for systematic errors in the drilling system and provide spectrophotometric calibration information. Next, three nominally 15 minute exposures make up the primary spectroscopic observation. Three exposures allow efficient cosmic ray discrimination, though it is possible that sufficiently powerful software will work well with two or even one; we will pursue this issue during the test year.

The sky brightness will be monitored with a guide camera through a wide-field (30 arcsecond) fiber bundle placed at a blank location in the field. This information, combined with the transparency and seeing data available from the ten guide stars will allow us to control precisely the exposure time needed to achieve the desired signal to noise ratio. The exposure times will be adjusted for atmospheric extinction, sky brightness, and galactic extinction with the aim of producing insofar as possible a uniform limit to the survey *outside the galaxy*; the range in exposure times is close to a factor of two.

Thus the sequence (with estimated timings) is as follows, starting with the telescope at the zenith between exposures, and assuming that the previous exposure has just ended:

1. Remove the old plugplate and install the new one (5 minutes -- this is our current target time, and it appears not unreasonable).
2. Slew to the position of the anticipated center of the new exposure. (1 minute) The flat-field screen is being drawn over the aperture of the telescope during this move.
3. Take and read the wavelength calibration exposure (1 minute).
4. Take and read the flat field exposure (1 minute).
5. Move to the position of the field, acquire the guide stars and make the necessary adjustments to position, rotation, and scale (2 minutes).
6. Take the four offset raster exposures (4 minutes total; the read time and offsetting time are negligible).
7. Set back to the center of the field, turn on the guider, and take and read the first exposure (15 minutes exposure plus 1 minute read).
8. Second exposure (16 minutes).
9. Third exposure (16 minutes).
10. Slew the telescope to the zenith in preparation for the plugplate change (1 minute).

The total time for this sequence is 63 minutes.

At the end of the series of observations, the data are collected, merged with the database information that identify the field, plug-plate, and fiber assignments, and written to disk. At the end of the night, all of the data are written to a DLT tape and shipped to Fermilab.

References

Limmongkol, S., Owen, R.E., Siegmund, W.A., and Hull, C.L., 1993, ASP Conference Series, Volume 37, *Fiber Optics in Astronomy II*, (BYU: Provo, UT), 127.

Tiling and Adaptive Tiling

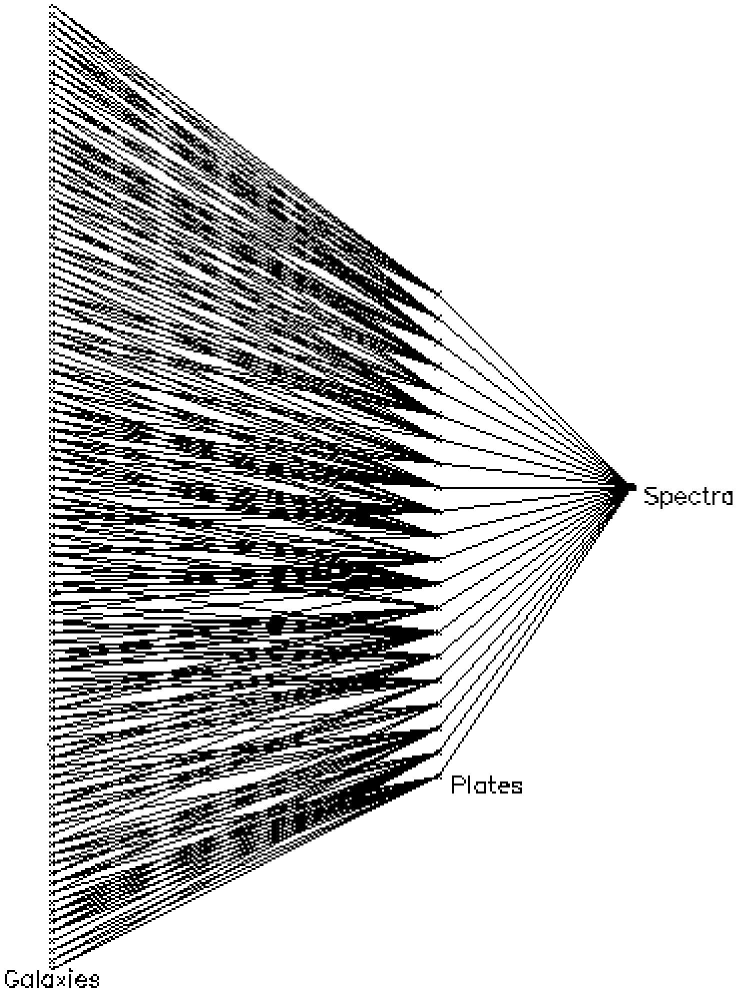
If galaxies were randomly distributed on the sky with a surface density of 100 galaxies per square degree, then each 3° circular field would be expected to contain 707 +- 27 galaxies. In fact galaxies are considerably more clustered than this, and many fields contain many more galaxies than we have fibers. While such fields could be repeated to obtain the remaining redshifts, the time overhead would be prohibitive, close to 60%. Therefore, instead of using a uniform grid of field centers which uniformly covers (`tiles') the sky, we plan to use adaptive tiling, in which the field centers are moved more closely together in regions of high galaxy density to provide greater coverage of these regions. It would obviously be desirable to arrange the plates in such a way that all density enhancements lay in the areas where many fields overlapped. In addition, if it could also be arranged that all pairs of galaxies closer than 55" , the limit set by the fiber size, were in the overlap regions, we would be able to circumvent this limit. In the following we will speak of the 3° fields as "fiber plates" or simply "plates."

In one dimension the optimization problem is trivial, but in two (or more) dimensions it may be shown to be NP-complete (which means for our purposes that there is almost certainly no efficient algorithm for solving it exactly, although there may be good approximate methods). We approached the problem by first solving the plate-assignment problem for given plate centers, in itself a non-trivial problem, and then generalizing our solution to find a near-optimal set of field centers.

Assigning fibers to plates, given the plate centers

The plates are circular and hence overlap; thus if one is given a set of fiber plate positions many of the galaxies will appear in the regions where two or more plates overlap, and for each of these galaxies one must decide in which plate to take its spectrum. Because one is using a fixed number of fibers such decisions can affect the total number of spectra that we can obtain with a given number of exposures ("If I take this spectrum with plate A that frees up a fiber in plate B, so I can deal with that galaxy in plate B not plate C, so ...").

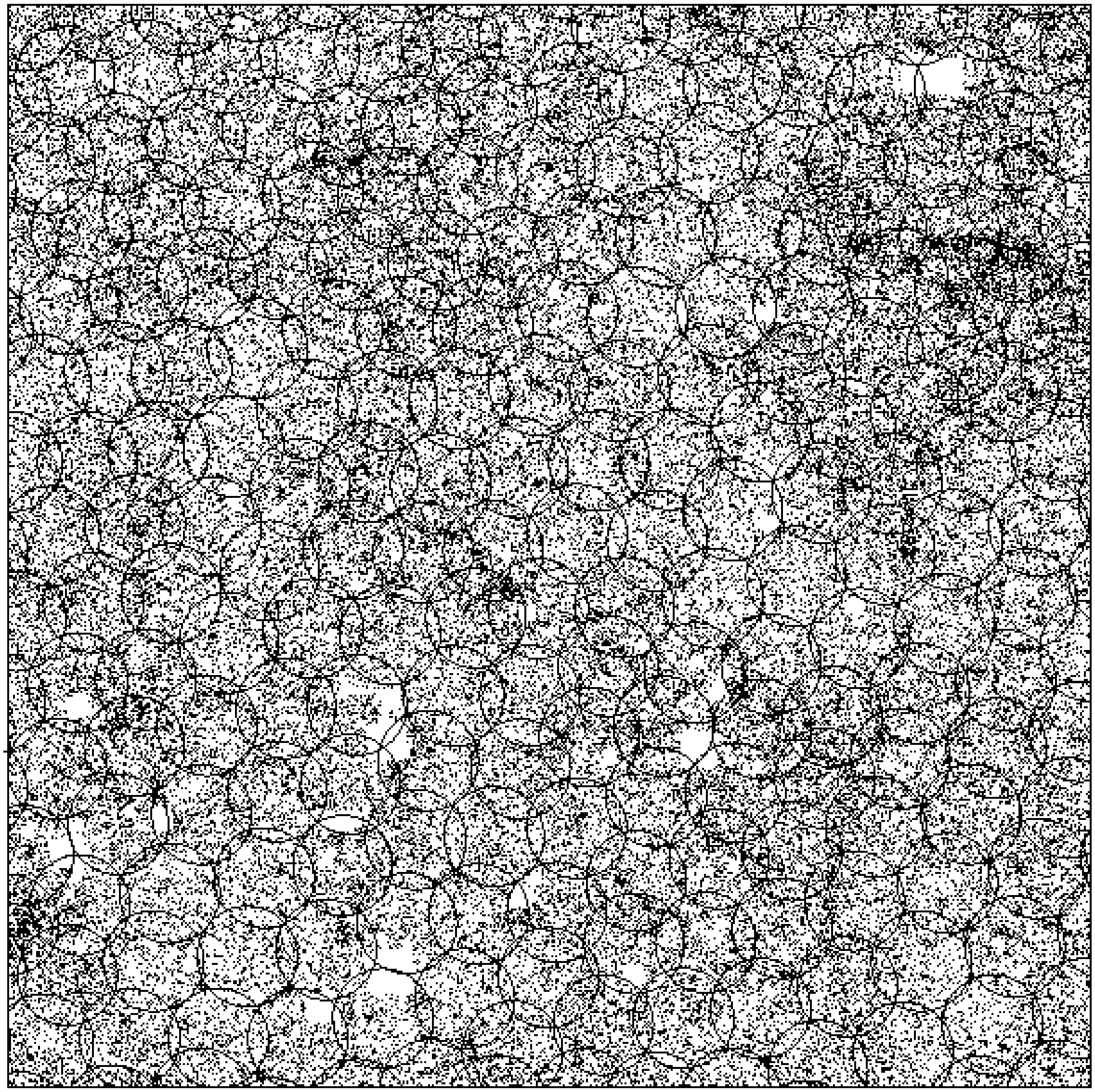
Figure 8.1



How the tiling problem maps onto a network flow. The leftmost column of crosses represents the galaxies, and each is connected to the one or more plates in which they lie. Each of these lines has a flow capacity of 1. Each plate is connected to a point labeled `Spectra'; each of these lines has a capacity equal to the number of galaxy and QSO spectra that we can take at a time, approximately 610 in the real survey.

This problem can be reduced to a network flow problem (e.g. Papadimitriou and Steiglitz, 1982) and solved efficiently in polynomial time. The construction consists of noticing that each galaxy can be thought of as a source for 1 unit of flow of some fictitious fluid, and that each plate can be thought of as a pipe that allows no more than n units of flow (where n is the number of galaxies per plate). Each galaxy is attached to each plate that it lies in (see Figure 8.1); if we can find a flow pattern that satisfies our constraints we have solved the problem. Fortunately such flows constitute a well-studied field of computer science and good algorithms exist for solving them in a time that scales as some low power of the number of plates (or, more precisely, the number of regions defined by the overlap of two or more plates).

Figure 8.2



The distribution of plates after running the adaptive tiling algorithm. This run used test data consisting of galaxies from the APM Galaxy Survey and QSO candidates distributed at random. Out of 365,931 targets, 365,711 are successfully assigned to 651 plates, each of radius 1.5° and capable of taking 610 spectra. This plot shows just the central 36° x 36° area. The blank regions are holes drilled out around bright stars.

In order to test our implementation of the flow problem, we have generated a test data set of 365,931 galaxy and QSO targets covering an area of sky of solid angle 1 steradian (just under a third of the area of the northern survey). The galaxies were selected from the APM Galaxy Survey (Maddox et al. 1990) and so reflect true clustering in the Universe. QSO candidates were generated at random, and comprise 15% of the targets. Each spectroscopic plate is capable of taking 640 spectra. Thirty fibers per plate are reserved for sky and for standard stars, leaving 610 fibers per plate for scientific use. Ignoring for the moment the fact that we cannot observe both members of a pair of galaxies closer than the minimum fiber spacing on one plate, using a near uniform covering of 651 plates enables 97.2% of targets to be observed. The remaining 2.8% of unobserved targets are far from randomly distributed and so would significantly bias measures of galaxy clustering. One can observe 99.5% of targets by increasing the number of plates to 770, but this is an additional overhead of almost 20%. Scaled to the SDSS area of pi steradians, this represents a year of operation and roughly three million dollars.

Adaptive Tiling: Finding a Good Set of Plate Centers

It has seemed to us that if we use only slightly more than the minimum number of plates, but allow the centers to move in such a way that we can have a higher density of plates locally to cope with high density regions, we might be able to do better.

In the network flow problem described above, we have no way of telling which perturbations of the plate centers would help and which hinder, but a slight modification of the flow problem provides this information; instead of assigning each galaxy to the set of plates in which it lies, assign it to *all* plates (or at least all the nearby ones), with an associated cost which increases sharply as the galaxy moves away from the plate. Then instead of solving the simple network flow problem, find the flow with the lowest total cost; again this is a standard problem in computer science and good solutions exist. We now know which galaxies are forced to be assigned to distant plates, so we can rearrange the plates accordingly and iterate towards a solution.

Starting with the uniform cover of 651 plates, the adaptive tiling code soon converged to a solution whereby 99.9% of targets were assigned a fiber; the resulting distribution of plate centers is shown in Figure 8.2. In practice, the observing strategy of interleaving photometry and spectroscopy will present us not with large, square areas for tiling, but with long, narrow regions elongated in the scan direction. We have investigated the effects of this by tiling the test data in a series of 10° wide regions. In order to minimize edge effects, we do not drill plates which extend over an internal boundary between tiling regions, but include the untiled targets left over from one run of the tiling code with the succeeding run. We lose a little efficiency (~ 5%) by tiling in sub-regions, but this is still much more efficient than using a fixed tiling pattern.

The Effect of the Finite Fiber Separation

In the plug-plate scheme, the plugs which hold the fibers will be 3.2 mm in diameter, which projects to 55 arcseconds on the sky. This clearly represents the minimum fiber separation; it corresponds to a distance of about 100 kpc $(H_0 100)^{-1}$ at the survey limit of about z=0.2 . If the galaxies were randomly distributed, the probability of another fiber plug overlapping a given one is about 3%; this is increased to about 7% on average by the correlations in the counts to this limit. Most of the objects lost are members of binaries. The losses are worse in the densest regions, such as the centers of rich clusters at intermediate distances.

These losses do not affect subsequent large scale structure studies, because if we are unable to target both galaxies of all close pairs we must remove all such pairs from the analysis. However, there are other problems, most particularly the dynamics of binary galaxies and the study of subclustering and velocity dispersions in rich clusters for which this minimum separation is a real hindrance. It is important to note, however, that more than 40% of the sky is covered by two or more plates. In these overlap regions, we can observe nearly all pairs of galaxies closer than the minimum fiber separation by assigning each galaxy to a different plate. Currently, this re-allocation of fibers to close pairs in overlap regions is carried out by a process which runs *after* the adaptive tiling, and enables us to observe 95.6% of targets. We are in the process of incorporating the minimum fiber spacing constraint into the tiling algorithm itself, which will hopefully allow us to observe an even higher fraction of galaxies in close pairs by placing plate overlaps in regions where there is a high density of close pairs.

References

Maddox, S.J., Sutherland, W.J. Efstathiou, G., and Loveday, J., 1990, MNRAS 243, 692.

Papadimitriou, C., and Steiglitz, K., 1982, *Combinatorial Optimization, Algorithms, and Complexity*, Prentice Hall.

Simulations of Survey Data

We have carried out extensive efforts to simulate the data that will be produced by the SDSS. These simulations serve several interrelated purposes. First, they provide data that we can use to test the survey's data acquisition and reduction software. Second, they help us make informed decisions about survey strategy -- target selection criteria, geometry, scheduling, and so forth. Third, they offer qualitative and quantitative predictions about what the SDSS should see, under specified theoretical assumptions. Fourth, they provide an artificial database that we can use to develop and test methods that we will eventually use to extract scientific information from the survey itself. Because the survey test year is fast approaching, we have so far focused our attention on the first of these applications, testing survey software. Once the survey is underway, the fourth application will assume great importance, especially for large scale structure studies. We will use simulations like those described below to calibrate uncertainties and check for biases in our methods of analysis, and to test the ability of the analysis to distinguish theoretical models of the large scale structure.

There are a number of survey software tasks that require simulated data. Of course we want to test the software on real data as well, but the real data will not be available until the survey telescope is in operation, and it is in any case useful to have some test data for which the "right" answer is known in advance. At the most nitty-gritty level, we need simulated data to test the photometric reduction software, everything from flat-fielding the CCD images to classifying objects and selecting spectroscopic targets. For this purpose, we need artificial images with realistic numbers of objects, realistic magnitudes, colors, sizes, and profiles, realistic clustering, and realistic seeing, noise, and defects. In the end, we can test for any biases in the selection of spectroscopic targets by passing artificial catalogs through the complete data reduction pipeline and comparing the actual target list to the list of targets that would have been selected if the data reduction were perfect.

For many other tests, we need only simulated catalogs , i.e. lists of objects with photometric properties and positions on the sky. We will use such catalogs to help us decide the selection criteria for spectroscopic targets. For galaxies, we want to know how different choices of selection criteria will affect our ability to measure large scale structure, and for quasars we want to estimate the impact of stellar contamination on the target lists as a function of Galactic latitude and longitude. We will also use simulated catalogs to test the efficiency of the tiling software described in Chapter 12, and to study the effects of the minimum fiber spacing on our ability to sample galaxy clusters, groups, and binaries. The simulations also allow us to examine the possible impact of photometry errors and uncertainty in Galactic extinction on measurements of galaxy clustering.

We have designed a three-tier system for creating simulated data; we have implemented many parts of this system, and are continuing to add new elements and refine existing ones. The first tier creates artificial sky catalogs -- lists of objects with positions and photometric parameters. We use a large, cosmological N-body simulation to provide a clustered galaxy distribution, and the Bahcall-Soneira (1984) model of the Galaxy for star counts. The second tier turns the catalog for a specific patch of sky into artificial images in the survey filter bands. Spiral galaxies are represented by inclined exponential disks, elliptical galaxies and bulges by spheroids with a de Vaucouleurs-law profile, and stars and quasars by point sources. The images are convolved with a point-spread function (PSF), and noise and other artifacts (cosmic rays, diffraction spikes, etc.) are added, as are the defects of real CCDs (see below). The third tier of the simulation system turns a series of artificial images into a data stream that looks like it is coming from the telescope's data acquisition system. We can thus run simulated data all the way through the photometric data reduction pipeline. The next section describes the first tier of this system, sky catalogs, in greater detail. Section 9.2 describes the other two tiers, in particular our techniques for creating artificial images.

Simulated Catalogs

A complete simulated catalog would list positions and parameters of galaxies, stars, and quasars, about a magnitude deeper than the depth of the photometric survey, covering the whole SDSS footprint on the sky. Such a catalog would be rather unwieldy, and in practice we use programs to create either a photometric-depth catalog over a rather limited area of sky or a catalog over the whole survey region but only to the depth of the spectroscopic survey. It is nonetheless useful to think in terms of "the" catalog that in principle contains all of the objects that the survey might observe in a simulated universe. Our convention is that the catalogs list "intrinsic" apparent parameters, with no instrumental or atmospheric effects, the sort of numbers that we want our data reduction software to produce and store in the data archive.

Galaxies are the most challenging component of the artificial catalog, since we require a population of objects with realistic spatial clustering, luminosities, diameters, and colors. Trickier still, we need to include any important *correlations* between these properties, to the extent that they are known observationally. The basis of the galaxy catalog is a large N-body simulation run by Changbom Park and Richard Gott. The simulation uses Park's (1990) particle-mesh, N-body code to evolve 380^3 ~ 55 million particles on a 600^3 mesh that represents a comoving, periodic cube 600 h^{-1} Mpc on a side. The background cosmological model is a flat universe with Omega_0=0.4, a cosmological constant Lambda_0=0.6 , and a Hubble constant h=0.6 . The initial conditions are Gaussian fluctuations with a cold dark matter (CDM) power spectrum corresponding to these values of Omega_0 and h . The normalization of the initial fluctuations corresponds to a "bias factor" of 1.3 in spheres of radius 8 h^{-1} Mpc , and "galaxies" are represented by a weakly biased subset of the full particle distribution, corresponding to nu >0.7 peaks of the initial density field, so that their clustering amplitude matches observations. The number density of the biased particles is sufficient to represent galaxies brighter than about 0.064 L_*. There are just over 8 million galaxy particles in the simulation volume.

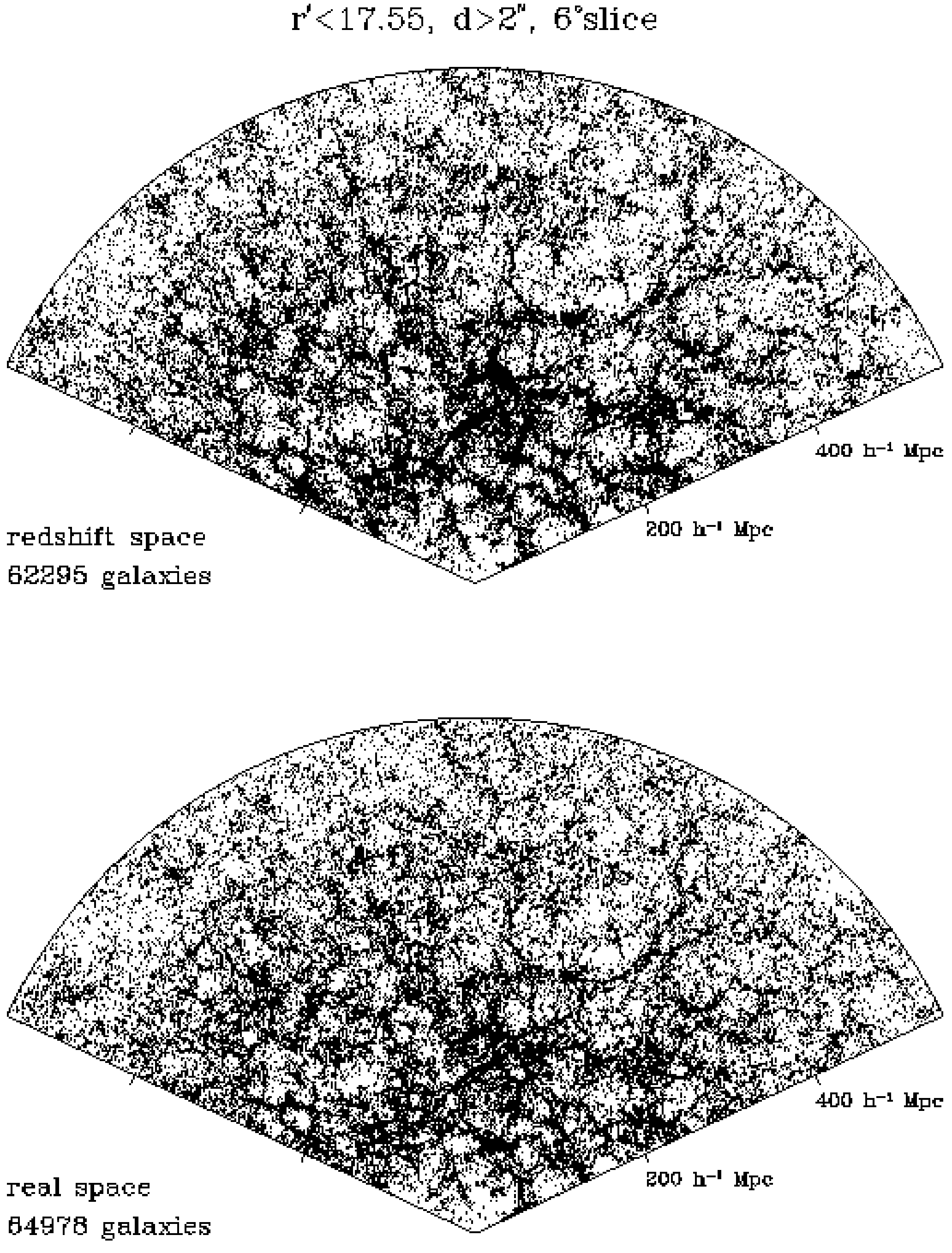
The logic behind using the CDM + Lambda model is simple: it produces a galaxy distribution whose clustering properties match existing observational data quite well over a wide range of scales. For the purpose of providing test data, the physical attractiveness or unattractiveness of the model is not particularly important -- we just want something that looks reasonably like the observed universe. To go from the distribution of particles in the N-body simulation to a galaxy catalog, we need to assign apparent fluxes and diameters in the survey bands, for an "observer" located in one corner of the simulation cube. Each galaxy particle is assigned an absolute blue (B_T) luminosity drawn randomly from a Schechter luminosity function. The luminosity distribution is cut off below 0.064 L_* (about 3 magnitudes below M_*) in order to match the galaxy density of the simulation. Each galaxy is also assigned a randomly chosen Hubble type from E to Sc, with relative probabilities determined from the local galaxy density using the observed morphology-density relation in the form given by Postman & Geller (1984). The galaxy can then be assigned an apparent flux in each of the survey bands using the appropriate colors and K-corrections; these are computed for each Hubble type using Coleman et al.'s (1980) galaxy spectrophotometry and the filter-response curves illustrated in Figure 4.1. To assign diameters in accordance with the luminosity, we use a mean relation based on the fundamental plane (for ellipticals) and a Freeman law (for disks). We add scatter about these mean relations, so that our simulated catalog includes both low surface brightness galaxies and compact, high surface brightness galaxies.

The N-body simulation is large enough to contain almost the full galaxy redshift sample, without periodic replications. To reach the depth of the photometric sample, we must wrap around the periodic boundary several times. A random line of sight through the box is unlikely to pass near itself on later traversals, so this is not a serious problem.

For this example, we adopted an apparent magnitude limit of 17.55 in the r' band, and a minimum angular diameter (at the half-light radius) of 2 arcseconds (these simulations were done based on an earlier version of the target selection criteria). These limits yield just under a million galaxies over the survey region; about 2% of the galaxies brighter than the limiting magnitude are rejected because they are smaller than the imposed minimum diameter.

The upper panel of Figure 9.1 is a 6 degree by 130 degree redshift-space slice. The lower panel shows the same slice in real space, i.e. with peculiar velocities set to zero. The redshift-space slice clearly shows "fingers-of-God" caused by velocity dispersions in virialized groups and clusters. These thicken the extended filaments and sheets, making them more prominent.

Figure 9.1



Slices from a simulation of the SDSS redshift survey. The upper panel. shows a 6 degree by 130 degree slice in redshift space -- each point represents a galaxy, plotted at the distance indicated by its redshift. The lower panel shows the same slice in real space, with no peculiar velocity effects.

The star catalog is based on the Bahcall-Soneira (1984) model of the Galaxy, which provides star counts in V and B-V . These counts determine the local star density at specified latitude and longitude; we choose the actual number of stars in a given area by Poisson sampling this distribution. Stars in the spectrophotometric atlas of Gunn & Stryker (1983) define fairly tight relations between B-V and colors in the SDSS filters, so given a V magnitude and a B-V color we can assign magnitudes to a star in all five bands. We include binary stars with separations of 0.1 - 3 arcseconds and magnitude differences up to 2.5 mag using the normalization and separation distribution found by Gould et al. (1995) in the HST SnapShot Survey.

Our simulated catalog also includes quasars, for which a random angular distribution is an adequate approximation; the tricky thing is to get the appropriate distribution of colors, since this distribution is essential when we try to estimate the efficiency of selection criteria at differentiating quasars from stars. Another important element still to be added is Galactic extinction; we will use the simulations to check the procedures for computing and correcting this effect, as well as for checking the algorithms for selecting stars whose colors can be used to measure the extinction.

Simulated Images and Test Data

Images

The second tier of our simulated data system turns a catalog representing a specific patch of sky into artificial images in the survey filter bands, using code adapted from J. Gunn's Mirella/Mirage image-processing package. In the current version of the image-making code, an object is represented by a sum of up to three profiles: a PSF, an inclined exponential disk convolved with a PSF, and a de Vaucouleurs-law bulge convolved with a PSF. Each catalog entry specifies the fraction of light in each of these profiles (in each band), the effective radii and position angles of the bulge and disk components, and the inclination (for disks) or axis ratio (for bulges). We describe the adopted profiles below, then summarize our procedures for adding noise and defects to the images, and list some of the parameters that are used to define the images. We show a few representative images in Figure 9.2.2.

i) PSF

The simulations to date have used a PSF consisting of a double-Gaussian core with an r^{-3} power-law tail. Thus, the intensity pattern produced by a point source of total flux F_0 is

$$I(r) = \frac{F_0}{K}\left[\alpha e^{-r^2/2\sigma_1^2} + (1-\alpha-f)e^{-r^2/2\sigma_2^2} + f\left(1+\frac{r^2}{\beta\sigma_2^2}\right)^{-\beta/2}\right],$$

with alpha = 0.9, beta = 3, sigma_1= 0.94 (pixel), and sigma_2 = 2.7 (pixel). K is a normalization constant, and f determines the fraction of the total light in the power-law component, which we set to 0.075. With these parameters, the PSF corresponds approximately to 1 arcsec seeing. We can change alpha , beta , sigma_1, and sigma_2 so that we obtain images with various seeing conditions.

A recent paper by Racine (1996) finds somewhat shallower power-law wings (this is an important consideration for the effect of these wings on sky brightness fluctuations; see Chapter 10). The new simulations will use instead the PSF model suggested by Racine (1996), consisting of a superposition of three Moffat (1969) functions. This model produces a smoother PSF and a better representation of the observed power-law wings. However, the photometric pipeline (Section 10) will continue to characterize the PSF using the formulation given above because this separable function is computationally very efficient.

We add diffraction spikes for stellar sources. The intensity pattern produced by a diffraction spike is:

$$I_{\text{spider}}(r,j) = \text{spk}_{\text{amp}}(r) * \text{spk}_{\text{cross}}(j),$$

where

$$spk_{\mathrm{amp}}(r) = \frac{I_0 \gamma \delta}{\gamma^2 + r^2} ,$$

gamma = 10 (pixel), delta = 0.01, I_0 is the central intensity of the star, and

$$spk_{\mathrm{cross}}(j) = \frac{1}{\sqrt{2\pi}} \left[\alpha \sigma_1 e^{-j^2/2\sigma_1{}^2} + (1-\alpha) \sigma_2 e^{-j^2/2\sigma_2{}^2} \right] .$$

Here r represents a distance from the central star along an arm of the diffraction cross, and j represents a distance transverse to the arm. Thus $\mathrm{spk}_{\mathrm{amp}}(r)$ represents the "raw" intensity pattern of the diffraction spike, and the convolution with $\mathrm{spk}_{\mathrm{cross}}(j)$ incorporates the effect of seeing. For stars that saturate the full-well of the CCD, we add saturation tracks along the direction of the charge transfer (the scanning direction).

ii) Galaxy Profiles

Observed disk galaxies have profiles that are close to exponential out to about four scale lengths, at which point they tend to cut off rather sharply (van der Kruit and Searle, 1981). We therefore represent disks by a truncated exponential profile:

$$I(r) = \begin{cases} I_0 \left(e^{-r/\epsilon r_e} - e^{-4/\epsilon r_e} \right), & r \leq 4\epsilon r_e \\ 0, & r > 4\epsilon r_e . \end{cases}$$

Here epsilon = 0.59584 is the conversion factor between the exponential scale length and the effective radius r_e containing half the total light: r_s= epsilon r_e . For the inclined profiles we assume that the disks are optically thin. Ellipticals and bulges are represented by a de Vaucouleurs profile (de Vaucouleurs, 1948),

$$I(r) = I_0 e^{-7.6692 (r/r_e)^{1/4}} ,$$

The above profiles are convolved with the PSF before being added to the image. In practice, we interpolate and scale from a catalog of template images. In future versions of the simulations, we plan to represent the the brighter galaxies using suitably scaled CCD images of nearby galaxies instead of idealized models. (cf. Frei et al. 1996). We will also adopt a PSF form that corresponds more exactly to the optical design of the telescope.

iii) Noise and Defects

For each catalog and filter, we make images with varying degrees of noise and other defects. At the simplest level, we have noiseless images that precisely represent the objects in the input catalog. We can easily make images with Gaussian photon and readout noise. At a higher level of sophistication, we can make images with photon noise and a background taken directly from one of the test CCDs, including dark current and cosmic rays. We can also `un-flatten' images using flat-field vectors obtained from test CCDs. We have developed code to include in the simulations nastier problems such as the first and second order filter ghosts (which begin to appear around stars of magnitude 15^{m} and 7^{m} respectively), as well as satellite tracks, airplanes, and so forth. The photometric pipeline will have to identify and handle all of these complications, and we must test its ability to do so.

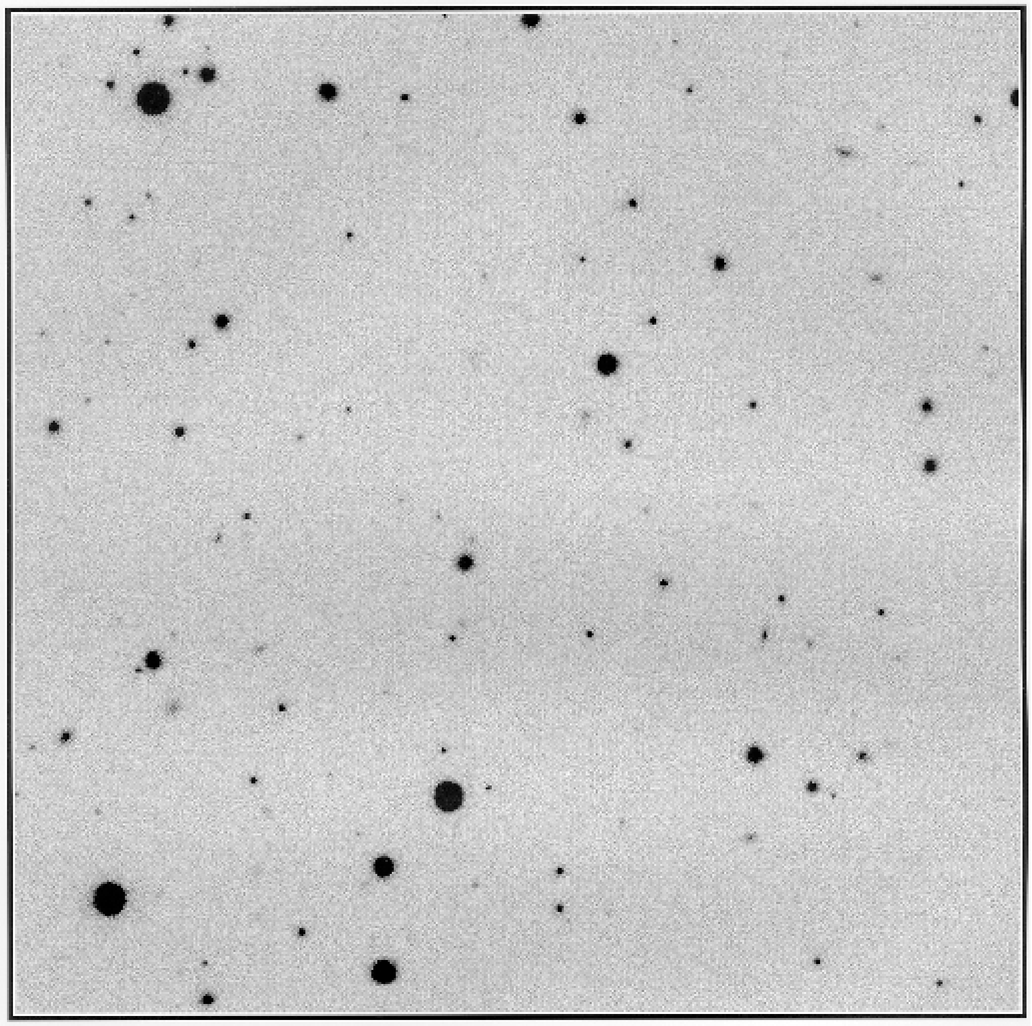
iv) Parameters

At present, we create images using the following parameters. The image scale is 0.400 arcsec/pixel. Readout noise is 5 electrons rms. The AD unit is determined so that 1 DN corresponds to 2, 4, 6, 6, and 6 electrons from u' to z' . All magnitudes are on the AB system (where 0 mag represents a flux-density of 3630 Jy). A 1 Jy flux density gives qfactor x qt dl l electrons, where qfactor = 3.02x 10^9 and qtdll = 0.047, 0.121, 0.120, 0.082, and 0.021 for u' , g' , r' , i' , and z' , respectively. This value of qfactor assumes an exposure time of 55 seconds and secondary obscuration of 25%. For the southern survey, we adopt a typical exposure time of 1650 seconds. The assumed sky brightness (V=21.7mag arcsec^{-2}) based on the sky spectrum at the Palomar (Turnrose, 1974) is 23.1, 22.1, 21.1, 20.3, and 18.9 mag arcsec^{-2} for u' , g' , r' , i' , and z' . Atmospheric extinction is 0.675, 0.223, 0.111, 0.097, and 0.135 from u' to z' at 1.0 airmass; we typically adopt an airmass of 1.2. Image files are written in FITS U16 format.

v) Examples

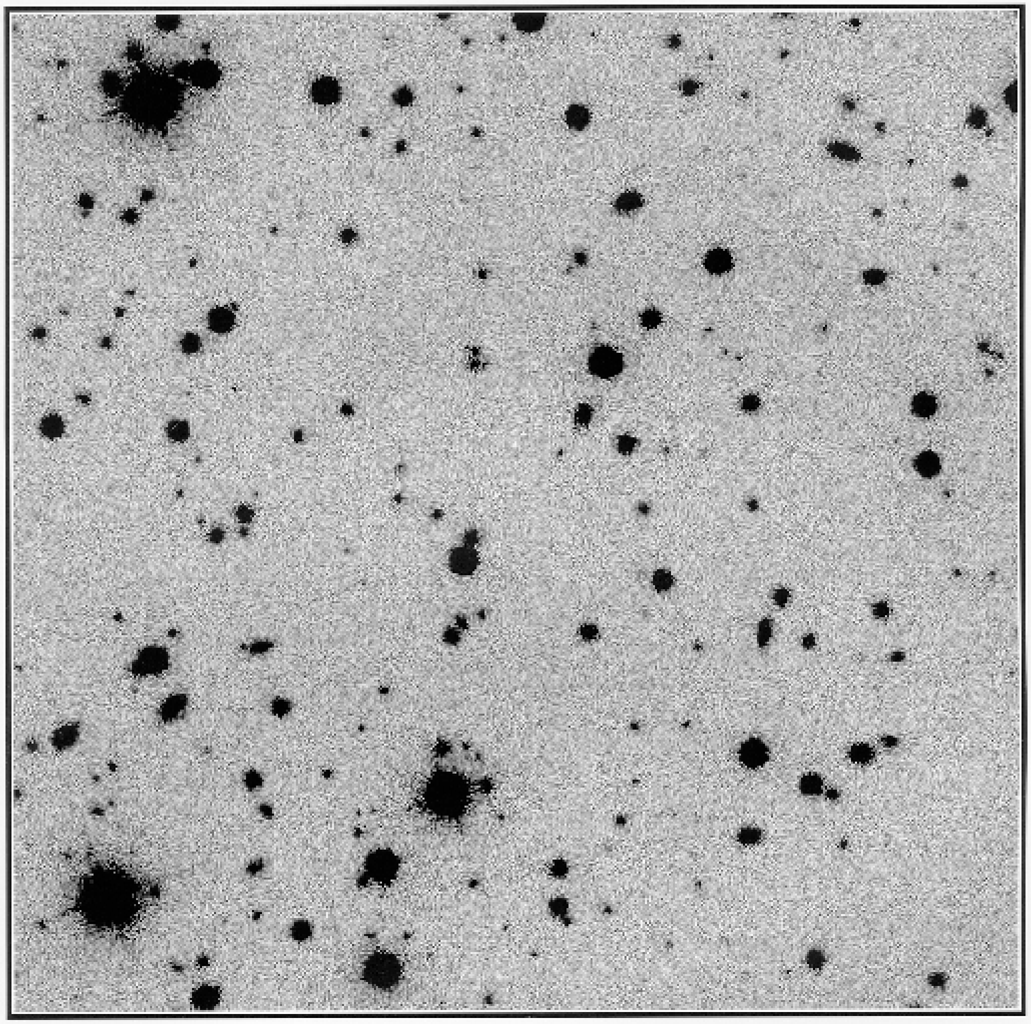
We have also made simulated images of the deeper, southern survey, for which the exposure time is about 30 times that of images in the northern survey. Figure 9.2.2 shows examples of the simulated images for fields from the northern and southern surveys. The simulated field has a star density appropriate to (l,b) = (0°, 40°) . The horizontal axis is the east-west (scanning) direction. Figures 9.2a and 9.2b show portions of the g' images, 512 x 512 pixels, with noise levels corresponding to the northern and southern surveys, respectively. We use the same object catalog for the two images, to allow direct comparison. Figures 9.2c and 9.2d show the corresponding r' images.

Figure 9.2a



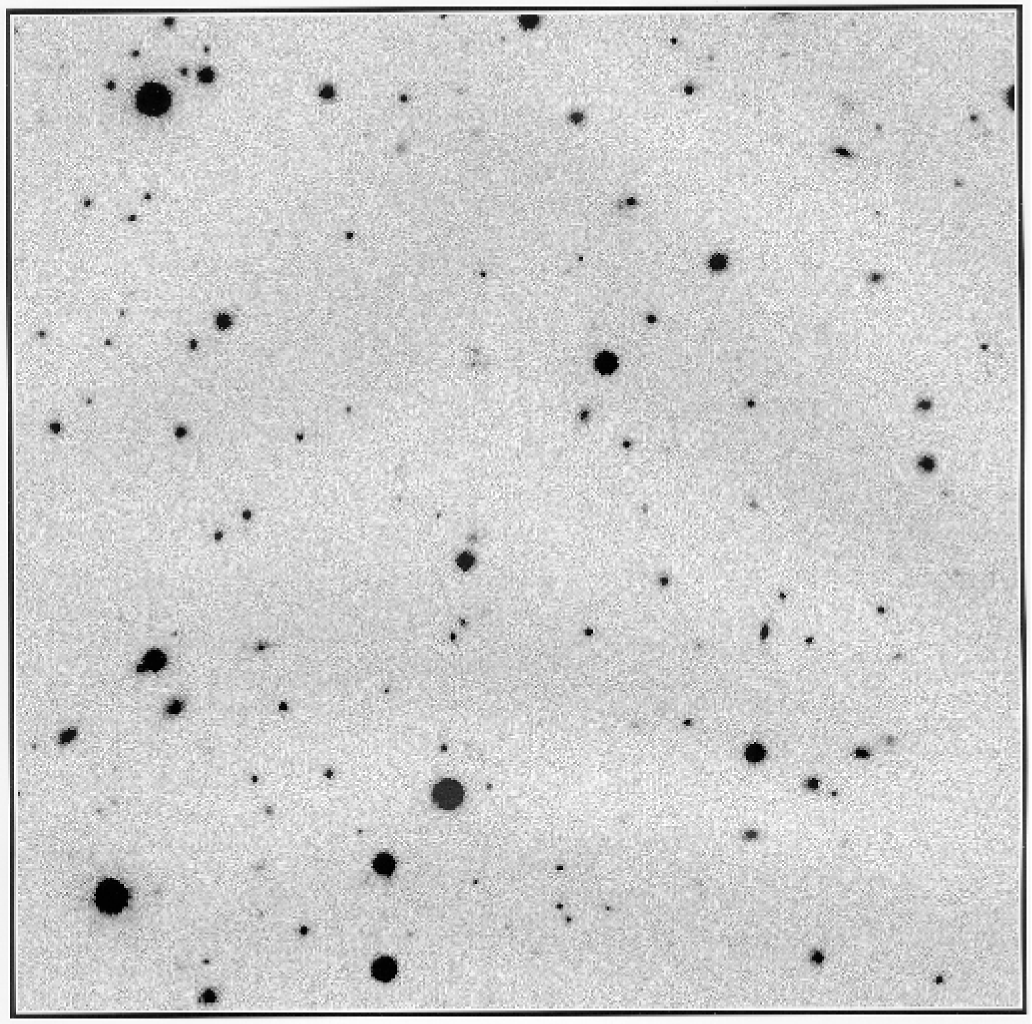
Simulated g' image for the Northern survey. The area shown is 512 x 512 pixels, and the exposure time is 55 sec.

Figure 9.2b



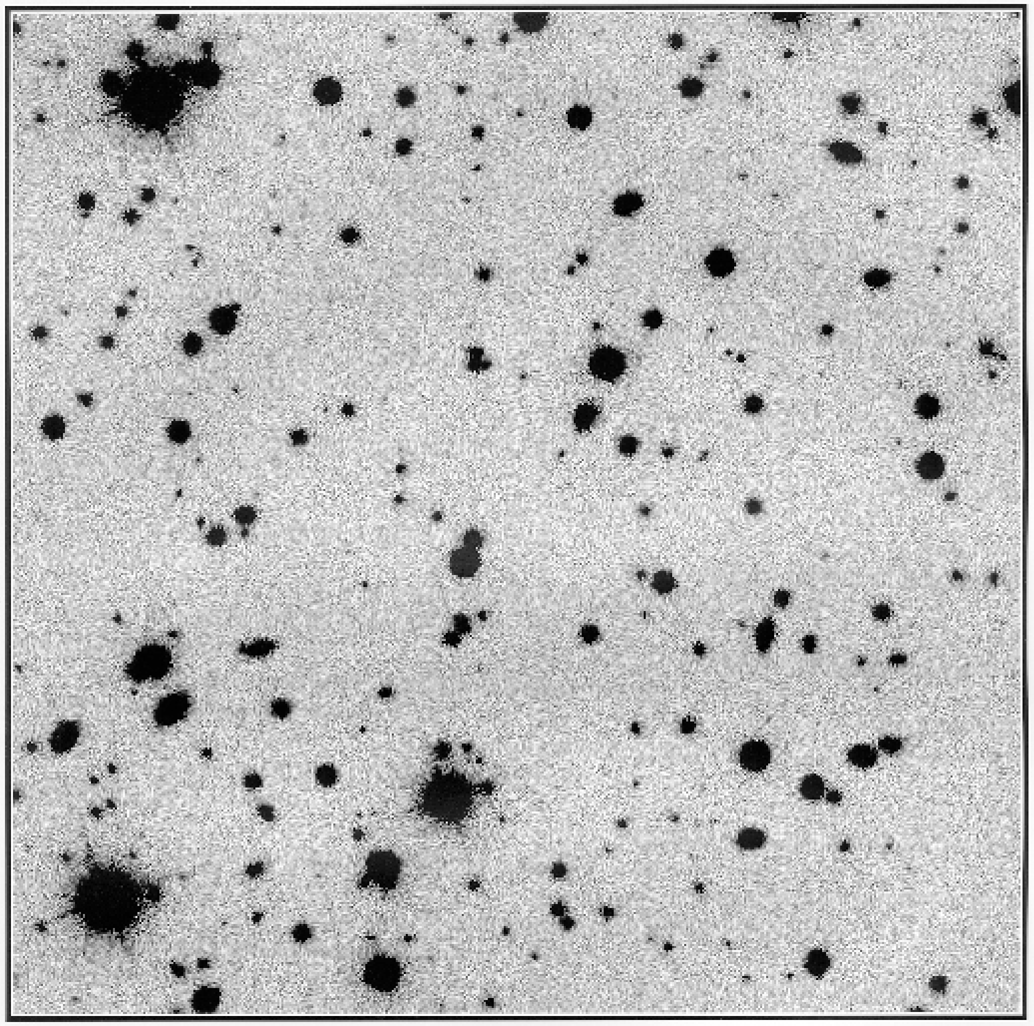
Simulated g' image for the deep Southern survey. The area is 512 x 512 pixels, and the exposure time is 1650 sec.

Figure 9.2c



Simulated r' image for the northern survey.

Figure 9.2d



Simulated r' image for the southern survey.

Test Data

We can test individual components of the data reduction software using simulated images like those described above. However, we also want to test the operation and integration of the full photometric pipeline, and for this we need a complete data stream that looks as if it is coming from the telescope. The third tier of our simulated data system turns a catalog covering a long strip of sky into such a data stream, including observation logs, astrometric and photometric parameters, postage stamps, quartiles, star lists, and raw (un-flattened) image frames (see Chapter 10, Data Systems, for details). We use realistic values for atmospheric extinction, sky brightness, atmospheric refraction, and so forth, and even include real astrometric standard stars from the standard catalogs. We prepare the test data in the actual format of the survey system, on DAT tapes.

We can use these simulated data to ask fairly detailed questions. If the sky brightness slowly increases over the duration of a photometric scan, does the photometric calibration software correct properly? What is the relative performance of the system at low and high Galactic latitudes? While the test year will no doubt bring some software surprises, the use of simulations has allowed us to have the data system integrated and largely debugged before the telescope itself is fully operational. The ability to use the same underlying data with varying degrees of complication will help isolate problems during debugging. The existence of a catalog with the "right" answers corresponding to a given simulation allows us to do regression testing in a detailed and quantitative way.

References

Bahcall, J., and Soneira, R. 1984, ApJSuppl 55, 67.

Coleman, G.D., Wu, C.-C., and Weedman, D.W. 1980, ApJSuppl 43, 393.

de Vaucouleurs, G., 1948, Ann. D'Astrophys. **11**, 247.

Frei, Z., Guhathakurta, P., Gunn, J.E., & Tyson, J.A. 1996, AJ 111, 174.

Gould, A., Bahcall, J.N., Maoz, D., and Yanny, B. 1995, ApJ 441, 200.

Gunn, J.E., and Stryker, L.L. 1983, ApJSuppl 52, 121.

Moffat, A.J.F. 1969, AstrAp 3, 455.

Park, C. 1990, Ph.D. Thesis, Princeton University.

Postman, M., and Geller, M.J. 1984, ApJ 281, 95.

Racine, R. 1996, PASP 108, 699.

Turnrose, B.E., 1974, PASP 86, 545.

van der Kruit, P.C., and Searle, L., 1981, AstrAp 95, 105.

The SDSS Data System and Data Products

Introduction and Overview

The SDSS will collect, process and distribute vast amounts of data. Indeed, the size of the data handling task is such that we could not reasonably have contemplated doing this survey even a few years ago with the compute power available then, even had all the other necessary technical developments (detectors, optics etc.) been in place. The task is far from trivial today, but it is doable; indeed, as we discuss in this chapter, the software is basically in place to run the survey.

The organization and processing of the SDSS data require techniques similar to those used in large high-energy particle physics experiments, and the entire data processing activity is managed by scientists in the Experimental Astrophysics and On-line Systems groups at Fermilab. Much of the science software is also written by these groups, but some self-contained tasks are being done at the member institutions, as described below, because of the presence there of small groups of people with talent in the appropriate area. These groups report their activities directly to the Fermilab group, change activities via negotiation with this group and maintain compliance with the data model, which contains the definitions of the data outputs and interface data for the Survey. The management activities will be described later in this chapter.

The data flow through the acquisition, analysis and archiving phases is organized according to the following principles:

1. The raw data will be permanently recorded, with some small exceptions to be discussed below.
2. The data will be available in its entirety, in both raw and various reduced forms, to the entire collaboration and, ultimately, to the entire educational, astronomical and public communities.
3. On-line processing (the Data Acquisition, or DA, system) will control the operation of the survey instruments, acquire the data and do enough analysis to monitor the data quality in real time. The processing will also prepare the data for off-line analysis.
4. The time between the acquisition of the photometric data and of the spectroscopic data for the same region of the sky can be as short as one month. In other words, the data reduction, processing, archiving and target selection must keep up with data acquisition.

Table 10.1:

	Spectroscopic	Photometric	Postage	Quartiles	Astrometric Stamps
Bits/pixel	16	16	16	3x 16	16
Number of CCDs	4	30	52	30	24
Number of pixels per CCD	2192x2068	2192x1354	40x29^2		2192x1354
per frame time	18.1 Million	89 Million	1.7 Million		71 Million
Peak data rate per CCD	153 kBy/sec	153 kBy/sec	1.9 kBy/sec	340 By/sec	153 kBy/sec
Average data rate	48 kBy/sec	4.6 MBy/sec	99 kBy/sec	10 kBy/sec	3.7 MBy/sec
Raw data per night(10 hours)	1.7 GBy	170 GBy	3.6 GBy	360 MBy	
Total raw data	360 GBy	12 TBy	254 GBy	25 GBy	

Table 10.1 summarises the data acquisition rates and the total data storage requirements for the Survey. The meaning of some of the terms used in Table 10.1, and the assumptions used to calculate the data rates, are as follows. The data stream from each photometric CCD is cut into `frames', 2048 x 1362 pixels, for ease of handling in subsequent processing. The set of five frames in each of the SDSS filters for a region of the sky is called a `field'. We assume that we are reading from two amplifiers for each of the photometric CCDs, with 20 extended register pixels and 20 overscan pixels read through each amplifier, in addition to the 1024 data pixels, for a total of 1064 per amplifier. The extended register and overscan pixels contain no light signal and are used to establish the electronic zero points and baselines for the system.

The `postage stamps' are 29 x 29 pixel subimages of bright stars cut from the imaging data stream by the DA system, which are analyzed on line to monitor the image size and are also passed to subsequent data reduction. (Historical note: US postage stamps for first class mail cost 29 cents when this system was designed). Quartiles are taken of the data in each column for each CCD and are used for the construction of flat fields.

The quantity of spectroscopic data was calculated by assuming that flat field and wavelength calibration frames accompany each pointing and that the data exposure is split into three parts. This gives a total of about 160 MBy per spectroscopic frame, and we hope to complete one field per hour. We have assumed that two amplifiers are used for each spectroscopic CCD and that they are read by a split serial register as are the imaging CCDs and that, in addition, they have a 20 pixel vertical overscan, used to establish the electronic zero points and baselines.

We will not keep all of the data from the astrometric CCDs (which have a total data rate about 2/3 that of the photometric CCDs). There will be about 500 stars per square degree averaged over the survey region which will be bright enough for astrometry, and we will keep centroid and shape information, plus a 29 x 29 postage stamp for each star in each of the 22 astrometric CCDs separately. The `total data' entries include all of the overlaps, both between CCDs in a stripe and between stripes. We will keep postage stamp data from the focus CCDs just as we do from the astrometric CCDs for quality control.

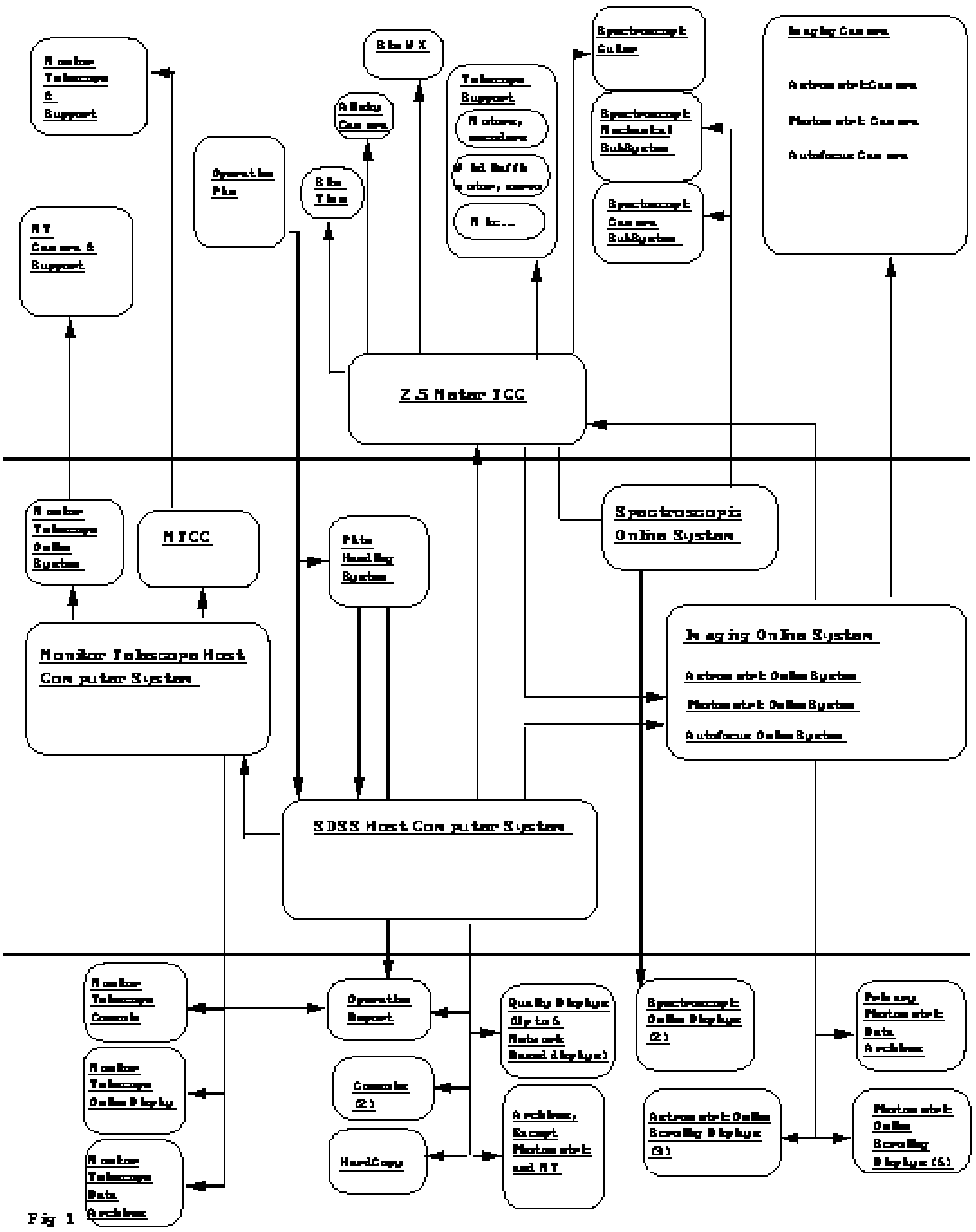
The total amount of data to be gathered and archived by the SDSS is huge, and is dominated by the imaging data. The quantities listed in Table 10.1 do not take data compression into account. The catalogues will be stored in efficient format in the data base, and the images and postage stamps compressed. Gains of a factor between 2 and 3 are likely in storage efficiency. Further, as we discuss later in this chapter, we may implement the construction of highly compressed data cubes which can encapsulate a description of certain fundamental aspects of the survey data (e.g. the galaxy distribution for the entire survey) in much more compact form for subsequent analysis.

Our software falls into two main categories, data acquisition and data processing.

Data Acquisition Systems

Figure 10.1 gives a simplified view of the Data Acquisition (DA) systems and how they interact with the instruments and the follow-on processing. We will have four major instruments to deal with at APO: the photometric array (30 CCDs), the astrometric/focus array (24 CCDs), the spectrographs (4 CCDs), and the monitor telescope (1 CCD). Each instrument has its own VME-based realtime control system with a backend UNIX workstation. The control systems have similar architecture and share much software, although each is tailored for its specific application. The photometric, astrometric, and spectroscopic systems share an SGI Crimson host computer; the monitor telescope system has its own SGI Personal Iris computer.

Figure 10.1



Top-Level Survey Operations

The Imaging Arrays

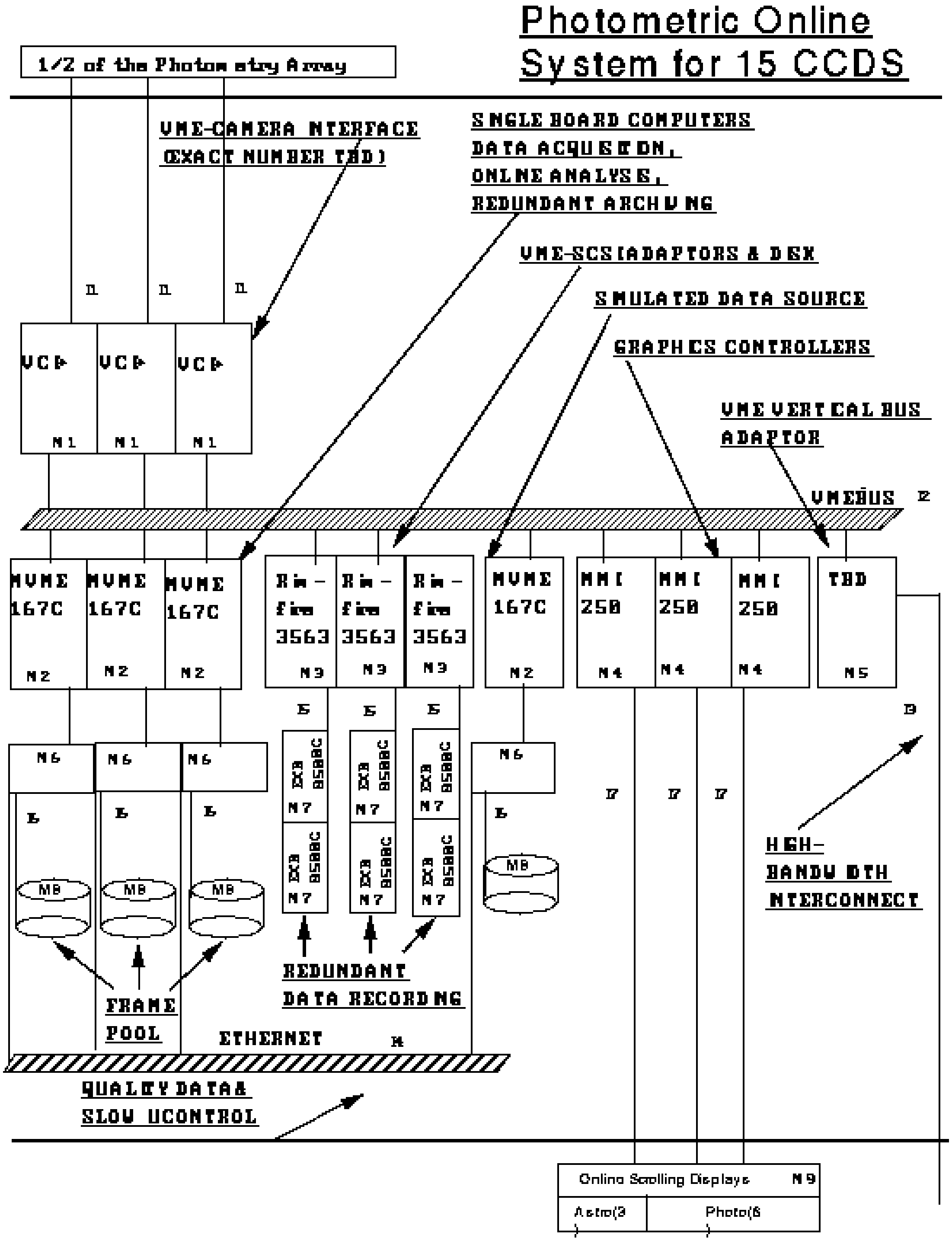
Data for a particular location in the sky will come from one column of CCDs (we define columns to be parallel to the scan direction, just as the columns of the chips are). The focal plane showing the photometric and astrometric chips is illustrated in Figure 4.2. Each column of photometric CCDs is associated with two astrometric chips, a leading one and a trailing one which scan the same region of sky as the photometric column, the leading immediately before the first chip and the trailing immediately after the last chip of the photometric column. Between these astrometric devices are the astrometric bridge CCDs, of which there are also leading and trailing sets and which establish the astrometric link between columns. For the purposes of data acquisition and organization, it is convenient to divide the CCDs along slightly different lines; we will treat the CCD array as ten sets of CCDs. Each of six photometric sets contains the five CCDs from a column. The four astrometric sets consist of the four rows of astrometric chips. In each astrometric dewar the five "bridge" chips and the focus chip are one set and the six chips corresponding to the photometric chips the other. Thus there are six identical photometric sets of five chips, two identical astrometric sets of six chips, and two identical astrometric/focus sets of six chips. Each set has its own acquisition processor and communicates with it over a single optical fiber link.

The map of the sky will be assembled from data written in blocks of runs where each run corresponds to, say, 1 hour (although the length of a given run is arbitrary and will depend on observing conditions). A run covers a portion of a strip, and two interlaced strips are used to form a filled stripe. For planning purposes, we assume that there will be 45 stripes and 540 runs.

The Photometric Array

Figure 10.2 shows a diagram of the DA system that operates 3 columns of the photometric array (there will be two such systems for the photometric array). Data arrives from the telescope over a fiber optic link and is captured by a custom-built board (the VCI+). The data are then spooled onto disk in a `frame pool' that can store 45 minutes of data. This disk holding area is used to buffer data to cover tape drive failures. A scrolling display will show a portion of the data from one or more CCDs in realtime for diagnostic viewing. An MVME167C single board computer (based on a Motorola 68040 processor) handles one column of 5 CCDs. This computer will perform online processing as described below and will write the data to tape.

Figure 10.2



Online system for 15 photometric CCDs

We have chosen to use DLT 2000 tape drives to record the data. These drives combine high bandwidth with reasonable cost. A single drive can sustain a rate of 1 MBy/second, and a single tape can hold 10 GBy. The drives implement hardware LZW compression. We can record the data from a single photometric set of CCDs onto one tape, so we will run 6 DLT 2000s in parallel. Under dark sky conditions, the data will be compressible by at least a factor 2 using this algorithm, so the effective rate is ~ 2 Megabyte/second, more than adequate for the task. Because tape drives generically have reliability problems, we have chosen to record in parallel onto a second set of 6 drives; this will also provide us with a backup copy of the raw data. To ease the offline processing, the data from a single CCD are divided into frames of 1362 rows each (one half the separation between CCD centers) and recorded in a staggered fashion such that frames from the 5 filters for a given piece of the sky are written sequentially on tape. A set of tapes can record approximately 7 hours continuous imaging. The frame pool provides enough buffering to permit changing tapes without halting observing.

The processing of the data is split between the online system and the offline processing. The split is done for two reasons. First, if one simply recorded the data with no real-time processing, it would be impossible to perform even simple quality control procedures (What is the seeing? Are all the sensors actually working?). Second, there is information in the stream that is time-variable, (the forms of the PSFs due to seeing changes, focus and tracking imperfections, the sky brightness and color, etc) which may or may not be accurately determinable from a single frame, and for which in any case it is desirable to enforce continuity in the data stream. These quantities need to be extracted in a global way; part of this is done in the online processing and part offline, as will be described below.

The online system performs three steps of processing. First, the median, first, and third quartile points are computed for each column of pixels in a CCD frame (1362 rows). These quartile arrays are saved to disk. Second, a simple object finder identifies all bright stars with intensities that fall above some preset threshold, extracts 29-pixel square (12 arcsecond) postage stamps around each one, and computes shape parameters. These too are written to disk. In the subsequent processing, the quartiles are used to find the sky intensity and thus the flatfield vectors, and the stars are used for photometric, astrometric, and point spread function calculations. The data are also available to the observers and provide a convenient monitor of the sky intensity and seeing.

A full night of quartile vectors will occupy about 360 MBy of disk, and the postage stamps another 3+ GBy. One can now buy 9 GBy disks cheaply, so the volume of data can be handled without much trouble.

Data which are crucial to the calibration of the imaging data will be included in the data stream. These include the time; telescope pointing information; information about the health of the camera, including the chip temperatures and the readout noise derived from the extended-register pixels; and information from the Monitor Telescope regarding the seeing and transparency during the scan. Summary versions of these data will be recorded in a separate data base to facilitate the determination of long-term trends in the observing conditions and state of the instrument.

The Astrometric Array

The DA system for the astrometric CCDs is nearly a copy of that for the photometric CCDs. However, only the postage stamps and parameters for the detected stars will be written to tape, not the raw pixel data from the astrometric chips. Over the survey region there are an average of about 500 stars per square degree bright enough for the astrometric chips to record to interesting accuracy, and these are the only ones in which we are interested for this purpose. For each of the stars we will record an accurate pixel position, shape parameters, a flux, and (just to cover ourselves) a 29 x 29 pixel (12 arcsecond) postage stamp of the star. By saving the postage stamps, we have the option to go back and apply flatfields and derive more accurate centroids using more complex PSF fitting at a later time, but we may well find that the improvement in centroiding is negligible. Each chip covers 9.4×10^{-4} square degrees per second, and so encounters about 0.5 of these stars per second; the whole array of 22 chips encounters about 11 per second. The DA will preferentially skip fainter stars if the rate for any CCD exceeds about 1 per second, given the CPU limitations of the DA computers, but that does not have any impact on the subsequent processing.

The Spectrographs

The data rates for the spectroscopic survey are much more modest, and consequently the DA system is much simpler than that for the imaging system. We will use the same basic architecture for instrument readout (a VME crate with an MVME167 control computer), but the raw pixel data will be sent straight to the Unix host computer for storage.

We have two double spectrographs, so the 640 spectra are recorded on four 2048 x 2048 CCDs. After an exposure we read out a total of 36.3 MBy (4 CCDs worth). The readout time in 2 amplifier mode is 59 seconds.

The spectroscopic exposures will probably be split into three parts to allow cosmic ray removal, and will be accompanied by calibration data, probably one flat field and one wavelength calibration (if the night sky lines plus occasional spectral lamp exposures do not suffice), which multiplies the total data and average rate by 5. There will probably also be some highly binned spectroscopic frames (taken with the telescope slightly offset to discover exactly where the fibers fall on the galaxies -- cf. Chapter 7.8), which add negligibly to the total amount of data. Thus each spectroscopic frame and associated data amount to about 180 MBy and simultaneous access to more than one frame is not necessary. The total amount of spectroscopic data recorded depends on the number of targets. If we assume that each spectroscopic field looks at 5 square degrees of sky (the full 3 degree round field is seven square degrees, but the inscribed hexagon is 6.2 square degrees and there must be some overlap to allow for adaptive tiling), there will be about 2000 spectroscopic fields, 250 eight-hour plus (for overhead) nights, and a total database of about 360 GBy. These data will compress by a factor of two to three, although compression seems hardly necessary.

Data that are crucial for the calibration of the spectra and their correlation to the photometric survey will be written along with the spectra. These include the time, telescope pointing information, and fiber placement measurements, as well as any engineering data which are relevant, such as chip temperatures and read noise.

The Monitor Telescope

The monitor telescope DA system is essentially a smaller version of the spectroscopic system, since there is only 1 CCD to be controlled. The monitor telescope will operate in a semi-automated fashion. The default observing program will be to observe a sequence of primary standard stars to monitor atmospheric extinction. While the 2.5 meter is operating, the monitor telescope will receive positions of secondary star fields to be observed as well. (These can be done any time, so long as they are available for reduction of the photometric data; and it would be advantageous to get ahead of the imaging survey.) We will be able to observe roughly 6 primary standard stars and 3 secondary fields per hour. For the primary standards, we will record only the central 1024x 1024 of each frame. The maximum amount of data that could be collected in one night is about 2 GBy. Again, these data will be stored directly to disk.

The primary standard star frames will be processed in real time in order to extract instrumental magnitudes. Calculation of a photometric solution roughly once an hour, and comparison of the measured and expected counts for each star, will allow us to determine the time period when the night is photometric; this information will be relayed to the 2.5 m observer in real time to facilitate planning of the night's operations.

Online Processing

The photometric, astrometric and spectroscopic DAs will share a backend UNIX host workstation (an SGI Crimson). This workstation will provide the operator interface to the DA system, provide compute power for any `quick look' processing of images that is not already being done in the VME computers, and provide displays of critical information on the system performance. Most of the key information that one might want can be extracted from the star lists and quartile arrays, and so the monitoring task is simply to format this information for the observer. The realtime analysis of the camera data is performed by a collection of routines called `Astroline'.

Control Software

All of the control and analysis software is written using the TCL (Tool Command Language) based software framework that is described in Section 10.3.2.7. A common core DA system is provided for all the instruments, with the code split between the VME front end and the Unix back end computers. The common core DA can receive data from multiple CCDs, perform the online analysis tasks, spool data to disk and tape, run the scrolling displays, and pass information between the VME and Unix processes. Each major instrument has customized configuration files and TCL-based observing programs. The TCL language provides features such as multi-tasking, foreground/background process control, interprocess communications using TCP/IP, file I/O, X window GUI construction tools, and extensive interfaces to the Unix operating system. The observing program is broken into several processes that are run as independent tasks. For the imaging system, for example, the following processes are provided:

1. User interface - used to initiate tasks in other processes and to interactively view and perform interactive analysis tasks.
2. Telescope control - provides interface to the telescope control computer.
3. Camera control - sends and receives control and status information to the camera controller.
4. Executive control - execute observing programs that drive the other processes in a coordinated fashion (example - run a focus sequence).
5. QA control - this process fetches the postage stamp and quartile files from VME to Unix and formats the information for display on the observer's console.

For the spectroscopic observations, we will use IRAF to perform any online analysis (e.g. quick flatfielding, extraction, wavelength solutions) as required.

Data Processing

Overview

The bulk of the data processing will be done offline at Fermilab. Tapes from a night's observations will be shipped via Federal Express (which picks up at the Observatory). We feel that this approach is preferable to the alternative of doing the processing on the mountain, the reasons including the difficulty of maintaining the fairly large computing system required for the reductions and the necessity in any case of getting the data to the central archive. The effective baud rate of a box of DLT tapes carried by Federal Express is much higher and the transfer much more reliable than that afforded by current network protocols.

There are several constraints on the data processing that must be met. The imaging observations must be reduced in a timely fashion in order to identify spectroscopic targets. Ideally we would like to turn around the image processing on the time scale of a few days, but there are enough steps involved that this may not be practical; furthermore, a given spectroscopic plate will require of order 6 separate stripes to be combined in order to generate target lists if we are to take advantage of the benefits of adaptive tiling (see Chapter 8). Practically, we will aim to have all imaging data from a given dark run reduced in time to have spectroscopic targets available for the next dark run -- this gives us a required turnaround time of about a week. Given the large amount of data that must be processed, and the desire for uniformity, the algorithms must be sufficiently robust that minimal human intervention is needed. Finally, rapid turnaround is needed to verify the quality of the data and allow re-observation of a given field if necessary. Thus we will have accomplished our most fundamental goal the code implemented at the beginning of the survey correctly finds and classifies all spectroscopic targets. If, furthermore, we find all the objects present in the data to the desired significance level and extract large enough subimages on the first pass, we need never return to the full pixel data set for reprocessing, and can do improved processing on the much smaller subimage data set (cf., the discussion below in Section 10.3.5.10). However, if we find it necessary, we *will* be able to reprocess the imaging data (though we will go to great lengths to avoid having to do so), as long as it does not affect the spectroscopic selection.

Figure 10.3 gives an overview of the complete data processing system. Each tape of raw data from the mountain is fed into a "pipeline" that performs routine processing and produces some output, usually of much reduced size. The astrometric, monitor telescope and photometric pipelines are very interdependent. The astrometric and monitor telescope pipelines are fairly straightforward to design and implement, but the pipelines that process data from the photometric CCDs present the greatest problems because one needs calibration data available in order to process the data, but final calibrations, especially quantities like the PSF parameters and the flat field as a function of time, are not local quantities, and are thus not available until well into the pipeline processing. We have addressed this problem by breaking the processing into two stages: the `postage stamp' pipeline and the `frames' pipeline. The postage stamp pipeline takes the postage stamps of bright stars and quartiles that were extracted from the online system and the upstream Serial Stamp Collecting Pipeline and computes the flat-field vectors, point-spread functions, and gets Monitor Telescope calibration lists and an astrometric solution, sufficient for the frames pipeline to function in such a fashion that individual frames can be processed independently. The output contains sufficient information about each object that it is possible to derive and apply slightly revised calibrations after the fact. The six columns of CCDs are independent of one another for the purposes of data processing, so our pipelines are designed to process one column of CCDs at a time.

The Management and Organization of the Science Software Pipelines

Since the software is designed for automatic reduction, cataloguing and archiving of the vast SDSS data stream, it is largely via the software pipelines, in particular the code which selects the spectroscopic targets, that the scientists at the SDSS institutions carry out the scientific design of the survey.

The SDSS is driven by the requirement to carry out its observations in a uniform, accurate, well-controlled, well-documented and well-understood manner. As the software has evolved, it has proven most effective to place responsibility for each of the major software pipelines in the hands of a small group whose members, as far as possible, are located at the same institution. Almost as important as the development of the software is its testing, and this is organized so that the pipeline development is done by one group and tested by another. The institutional responsibilities are:

1. The *basic framework*, and the common science code, are written at Fermilab and Princeton and are maintained at Fermilab.
2. *Astroline* is written and tested at Fermilab.
3. The *astrometric pipeline* is written by one group at the U.S. Naval Observatory and tested by another group there.
4. The *monitor telescope pipeline* is developed at Princeton and Fermilab and tested by the JPG.
5. The *photometric pipeline* is developed at Princeton and tested by the JPG.
6. The *spectroscopic pipeline* is written at Chicago and tested at Princeton.
7. The *target selection pipeline* is written at Fermilab (with input from scientists from the entire collaboration) and tested by D. Weinberg (Ohio State).
8. The *simulations* for testing and integrating the pipeline code are run, and their outputs stored, at Fermilab. The code was written by scientists at several institutions.
9. The *operational data base*, which is used in the operation of the survey, has been developed, and resides, at Fermilab.
10. The *science data base*, which enables scientific access to and analysis of the survey data, is developed at Johns Hopkins.
11. The *survey strategy* software is written at Chicago and Fermilab.

The simulations are described in Chapter 9 and the science data base architecture in Chapter 11. The survey strategy is discussed in Chapter 1. The data processing pipelines themselves are described in the next section.

Integration

By making use of extensive simulations, we have managed to demonstrate that all of these pipelines are able to interoperate. Specifically, we have run 10°x10° of simulated data through the entire system, up to and including target selection. This is a non-trivial achievement, and we periodically repeat the exercise to confirm that no inconsistencies have appeared.

CVS (Concurrent Versioning System)

With so many people working on the software at so many locations, it was clear that some sort of software management tool would be essential. We have chosen to use the well known free product CVS, which has proved to serve our needs well. The code repository is resident at Fermilab. Specifically, CVS allows:

- Multiple developers working on the same pipelines, with suitable control over conflicts.
- The ability to recover the state of the code at any time, and to inspect logs giving the reasons for changes.
- Transparent access at remote sites. For example, the Princeton PHOTO group can tell the JPG in Tokyo "Get the latest version of PHOTO, and get our latest version of JPGtest" and the JPG can say "We found a bug in your code, get it from CVS."

The development of the software pipelines is designed to proceed in four stages, as follows.

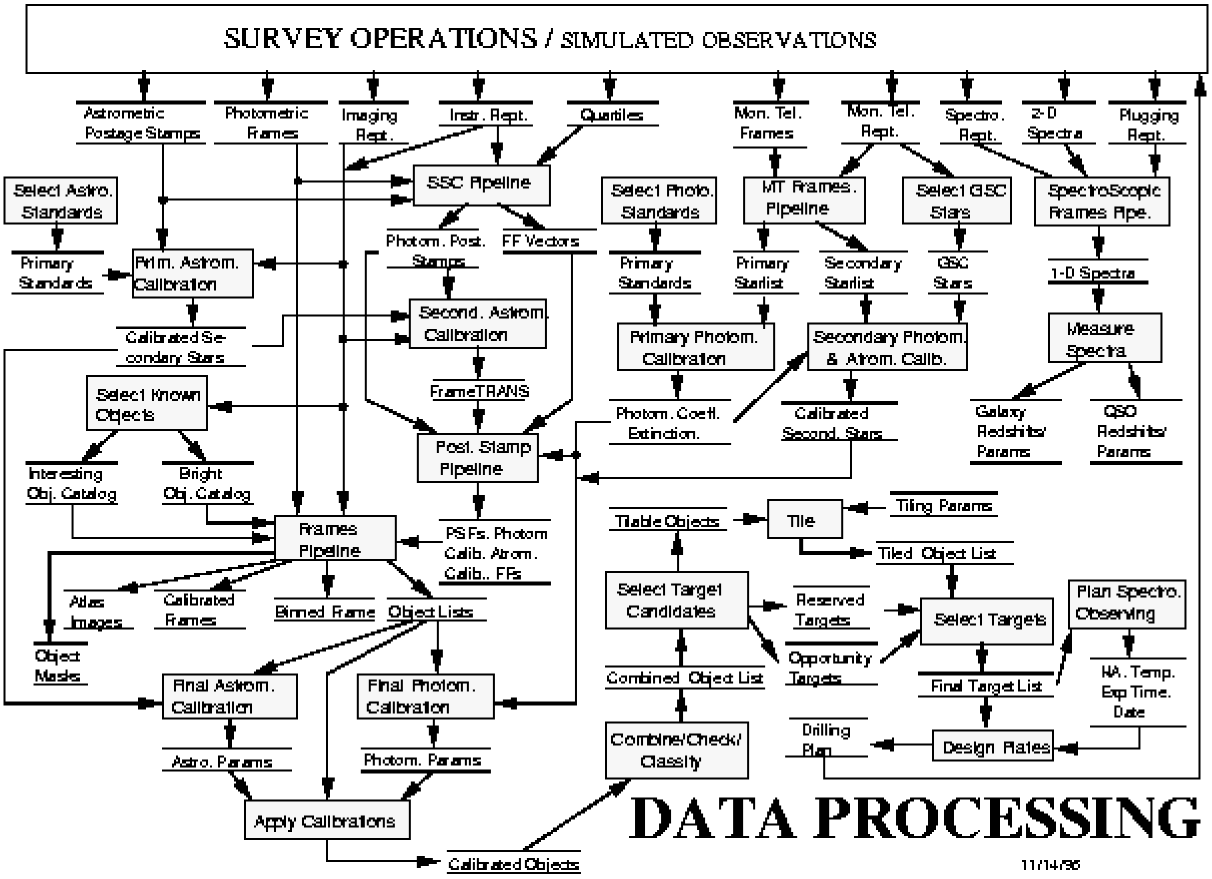
Prototype (completed January 1993)

The prototyping of the pipelines is, basically, a proof-of-concept exercise, and was carried out for the photometric, astrometric and spectroscopic pipelines only. These pipelines were written in outline form to analyze a small set of data which are not fully self-consistent and are far from reflecting the true system complexity (e.g. this software dealt with simulated data from only five photometric CCDs and one astrometric CCD; further, there was no requirement that the outputs be scientifically valid). The data base was used at this stage only in the development of the spectroscopic pipeline.

Level 0 (completed November 1994)

Level 0 is a test data processing system that is designed to ensure that the data procession framework is working correctly. The pipelines completed to Level 0 on the above date were the photometric, astrometric, spectroscopic and monitor telescope pipelines. They contained most of the necessary functionality and operated on a set of self-consistent test data. The imaging data pipelines were fully integrated, and the outputs were scientifically meaningful.

Figure 10.3



Data flow through the imaging and spectroscopic pipelines.

Level 1 (nearing completion at the end of 1996)

This system is as complete as possible without having actual data from the telescope. It is designed to have full functionality, i.e. to be able to run the survey, and operates on a set of simulations which have been designed to be as realistic as possible. The imaging pipelines (monitor telescope, astrometric and photometric) have been fully integrated, run efficiently and exhaustively tested, while the spectroscopic pipeline is ready to be integrated in early 1997. The imaging pipelines produce meaningful scientific output which has been successfully used to run the target selection pipelines and science and operational data bases.

Level 2

This is the software which will carry out the data reduction for the entire survey. During the test year (Section 1.3), the algorithms will be optimized for real data. Level 2 will be "frozen" at the beginning of survey operations proper.

Framework

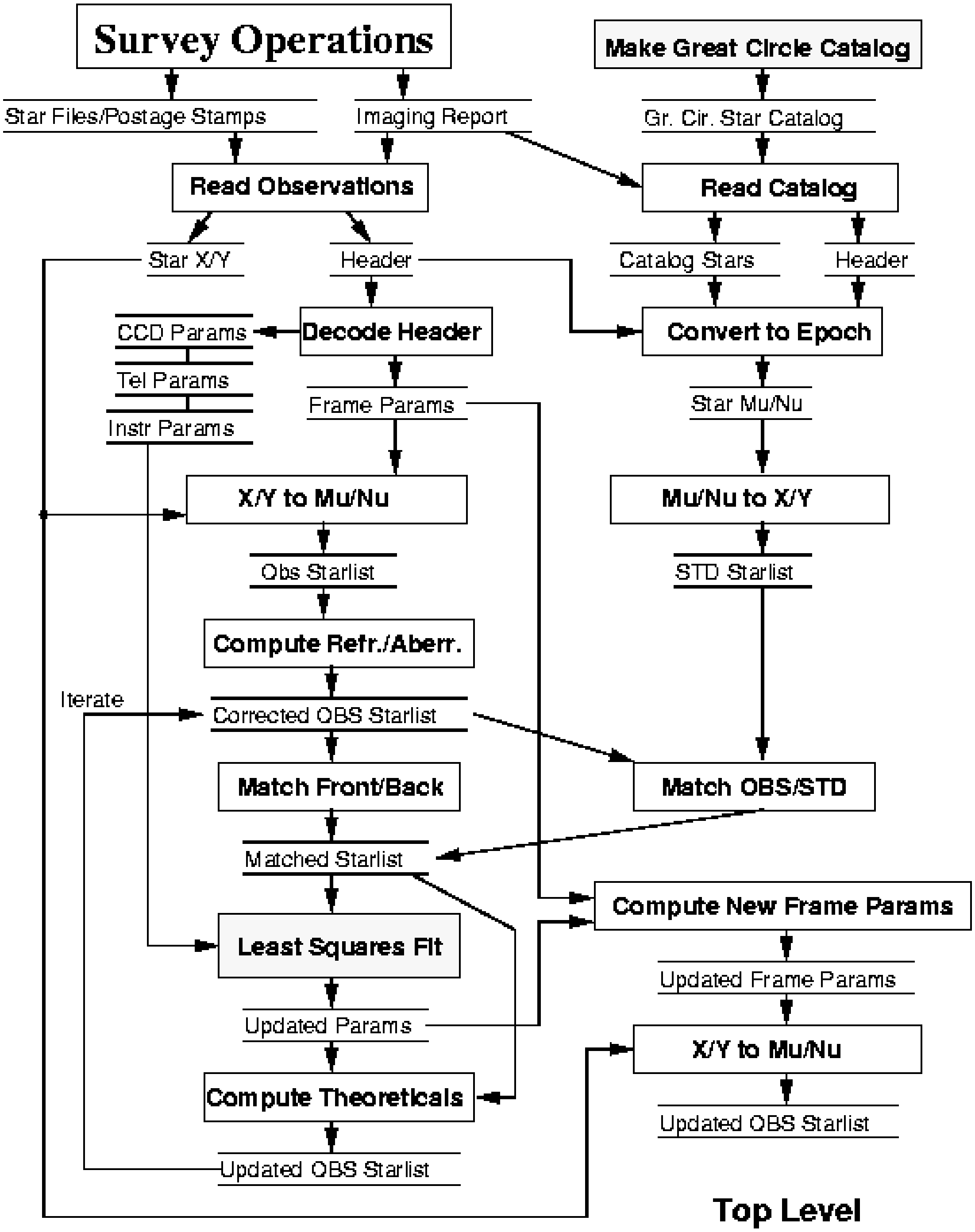
As will be seen in the discussion of the individual pipelines below, the functionality is usually provided by several sub-pipelines. The pipelines are coded in ANSI-C and TCL, and are built on a basic framework developed at Fermilab called DERVISH/SHIVA (the latter was the name until December 1996, hence the nomenclature in some of the figures in this chapter - it's a *long* story). The Astrometric (ASTROM), Monitor

Telescope (MTpipe) and Photometric (PHOTO) pipelines are integrated under DERVISH/SHIVA and run together to reduce a night's worth of photometric data - see Figure 10.3. The task of maintaining this operating environment and of integrating the pipelines and maintaining the integration is a major one. A very large amount of code is involved in ensuring that the pipelines talk to each other and to the data base, in monitoring the processing tasks and ensuring that the outputs from the processing are available when needed. Furthermore, the software must be embedded in a framework which provides image displays, command interpreters and so forth.

The Astrometric Pipeline

The online system detects bright stars above some preset threshold and saves both a postage stamp and the image centroid and shape parameters. No flatfielding is done. The present astrometric pipeline makes use only of the image parameters; processing of the postage stamps will be added later if deemed necessary. The centroiding algorithm makes use of the fact that the image of a point source is roughly Gaussian. The data are smoothed and interpolated using the standard seeing model approximation of a two component concentric Gaussian, the outer component having 1/10 the amplitude and about twice the width of the inner. Information useful for astrometry is contained entirely within the inner component. The centroid is then found by fitting a corrected polynomial expansion of the central Gaussian to the marginal x and y distributions of the data. The accuracy of this computation is < 3 mas due to systematic effects and ~ 20 mas due to photon noise at 20th magnitude. This is much better than the errors expected from classical seeing theory (Section 6.4.2) of about 30-40 mas. The identical centroiding algorithm is used in the Photometric pipeline (see below) to ensure that the astrometric information is properly transferred to the photometric data.

Figure 10.4



Architecture of the Astrometric Pipeline

Figure 10.4 shows a diagram of the astrometric pipeline. The major steps in the process are as follows:

1. Read output of online processing -- at present these are in ASCII format, one file per CCD per run. There are three instrumental quantities: x coordinate, y coordinate and counts.
2. Extract instrumental parameters (cf. Chapter 6 -- default values for all parameters are stored in the frame headers, but one can optionally insert new values if the headers are in error.)

3. Make Great Circle Catalog -- This step involves extracting all astrometric standard stars from online catalogs.
4. Read Catalog.
5. Convert to epoch -- This step applies proper motion corrections to the epoch of observation.
6. Find the transformations between CCD coordinates (X,Y) and great circle coordinates (μ, nu) -- For generality, observations and catalog stars are treated internally on an equal footing. Nominal positions for each star are computed based on the known scale of the telescope and the nominal pointing position as stored in the header.
7. Compute refraction/aberration correction -- These are the differential refraction and aberration corrections to the position of a star relative to the telescope boresight (see Chapter 6 for a description of the boresight).
8. Match front/back -- Match up duplicate detections of stars that cross both the front and back arrays.
9. Match obs/std -- Match up observed and standard stars.
10. Least Squares fit -- Solve for transformation coefficients for the astrometric solution.
11. Update parameters and iterate as necessary to find the new self-consistent astrometric solution.
12. Compute New Frame Params -- The transformation from CCD X and Y coordinates to sky coordinates is nonlinear and messy. Rather than store a large number of coefficients, we will store a set of linear coefficients, one set per CCD frame, that approximates the true transformation. Maximum deviations from the linear transformations should be of the order of 5 μ, about 80 mas. This is not good enough for the ultimate astrometric accuracy we hope to achieve, but the higher-order terms, which result mostly from uncompensated distortion introduced by the field flatteners, should be stable enough that they do not need to be recalculated often, and can be considered as fixed parameters in the transformations.
13. X,Y to μ, nu -- Compute improved coordinates for all stars.

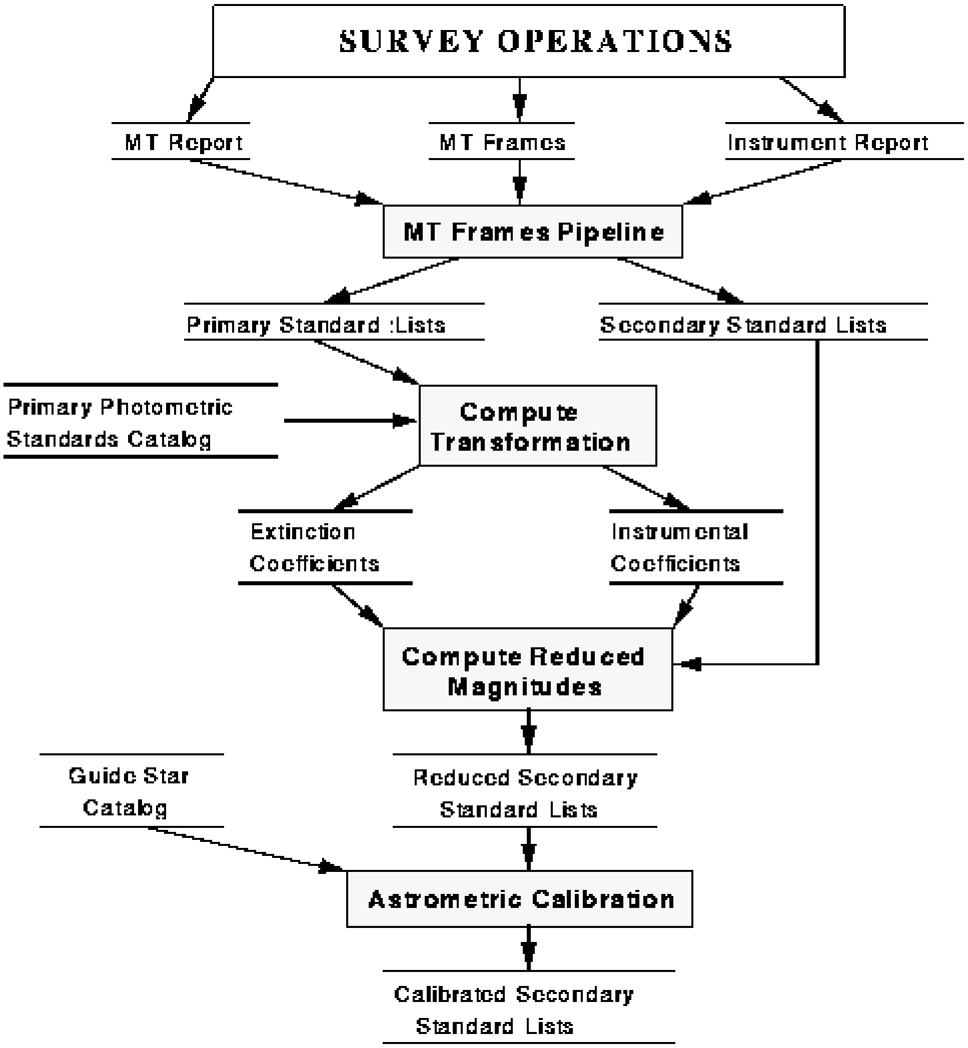
A related task is monitor and control of the focus, using the dedicated focus CCDs. The focus system is discussed in some detail in Chapter 4; here it need only be noted that only the fitting parameters for the focus images, which are like the astrometric images, need to be stored, though we may store postage stamps as well to allow analysis of the performance of the focusing system. We will update the focus more-or-less continuously, and so the computation of parameters will be done in real time.

The retained output of the astrometric pipeline therefore is a calibrated position for every observed star, a set of 6 coefficients for every photometric frame in the run. There will be in addition a detailed record of the reduction process and a large set of parameters from the least squares fit that will be archived, although they will not be used further.

The Monitor Telescope Pipeline

The purpose of the Monitor Telescope Pipeline (MTpipe, or MT for short) is: (1) to measure the extinction at the site as a function of time and (2) to calibrate the secondary standard stars across the sky (see Chapter 5). These secondary standards are thereby tied to a small network of primary standard stars (Fukugita et al. 1996). This is done using semi-autonomous observations of standard stars and secondary standard calibration fields (see Chapters 1 and 5). The data flow through MT is shown in Figure 10.5. MT automatically reduces these observations and makes the measurements which the photometric pipeline needs to calibrate the imaging data.

Figure 10.5



Top Level

MONITOR TELESCOPE PIPELINE

10/10/93

Data flow through the Monitor Telescope Pipeline.

MT consists of three sub-pipelines: MTframes, Excal and Kali, which are invoked in this order. MTframes, analogous to the Frames routine in the Photometric pipeline, does the bulk of the reduction of the raw MT data on a frame-by-frame basis. It bias-subtracts and flat-fields the data, searches for objects, measures them and outputs the results.

Excal is a TCL procedure which performs the photometric solution by a least-squares fit to the output of MTframes. To do this it must identify the standards automatically from the frames. This routine works on the primary standards, and outputs extinction measures for the Postage Stamp Pipeline (part of the photometric pipeline; cf. Section 10.7) to work with.

Kali is a TCL procedure which applies both a rough astrometric solution and the photometric solution to the secondary standards to run as input to the Postage Stamp Pipeline.

The Photometric Pipeline

The photometric pipeline (PHOTO) is required to carry out the following tasks:

Correct the Data: Flat field, interpolate over bad columns, and remove cosmic rays

Find Objects: Sky level measurement, noise measurement, sky subtraction, object finding

Combine Objects: put together the data from the five bands for each object.

Measure Objects: position, counts, sizes, shapes

Classify Objects: provide parameters for object classification, i.e. goodness-of-fit parameters from fits to simple models (point source etc.)

Deblend Objects: do simple model fits to overlapping images.

Output Results: Write out an object catalogue plus images and corrected data.

The pipeline operates on a frame by frame basis. The photometric data stream from each CCD in the photometric array is cut into frames of 2048 x 1362 pixels. Frames are then re-assembled by adding to the top of each frame the 128 rows from the next frame, so that the frames before processing are 2048 x 1490 pixels with 128 pixels overlap with the next frame. The five frames for each field are then combined. (Note that the individual bands are observed sequentially.) This is the same number of pixels in the side-to-side overlap when the two strips of a stripe are observed. Each set of five frames is then processed. However, as we mentioned above, the processing needs information for the entire run: quantities such as the point spread function, the sky brightness, and the flat field and bias vectors are not local quantities, but are determined as a function of time, smoothed where appropriate, and interpolated to each frame. This requires at least two passes through the data, and thus PHOTO consists of three pipelines. In order of execution these are:

Serial Stamp Collecting Pipeline (SSC): cuts out postage stamps (currently 65 x 65 pixels) from the photometric data stream for input to the PSP, and orders these files in an appropriate format.

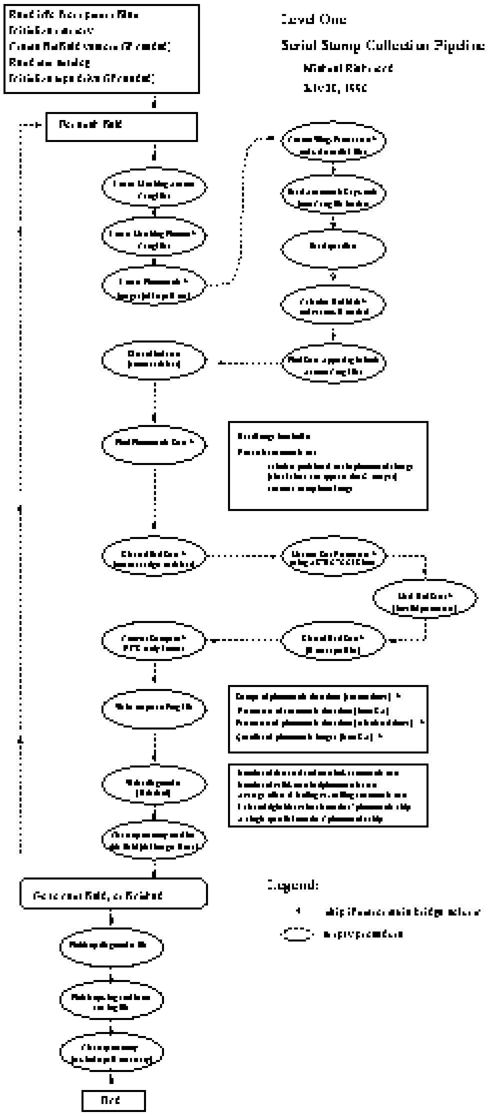
Postage Stamp Pipeline (PSP): analyzes these postage stamps, and characterizes the point spread function from these images. It calculates the bias, sky and flat field vectors for each row. It also takes input from the astrometric and monitor telescope pipelines to calculate preliminary astrometric and photometric solutions.

Frames Pipeline: Performs the analysis on the frames one at a time, using calibration information from the PSP.

PHOTO uses this architecture because the frames pipeline is very compute intensive. The architecture allows the Frames pipeline to run for one set of frames at a time regardless of the ordering of the frames - the calibration and instrumental information for the entire scan is carried by the Postage Stamp Pipeline. The compute engine for the production system is a multi-processor DEC Alpha (see below) which can run multiple copies of Frames, and pass a new set of frames to each processor asynchronously as the processing of the previous set finishes. Although the PSP originally used only the postage stamps cut by the DA system, further development showed that if postage stamps were cut from the imaging data for the stars detected by the astrometric CCDs, both the astrometric and photometric solutions are greatly improved, which prompted us to write the SSC. This, however, necessitates two passes through the data. As will be discussed below, we have fast enough machines and code to allow this.

The Serial Stamp Collection Pipeline

Figure 10.6



Flow diagram for the Serial Stamp Collection Pipeline

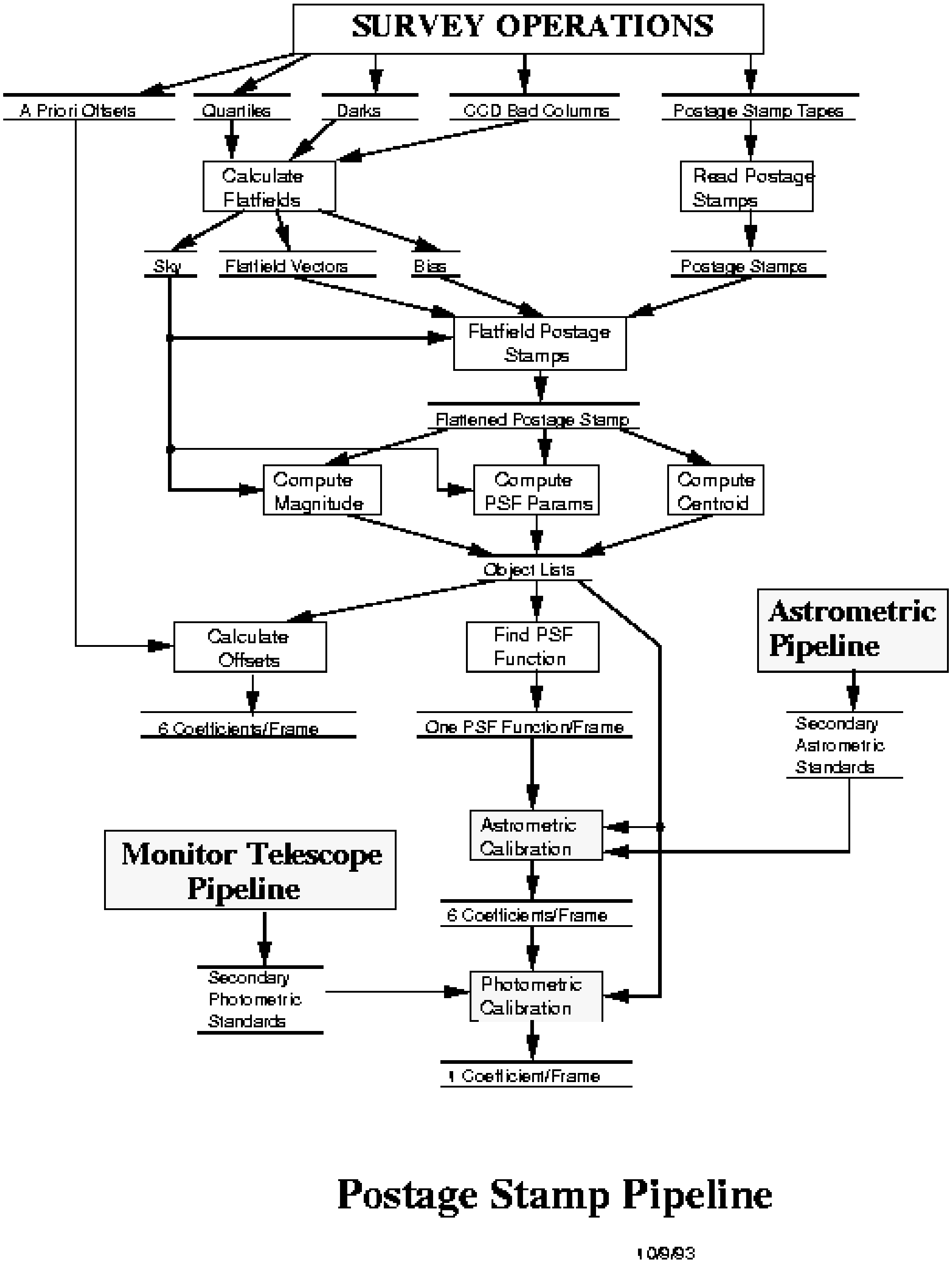
Figure 10.6 shows the architecture and data flow for the Serial Stamp Collection pipeline (hereafter SSC). The first job of the SSC is to cut postage stamps for three categories of star from the photometric data and to pass them to the PSP and Frames pipelines. These are: (1) stars detected by each pair of leading/trailing astrometric chips. The postage stamps are 65 x 65 pixels. (2) one 200 x 200 pixel postage stamp from each frame around a star brighter than 14^{m} (the central part of the image will be saturated). These stars are bright enough that they will show first order ghosts due to internal reflections in the camera optics. (3) an entire frame every hour or so containing a star brighter than 7^{m} ; these stars are bright enough to show ghosts due to secondary reflections in the camera.

Stars in category 1 serve two purposes; they allow the point spread function to be measured and they allow the astrometric solution to be transferred to the photometric data. To this end, the parameters of the unsaturated stars (centroids, widths) are measured using exactly the same code as does the online astrometric DA system. The outputs of the SSC are written to an output file for use by downstream pipelines.

The online DA system reads out each CCD in the imaging camera and groups together data from all of the images read out at one time into an output file. Since it takes about 10 minutes for a star to cross the astrometric and photometric arrays, the images in this file do not correspond to the same area of sky. The second purpose of the SSC is to re-arrange the DA data stream so that all of the data for a given piece of sky are collected into a single output file. The frames in the five different bands for the same part of the sky are called a *field*.

The Postage Stamp Pipeline

Figure 10.7

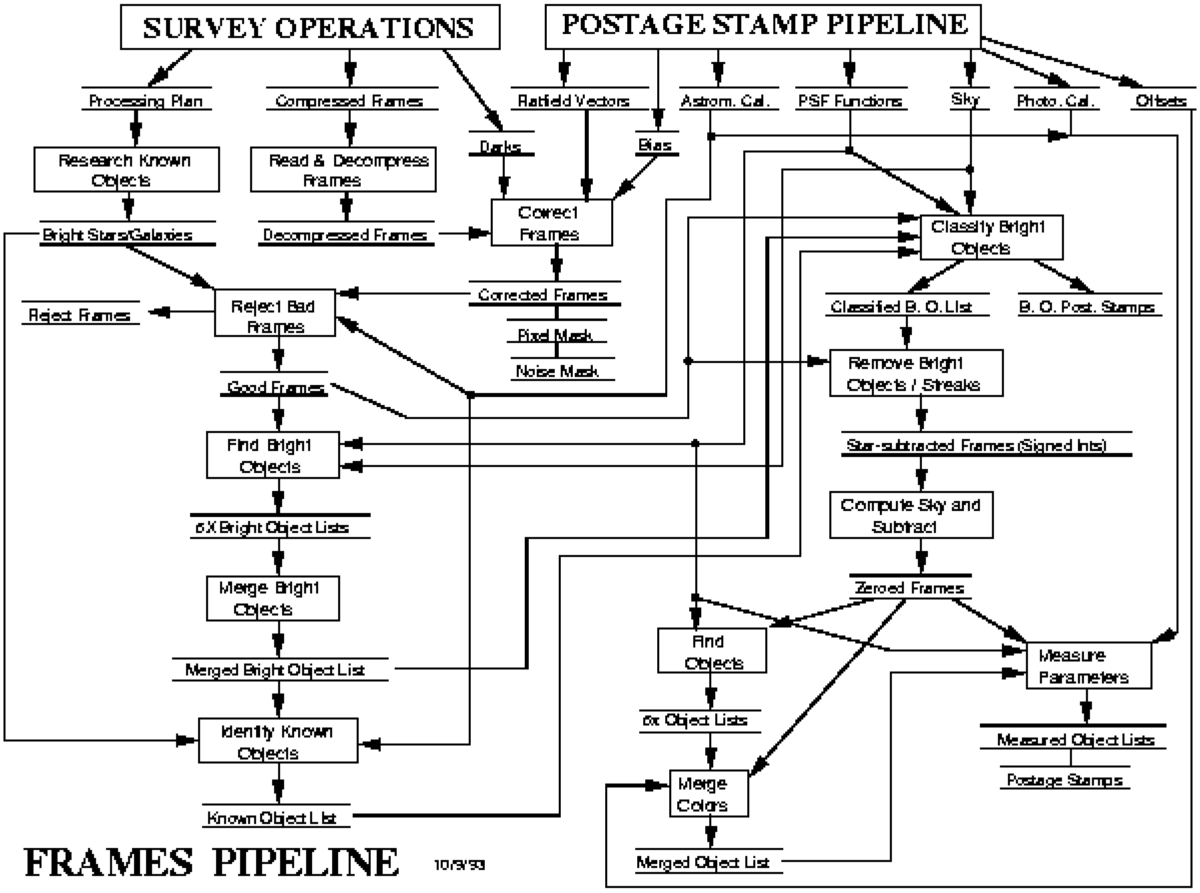


Flow diagram for the Postage Stamp Pipeline

The data flow through the PSP is shown in Figure 10.7. The PSP calculates the bias vector, flat-field vector and point spread function (PSF) and interpolates them to the center of each frame. It uses bias and data quartiles produced by the DA system and an input file describing each CCD, which contains information on bad columns etc., which are used to interpolate the data in the postage stamps. The postage stamps from the SSC are filtered for unsuitable objects (saturated stars, galaxies) and the point spread functions calculated by fitting a classified PSF model to the image, as described above in Section 10.3.3. The composite PSF for each frame is then calculated from the weighted mean of the PSF star data in that frame. If there are not enough PSF stars on a given frame (a likely occurrence in the u' band at high galactic latitudes) the mean PSF is found from the stars in several frames and interpolated to the center of each frame. The frame correction and calibration vectors are calculated from the input quartiles and overclock data. The outputs for each band are: a bias vector for the entire imaging run; a flat-field vector for each frame; bias drift values from both amplifiers in each frame; the sky value for each frame; and the sky gradient for each frame.

The Frames Pipeline

Figure 10.8



Flow diagram for the Frames Pipeline

The Frames pipeline operates on one set of frames (a field) at a time. The following is a much simplified outline of its operation (see Figure 10.8):

- Start

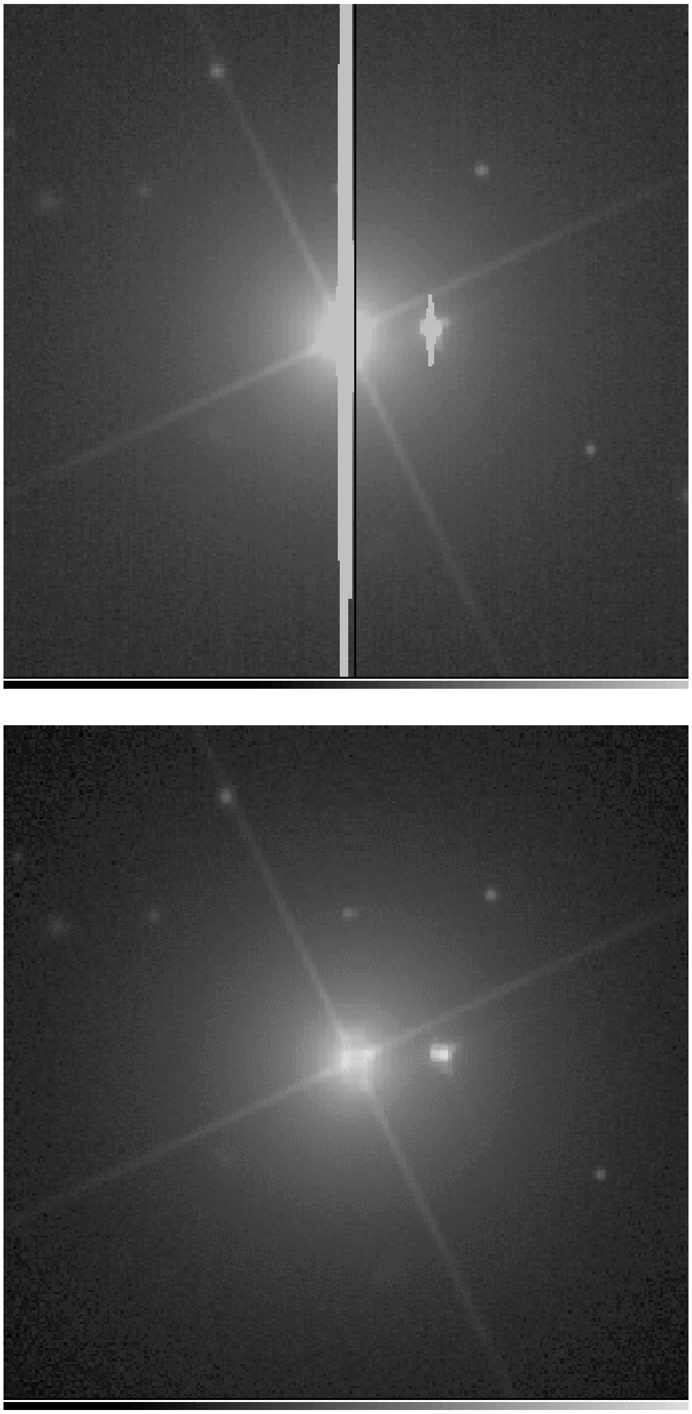
- *First Loop:* for each frame (u', g', r', i', z'):
  - Correct Frames
  - Find and Merge Bright Objects
  - Subtract Bright Objects
  - Find Faint Objects
- End
- *Merge 5 colors*
- *Second Loop:* for each object (u', g', r', i', z'):
  - Measure Object
  - Write output
- End
- Write remaining outputs
- End

The inputs required by Frames are: the CCD hardware specifications (locations of bad columns etc.); raw frames in each color with overlaps attached; bias vector for each CCD; flat field for each CCD; model PSF in each color; calibration data (the approximate flux calibration, and coordinate transformation for aligning the pixels in each frame to the fiducial frame, r' , because the astrometric CCDs observe in this band).

Correcting Frames

First, the frames are flatfielded using the flat field found from the data quartiles in the PSP. They are then corrected for defects. Three types of defect are corrected by interpolation: bad columns, bleed trails and cosmic rays. The survey CCDs are required to have any defects no more than one column wide, for which almost-perfect interpolation is possible because the pixel size samples the PSF at the Nyquist limit. The interpolation algorithm can also treat defects more than one column wide, such as bleed trails; however, these can be only partially recovered.

Figure 10.9



Interpolation over bad columns. The upper panel shows simulated images of a bright star with a bleed trail and a bad column just to the right of the trail. Lower panel: interpolated and corrected image. The faint object just above the center of the bright star has been recovered.

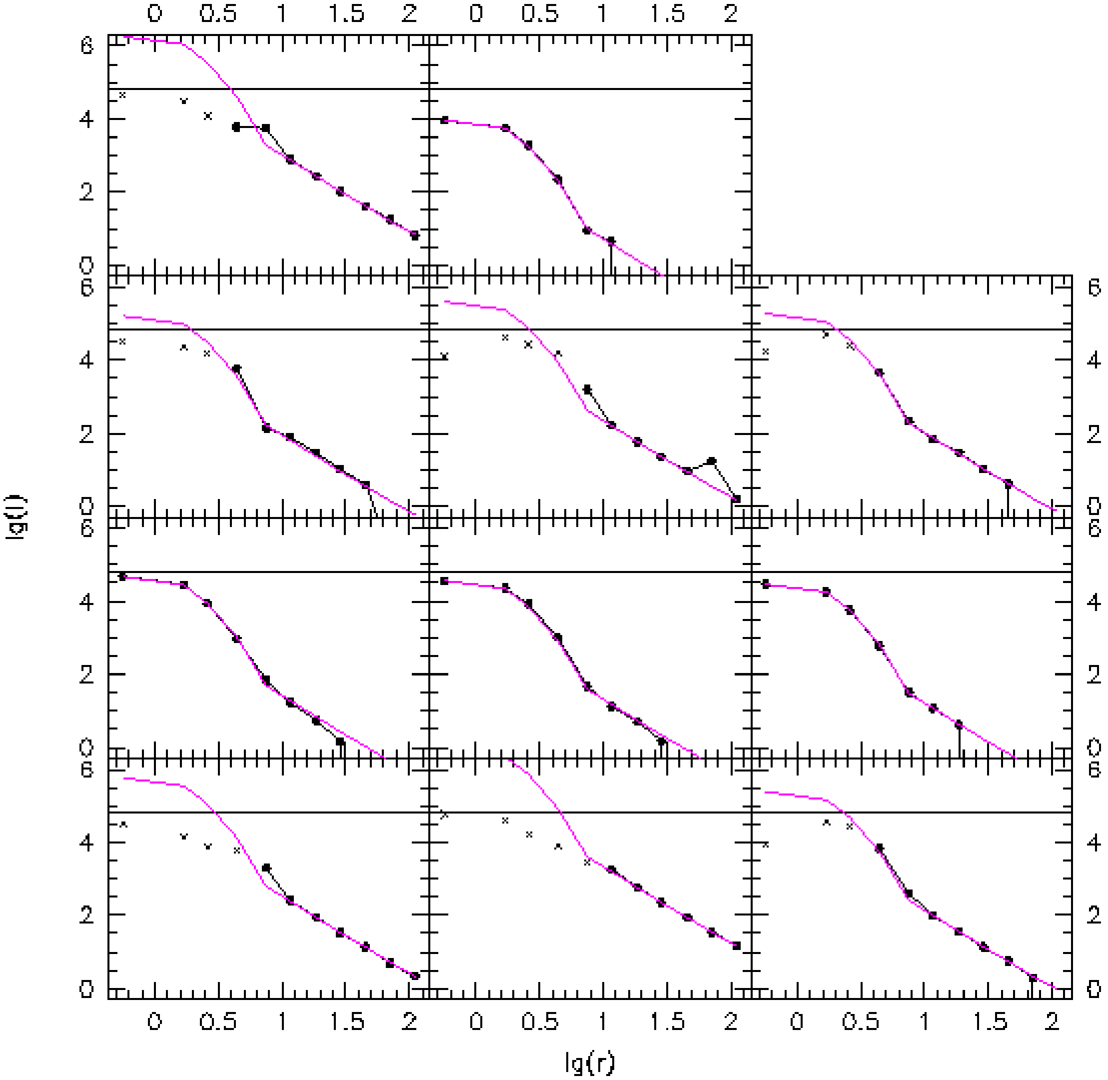
The correction of bad columns is done using linear prediction (see Press and Rybicki 1993 for a discussion) to interpolate the data across pixels where it is missing. This is done by calculating the interpolation coefficients for a seeing-convolved point source, i.e. the PSF, sampled by the pixel response function of the CCD. These can be calculated directly in the case of noiseless data, and the value at the location of the missing data interpolated using data on either side of the bad column(s). The algorithm used by PHOTO uses +- 2 columns on either side of the bad column, and, for bright stars containing a one-column defect, recovers the flux to about 1%. This method cannot be used for wide defects, such as bright star bleed trails; for these, the mean of the 2 values on either side is used. Figure 10.9 shows the correction of a frame with bleed trails and one bad column.

Cosmic rays are corrected in an analogous way. We do not have more than one frame of a given region of the sky, so cosmic rays cannot be found by comparison of two images in the same field. Rather, they are found because their signal is outside the band limit, i.e. the difference in counts between two adjacent pixels is larger than allowed by the PSF. Cosmic rays are found by comparing the intensity of each pixel with those of its eight neighbors, and removed by interpolation as described above. Information as to which pixels in an image are replaced with interpolated values is recorded in a mask image (which is highly compressible and therefore makes minimal impact on the data storage requirements).

Object Finding

Object finding is done in two stages: find (and remove if stellar) bright objects; and find faint objects. The reason for this two-stage process is that bright objects have large scattering wings, which on average cover something like one third of the sky at the 1-electron level in the Survey images. Since all object finding is done by threshholding, this could result in very different efficiencies for finding faint objects, especially close to the limit, as a function of position on the sky.

Figure 10.10



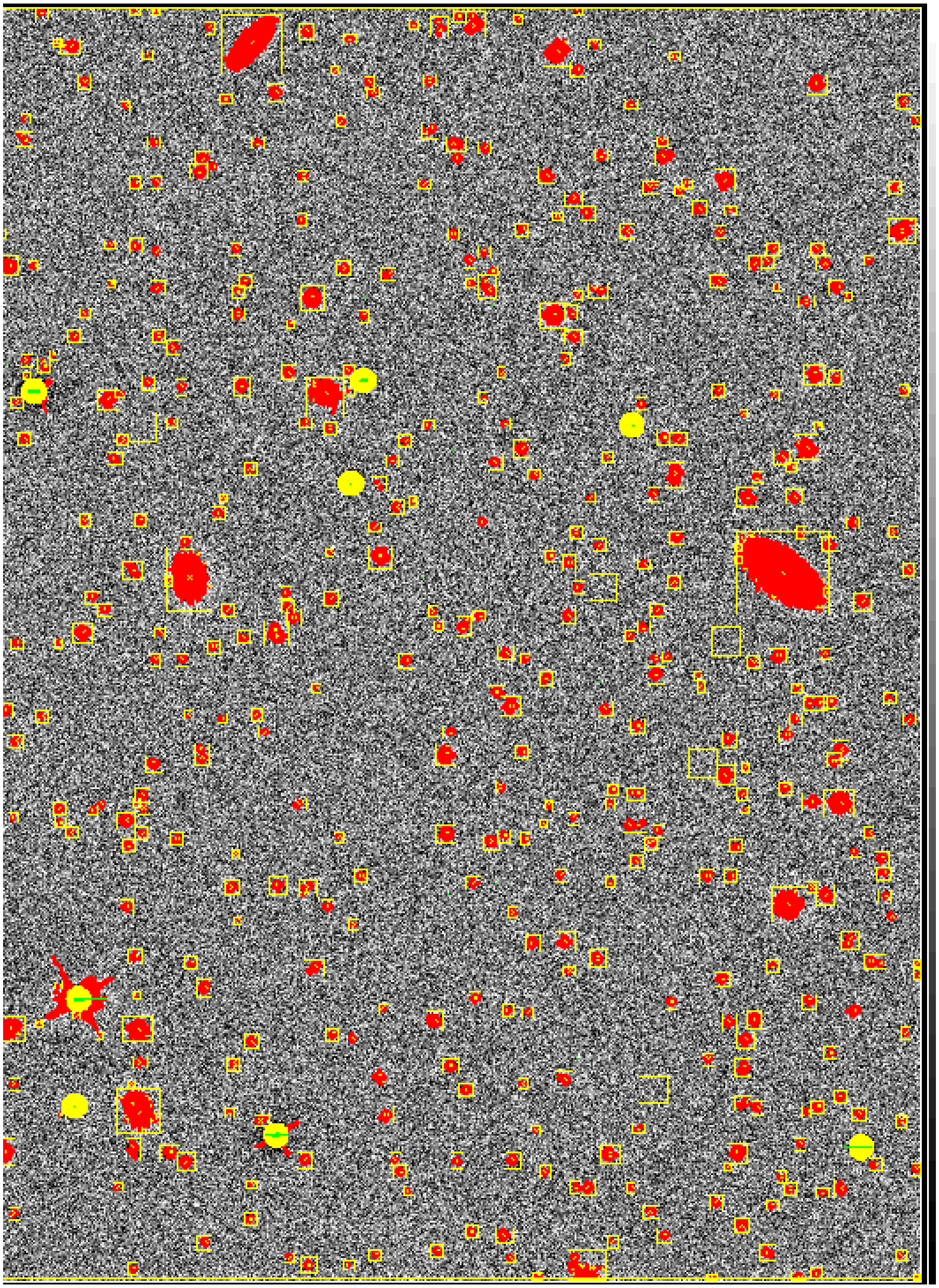
Radial profiles of bright stars. The (r' band) data are from a simulated photometric frame. The filled circles are those used to derive the composite profile, and the crosses mark points that may have been contaminated by saturation. The horizontal line is the saturation level. The magenta line shows the derived composite profile scaled to the data.

Since objects are found by the standard threshholding technique, the data are first smoothed with the PSF, to optimally detect stars. The mean sky level and noise are found by median-filtering the frame, and bright stars found by threshholding at a high level (currently 60 sigma , about 16^{m}). The bright stars so found are then removed by modeling, fitting and subtracting.

The aim is to remove the power-law scattering wings around bright stars. To do so requires constructing a model PSF which consists of (currently) two Gaussians plus power-law wings. Constructing this model PSF is complicated, however, by the large dynamic range of the data: stars which are bright enough to have power-law wings whose measurement is insensitive to the sky level are saturated in the core of the profile, while unsaturated stars, for which the core PSF can be measured, are not bright enough to have measurable power-law wings. Accordingly, a composite profile is measured by fitting together observations of saturated and unsaturated stars. Figure 10.10 shows profiles derived this way. These fits allow determination of the fluxes of even saturated stars. Tests with simulated data show that we can determine magnitudes accurate to 5% for saturated stars to about 12^{m} .

The code fits and removes the wings of bright stars. This subtraction is done to roughly 0.5 DN, with suitable dithering to obtain proper noise statistics in the other parts of stars. Note that some of the faint objects in the star's wings are found as `bright' objects because they are sitting on top of the bright wing and contain enough counts to be detected at the bright object threshold.

Figure 10.11



Object finding. A sample simulated r' frame is shown. Pixels deemed to belong to an object are colored red. Yellow pixels are the masked central regions of bright star images. Other masked pixels (measured to be saturated) are colored green. Squares are drawn around each detected object. There are also several blank fields to be observed by sky fibers during spectroscopic observations.

The corrected frames output by PHOTO will have bright stars removed. The parameters describing each star will be stored in the data base, so that the unsubtracted frame can be reconstructed if desired.

The object finder is now run again to detect faint objects. These are found as connected regions of pixels which lie above a given threshold. Since the data have been smoothed with the PSF, point sources can be single pixels at the limit.

The faint object finder is designed to work to the detection limit of the data. Its performance has been extensively tested by the JPG using simulations, and the algorithm works as theoretically expected; at magnitude 23.2 in the sensitive bands (g', r', i') the point source detection rate is about 50% and the contamination rate less than 5%. The detection and contamination rates depend of course on the threshold level. The optimal threshold can be fine tuned using simulations, and will be fixed during the test year when data from the imaging camera have been extensively run through PHOTO.

Figure 10.11 shows the results of object finding in a simulated r' frame. PHOTO also identifies several regions per frame with no detectable stars or galaxies. These `sky objects' are to be used to locate the fibers which measure the sky spectrum during spectroscopy.

At this stage PHOTO outputs the corrected frames and masks as described above, and a 4 x 4 binned frame (which is both useful for searching for low surface brightness objects and carries sufficient information about the noise characteristics of the data).

Merge Objects

In preparation for measuring the object parameters, PHOTO now merges the detections of the object in all five bands. To do this requires the coordinate transformations between the CCDs in each band from the astrometric solution.

Measure Objects

PHOTO first measures the centroid in each color by fitting the PSF. The image is then shifted by sinc-interpolation so that the center lies at the center of a pixel, and the radial brightness profile measured. To keep execution time to a minimum the shifting is done only for the inner ~ 5" in radius; the outer pixels keep their original values.

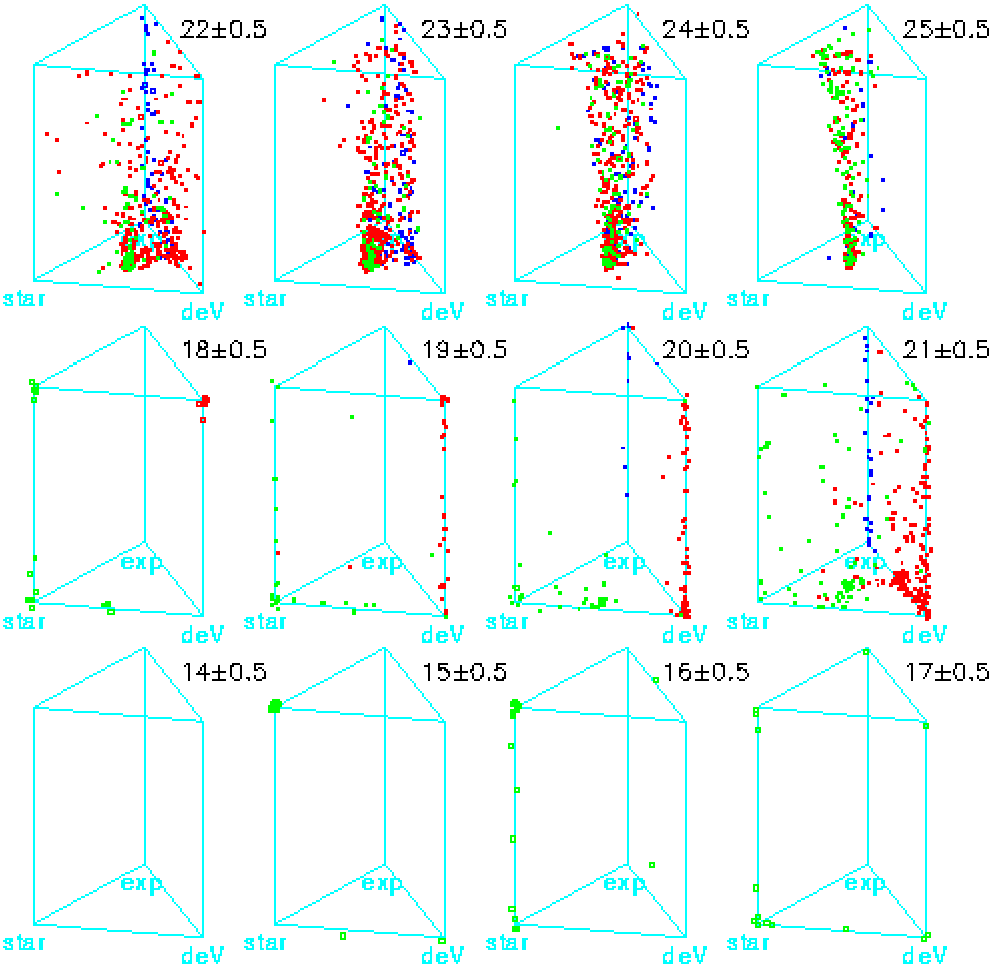
We measure the azimuthally averaged radial profile by measuring the mean and median of the data in a set of concentric logarithmically spaced annuli centered on the center pixel. The median rejects, say, bright stars projected onto an extended galaxy. However, this calculation fails for the case of highly elongated objects (e.g. edge-on galaxies) because the median value within an annulus will just be the sky. We could use the mean instead of the median, but this would make us very sensitive to poor deblending of overlapping images. Accordingly, we go one step further: each annulus is divided azimuthally into 12 sectors. Medians are calculated within each sector, and the radial profile is then the mean of the twelve sectors. Tests have shown that the resulting profile is robust to inclination, as well as various forms of contamination.

The ellipticity and position angle are measured within the 1 sigma circular boundary by calculating the intensity-weighted second moments of x/r and y/r. Several flux densities are measured: the best-fit point source flux (by fitting to the PSF); the 50% and 90% Petrosian fluxes and radii; the flux within a 3" aperture (the spectroscopic fiber diameter) under a fiducial seeing, and an isophotal shape, with the exact isophote level to be decided on during the test year. These aperture fluxes are all calculated by sinc-interpolation to properly account for the pixellation. The various fluxes are calculated in all bands, including those in which an object is not detected, to give a proper and consistent statistical description of the data in all five bands.

This last raises issues with representation of the data, because we will determine quantities like " -2+-4 units" which cannot be represented as magnitudes. Neither do we want to record linear units (Jy) because of the very wide range in brightness of real objects. We will likely use a pseudo-magnitude, i.e. "m" = $A + B \times \sinh^{-1}(F)$, a quantity which handles negative numbers, is linear for small F and tends to a logarithm for large F.

We now fit the median counts in the sectors to three simple radial profiles; a point source, a declining exponential and a de Vaucouleurs (1948) profile. This is done by a precalculated library of these functions with a range of scale lengths and inclinations, smoothed by the appropriate PSF. The best fit model parameters (peak flux, scale length, inclination, position angle) and the likelihoods are stored for each of the three types. The likelihoods can then be used to do a simple classification: star, spiral galaxy or elliptical galaxy. Any fancier classification or fitting, such as a bulge-disk decomposition will likely await the development of off-line pipelines to further manipulate the data, as described below in Section 10.3.5.10. In any case, this is the method we use to do the star/galaxy separation. Extensive tests on simulated data show that this classifier works well essentially to the data limit, as shown in Figure 10.12. The object classifications are shown as a function of their `real', i.e. input, type. In Figure 10.12 object likelihoods are plotted; an object lying at the `star' location is a point source, one halfway between `exponential' and `de Vaucouleurs' is extended but equally likely to be a spiral or elliptical galaxy, and so on. The vertical axis of the classification "prisms" is goodness of fit, ranging from 1 at the bottom to 0 at the top.

Figure 10.12



Performance of the object classifier. This is shown as a function of magnitude in g' , using simulated data. Green = input star; blue = input exponential disk; red = input de Vaucouleurs profile. The points corresponding to each object are located according to the model which best fits their radial profile. The vertical axis is the a measure of goodness of fit, with good fits at the bottom, and poor fits at the top.

Once a model galaxy profile is calculated, it is a simple matter to calculate a total flux. We believe that this is not likely to be a reliable flux because of template mismatch etc., but it is likely to provide good global colors. We will work these out by using the best-fit scale length in one band (probably r') to fit the peak amplitude in all five bands, and calculate the colors from these. These colors are of particular importance for determining photometric redshifts. We have learned the obvious from this work; the PSFs must be very accurately represented in order to make good fits. It remains to be seen what will happen when we get real data.

Further structure parameters under development are some kind of texture parameter, calculated from the residuals left from inverting an image and subtracting it, and the representation of a radial profile as a series of orthogonal functions (PSF, de Vaucouleurs, exponential) which may allow the fraction of light in a point source nuclear component to be defined as a function of color and be a powerful diagnostic of AGNs and fuzzy quasars.

Finally, overlapping images will be deblended using the above models. During object finding, overlapping images are tagged (there are interesting problems associated with the different level of overlap in the different bands). They are then deblended using object models as above to estimate the total flux in each object. The parameters for both parent and child objects are recorded in the data base, appropriately flagged.

The result of these calculations is a large output file containing a catalogue of objects with a large number of measured parameters and uncertainties, plus pointers to the atlas image for each object. These outputs will be stored in the operational and science data bases. The nominal performance goal is to allow the robust and carefully controlled selection of spectroscopic targets, but these data will obviously enable a vast amount of science. There has been a very large amount of fundamental algorithm development for this work; the code will be made available to the scientific community and a detailed description of the algorithms will be published. Descriptions of many of these algorithms have been written up as internal documents (Lupton, 1993-6).

Figure 10.13

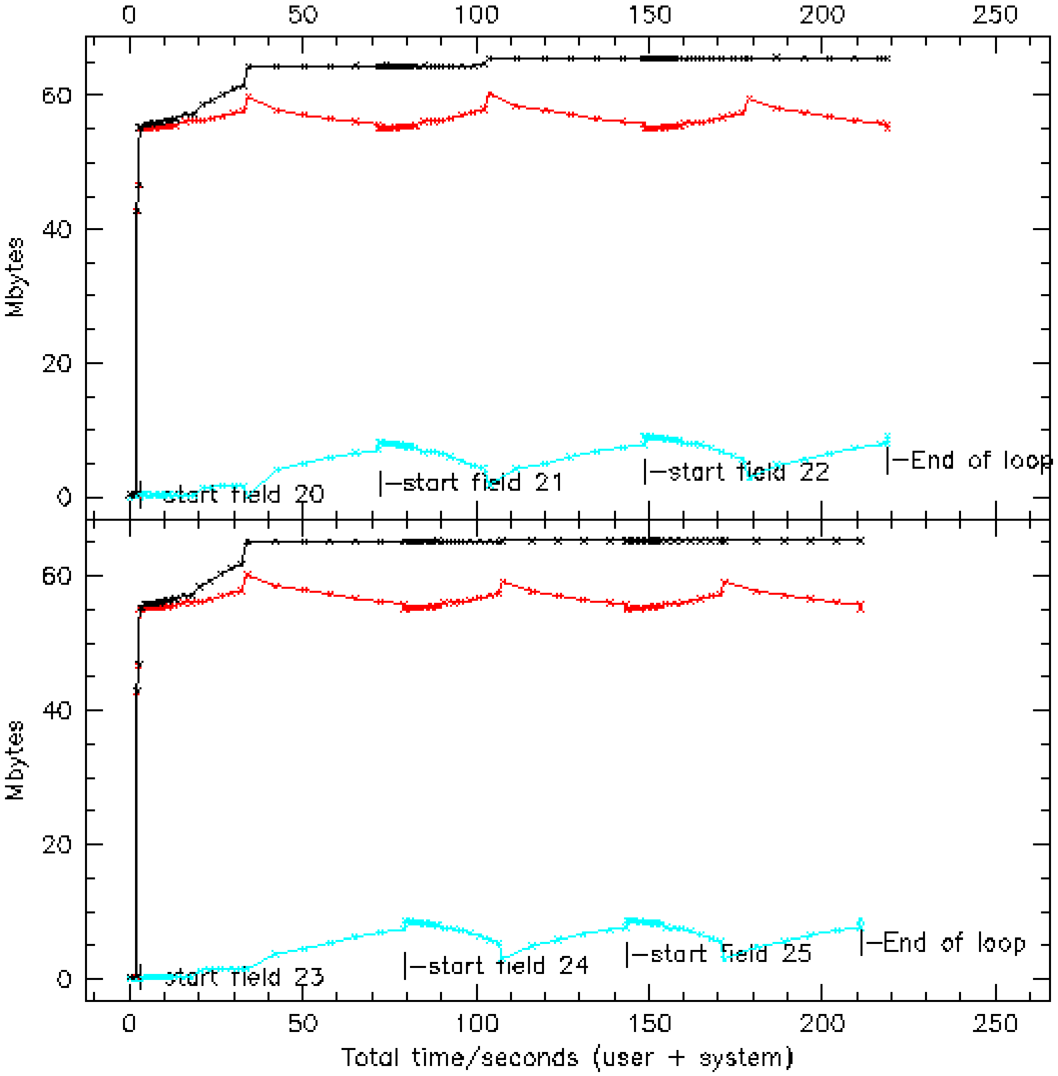


PHOTO performance. The graph shows the memory usage in MBy versus CPU time for a run of six fields, five frames each, on an SGI Origin 200. Black: total memory allocation by PHOTO. Red: memory usage. Blue:

residual. The executable occupies another 15 MBy approximately. The two panels show the outputs of the two processors.

Atlas Images

Atlas images are cut and stored for each object. Atlas images are subimages of size sufficient to contain the pixels belonging to the object, and the dimensions are set at the 3 sigma threshold, plus a border whose width is approximately that of the PSF. Unsaturated stars are thus well contained in a 29 x 29 pixel postage stamp. The dimensions of the atlas image for an extended object is set by those for the band in which the object is largest. Note that we cut an atlas image in all bands, even if we do not detect the object in all bands. Atlas images are also cut at the positions of sources from selected catalogues at other wavelengths (e.g. FIRST and ROSAT), whose positions are fed in from the data base on a frame by frame basis, with coordinates translated to pixels using the results of the astrometric pipeline. We will plan to cut these images without checking whether an image already exists because an optical object has been detected at that location; it is simpler this way and we can afford the overhead in storage required. Further, because objects look so very different at different wavelengths, it is useful to have an image cut to a size defined by the size of the object, e.g. it is likely that a larger image will be cut for a double-lobed `FIRST' source than would be cut for its associated optical galaxy. Further analysis which requires images of a larger area (e.g. searches for low surface brightness emission at large radii) can be made using the 4 x 4 binned frames.

Performance

As well as scientific and numerical accuracy, a prime requirement of the pipeline software is that it be robust and efficient in its use of memory and CPU. PHOTO handles the largest amounts of data in the SDSS by far, and it must be able to keep up. Development of fast, efficient code has been a prime concern; the calculations are done in integer arithmetic, memory allocation and use is very carefully controlled, and PHOTO operates on multi-processor machines in essentially parallel mode; fields are passed to processors as they finish the previous field independent of their ordering in the sky. The imaging survey takes data at 5 MBy/second; Figure 10.13 shows the memory and CPU usage by PHOTO running on a 2-processor SGI Origin 200 machine. Six fields (all columns of the camera) take 200 seconds to reduce on this machine and took 40 seconds to acquire; thus execution on the production machine, a DEC Alpha with 10 processors each roughly as fast as the individual SGI processors, can reduce the data essentially at the rate it was taken. Given the fact that the longest night's observing occupies only a third of 24 hours, and that photometric data can be taken only for a few nights a month, is is clear that the software and hardware are in place not only to keep up with the photometric data flow but to reduce it many times, if need be.

Future Development of the Photometric Pipeline

The entire focus of the photometric pipeline to date has been on readiness for the analysis of the Northern imaging survey; to analyze the photometric data rapidly, reliably and accurately enough that the reduction can keep up with the data rate, spectroscopic targets can be selected, and the performance of the imaging camera and of the conduct of the Survey can be monitored. Therefore, although the spectroscopic targets are brighter than 20^m , we have ensured that objects are detected to the limits of the imaging data; these data are scientifically interesting, but more immediately they are a powerful diagnostic of imaging performance. Further, were we not to find objects to the limits of the data, the entire raw data would need to be reprocessed at some future time with different software. However, we have paid minimal attention to the processing of the data for the largest ($D > 1'$) galaxies, which have internal structure of considerable scientific interest, given the difficulties of dealing with frame edges. For these objects PHOTO will find the centroid, cut atlas images, and tag them as a `large' galaxy for automatic spectroscopic observation. For the same reasons, the Southern photometric data will initially be processed like the Northern data - all imaging data will be run through PHOTO and separate object catalogues produced for each separate observing session.

Once PHOTO is developed to Level 2 and the Survey is well underway, further software development will take place. The following tasks, which do not affect survey operations, allow us to carry out further scientific goals of the survey.

Southern Survey Coaddition

The Southern imaging survey will repeatedly scan strips of the sky (Section 1.5), and to exploit the increased depth afforded by these data they must be co-added. This is not straightforward because data from different nights will not exactly map onto each other and each night's data must be fit to the cumulative coadded map before being added in. This fitting will also, however, allow us to add in data taken under non-photometric conditions, since it can be accurately calibrated by fitting to photometrically accurate data. We can therefore use more of the time for photometry in the South than we can in the North.

Southern Survey Subtraction

Just as we wish to co-add individual exposures on the Southern stripe, so too do we wish to carry out subtraction between them. Although many variability and proper motion studies can be done can be done at the catalog level, if we wish to search, for example, for supernovae close in to the cores of galaxies, we will need to be able to subtract "before" and "after" images of the galaxy from one another. This probably can all be done with Atlas Images, rather than having to go back to the corrected frames.

Compressed Data Cube

It will be of interest for many reasons to build a compressed version of the object catalogue as the survey progresses, to enable science analysis of several kinds. If each object found by PHOTO is described by a series of indices (position to a few arcseconds accuracy, magnitude to 0.1^m - 0.5^m , colors to 0.1^m and, for galaxies, color redshift) we can build up two cubes describing the density of stars and galaxies in summary form. The stars cube describes color and magnitude distributions as a function of position in the sky and can be used in quasar selection as well as star count, cluster finding and correlation studies. The galaxy cube can be searched for clusters (it will be interesting to see how many galaxy density enhancements coincide with X-ray sources and with Bright Red Galaxies) and may be adequate for many statistical studies of large scale structure. The PHOTO group has developed code to construct and analyze these cubes (Kepner et al. 1997); the code will be incorporated into the data base at some suitable time.

Merged Pixel Map

The description of target selection above discusses how the object catalogues produced by PHOTO will be merged. We plan also to merge the reduced frames into a continuous five band image of the sky and to provide tools to access any part of it.

Sky Map

A second map which can be made as part of the development of the Merged Pixel Map is the sky map with all detected objects subtracted. This can then be searched for extended, low surface brightness objects.

Atlas Image Pipeline

The atlas images of the brightest, largest galaxies are likely to be of widespread interest. As described above, we will not engage in any extended analysis of these images in the routine data processing, but plan to investigate automated morphological analysis of the images, to produce analytical classifications, radial color profiles, color maps etc of a very large number of galaxies, to relate their morphological properties to the underlying dynamics and stellar populations. If this effort is successful, we can investigate the robustness of the classifications by degrading the images to simulate the effects of distance and redshift. This effort may allow a real comparison between the properties of nearby and distant galaxies.

The Southern Survey

Processing the imaging data for the Southern Survey will be substantially more complex than processing that for the Northern Survey. The DA system remains the same, though. The imaging for the Southern Survey has two main goals; to find variability from scan to scan and to make a deep survey using all of the data. Achieving the former goals means processing the data through PHOTO as for the Northern Survey and doing image matches and comparisons via the data base, where the data from previous scans are stored. There are some subtleties here, because we can use data taken under pristine photometric conditions to calibrate data taken under less good conditions, as outlined above. For the deep survey, we will need to do a carefully registered composite map as described above, and the incidence of blended objects (with overlapping images) will be about 10% higher.

Extending the Survey to Lower Galactic Latitude

There is considerable interest in the eventual extension of the imaging survey to lower galactic latitudes. Although this is not part of the current SDSS planning and in any case will not be done for many years, it is of interest to consider the data handling problems. In the survey region, there are already about three times as many stars to our limit as there are galaxies, and that ratio will increase dramatically at lower latitudes, especially at faint levels. At about 20th magnitude, there are an order of magnitude more stars per unit area in the Galactic plane than at the boundary of the Northern survey, and about 30 times more than the average density over the survey region. There is also a very large variation with longitude, and there will clearly be regions in which no algorithm without highly sophisticated deblending built in could successfully cope with the data.

The Spectroscopic Pipeline

The spectroscopic pipeline (SPECTRO) is a large data reduction software package designed and written to completely and automatically reduce all spectroscopic observations made by the SDSS, and will be one of the main software workhorses during the SDSS operations. As described in Chapter 1, the SDSS will obtain about 10^6 galaxy spectra, 10^5 quasar spectra, and spectra of a wide variety of other astronomically interesting objects; X-ray sources, radio sources, stars, sources with unusual properties and of course quasar candidates which turn out not to be quasars. Further, we expect that a substantial number of spectra of different kinds of objects (especially, again, quasar candidates) will be obtained during the test year. These points highlight the two main performance requirements for SPECTRO; first, the systematic, automatic and uniform reduction of more than a million spectra taken at rates as high as 5000 per night, and second, the ability to recognize, and deal with, a host of different types of spectra, from low-metallicity (or flaring) M stars through normal stars and galaxies of all types to powerful AGNs. This is a unique and unprecedented challenge.

SPECTRO has already gone through two levels of development. Level 0 designed and constructed the algorithms necessary for obtaining scientifically useful spectra and redshifts from the raw data and was completed in late 1994. Level 1 was largely complete by late 1996 and addresses the key issues of computer compatibility, operational speed, efficient memory usage and integration with the other software. Further algorithms were also included. This code is basically sufficient to run the SDSS test year operations. Some refinements to the algorithms are bound to result from the analysis of real data, but the core of the software is expected to remain the same for the survey proper.

Performance Goals for SPECTRO

SPECTRO currently consists of two pipelines. The first reduces the raw 2D spectral frames from the DA system to 1D calibrated spectra. The spectra are to be obtained in three 15-minute exposures to allow for cosmic ray rejection, reduce the amount of data lost to such events as changing weather, avoid saturation of the night sky lines, provide various other internal checks and (perhaps) allow the total exposure time to be built up out of observations taken on more than one night. The first of the SPECTRO pipelines produces a single calibrated spectrum for each object. The second SPECTRO pipeline classifies the spectra and performs various scientific analyses on them, including obtaining the redshifts.

The operational goals of the 2D SPECTRO pipeline are:

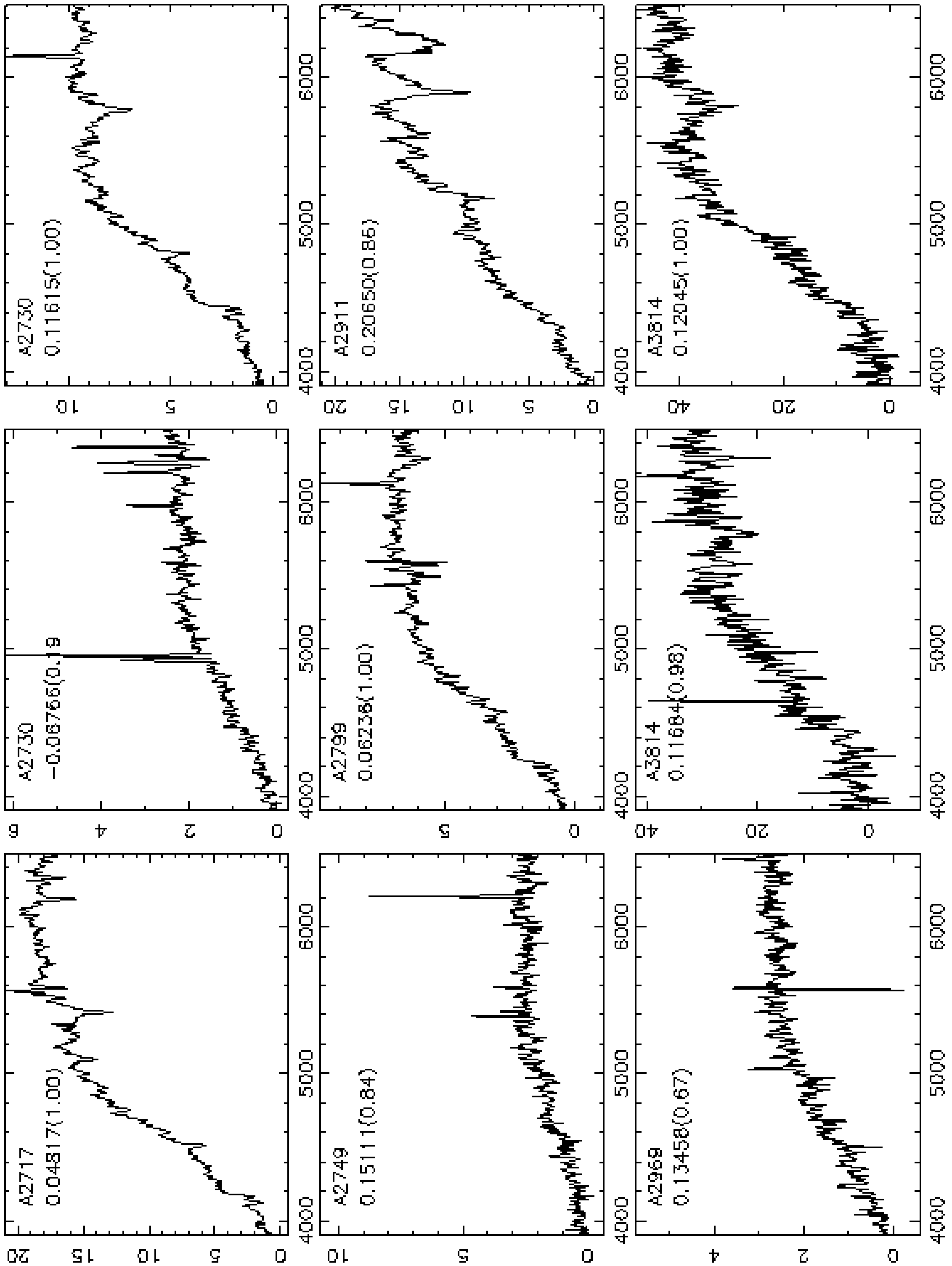
1. To interpolate over bad pixels
2. To bias and dark subtract the raw 2D data frames
3. To trim the frame
4. To flat-field using calibration frames taken at the same telescope pointing position before and after an exposure on the sky.
5. To optimally extract 1D spectra from this 2D frame.
6. To apply wavelength calibration, rebin to a common wavelength solution, and sky subtract.
7. To combine the three individual exposures for each object.

Many of these tasks are carried out using the IRAF routines CCDRED and SPECRED. These have been slightly altered to operate within the SDSS software framework and to allow both interactive and automated processing.

The operational goals of the 1D SPECTRO pipeline are:

1. To put the red and blue halves of the spectrum together.
2. To mask all bad pixels and pixels contaminated by strong sky emission.
3. To fit the continuum of the spectrum.
4. To find and fit emission lines. This process includes measurement of the centroid wavelengths, equivalent widths and peak heights. The current software contains two line finding algorithms, the first being the standard peak-finding algorithm and the second a wavelet-based method that helps deblend close pairs.
5. To determine an emission-line redshift and classify all detected emission lines. A flag will also be set to identify any expected emission lines (on the basis of the spectral classification) which were not seen.
6. To classify the spectrum using a set of template spectra ranging from stars to quasars. The present suite of spectra contains 30 templates, but this collection will grow substantially as a result of the test year observations, which will obtain measured templates using the actual SDSS hardware. A principal component analysis similar to that of Connolly et al. (1995) has also been implemented.
7. To cross-correlate the spectrum with the above set of templates and obtain the absorption-line redshift. The redshifts of the three highest peaks in the cross-correlation function (CCF) will be recorded, along with their heights, widths and confidences (Tonry and Davis 1979, Heavens 1993). Results for the five templates that best match the spectrum will be recorded as well. The internal velocity dispersion of galaxies will be estimated using the width of the CCF peak.
8. To flux calibrate the spectrum to obtain crude (~ 15%) spectrophotometry, using the calibrated photometric images.

Figure 10.14



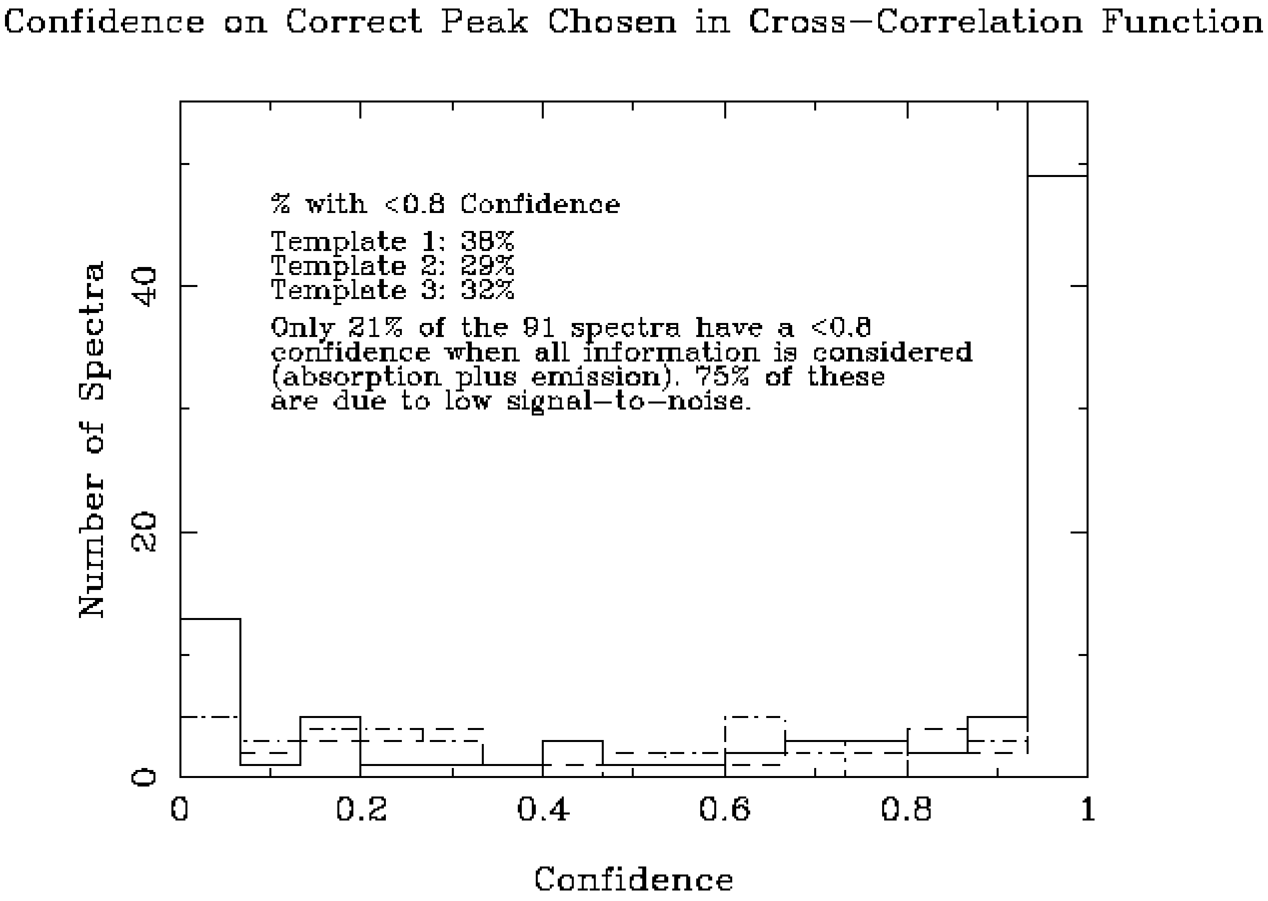
Spectra of brightest red galaxies in Abell clusters. The data are from Collins et al. (1995).

The 1D pipeline has been tested using both simulations and real data. The most informative of these tests to date is the re-reduction of 100 1D galaxy spectra taken from the Edinburgh/Milano Cluster Redshift Survey (Collins et al. 1995). These spectra (some of which are shown in Figure 10.14) represent a fair sample of those from the E/S0 galaxy population we expect to observe with the SDSS; spiral galaxies, however, are under-represented in the sample. We have also constructed an extensive library of simulated spectra of both quasars and galaxies. Testing using this library has been ongoing since Fall 1994; we find that we obtain unbiased redshifts to well below our spectroscopic limit for galaxies.

Both pipelines will have a pre- and post-processor. The pre-processor for the 2D pipeline will ensure that all required frames are present and in order, and will generate the IRAF scripts necessary for running the pipeline. The postprocessor will write the 2D frame to the database and update the operations log. The preprocessor for the 1D pipeline will combine repeat exposures, join the red and blue halves, and rebin the 1D spectra to log(lambda) spacing, thus producing a seamless moderate resolution (2-3 Å) spectrum from 4000 Å to 9000 Å . The postprocessor writes the results to the database (including the binned and unbinned spectra) and assesses the success of the overall reduction: for example, did we get a satisfactory answer that is internally consistent? Do the emission and absorption line redshifts agree? Does the photometric object classification (roughly, star, quasar, elliptical galaxy, brightest red cluster galaxy, spiral galaxy, weird thing) agree with the spectroscopic classification (here is a rich lode for serendipitous discovery). Such intelligent software is required to identify potential problems in both pipelines and will hopefully reduce the number of human interactions, which has been nominally set at 1% (10,000 spectra !).

Preliminary tests of the 1D spectroscopic pipeline using 100 spectra from the Collins et al. (1995) data described above used 5 spectral templates and took 343 seconds on an SGI workstation. Thus a whole nights' worth of data, cross-correlated against 30 spectral templates, will take 24 hours on the same workstation and, on the SDSS production system, will be faster by about an order of magnitude. It is clear from these tests that the spectroscopic data can be reduced in a timely manner.

Figure 10.16



Confidence estimates for the measured redshifts. The sample consists of 91 E/S0 galaxy spectra.

Figure 10.16 shows the histogram of confidences derived from the cross-correlations from this sample (Heavens 1993). These confidences quantify our security that the derived redshifts are real, given the height of the chosen CCF peak relative to the background noise. Over 60 of the spectra have at least one template with a significant confidence. Including information from emission lines, only 21 galaxies, all with signal-to-noise ratio well below that of our faintest galaxies, had no confident redshift determination.

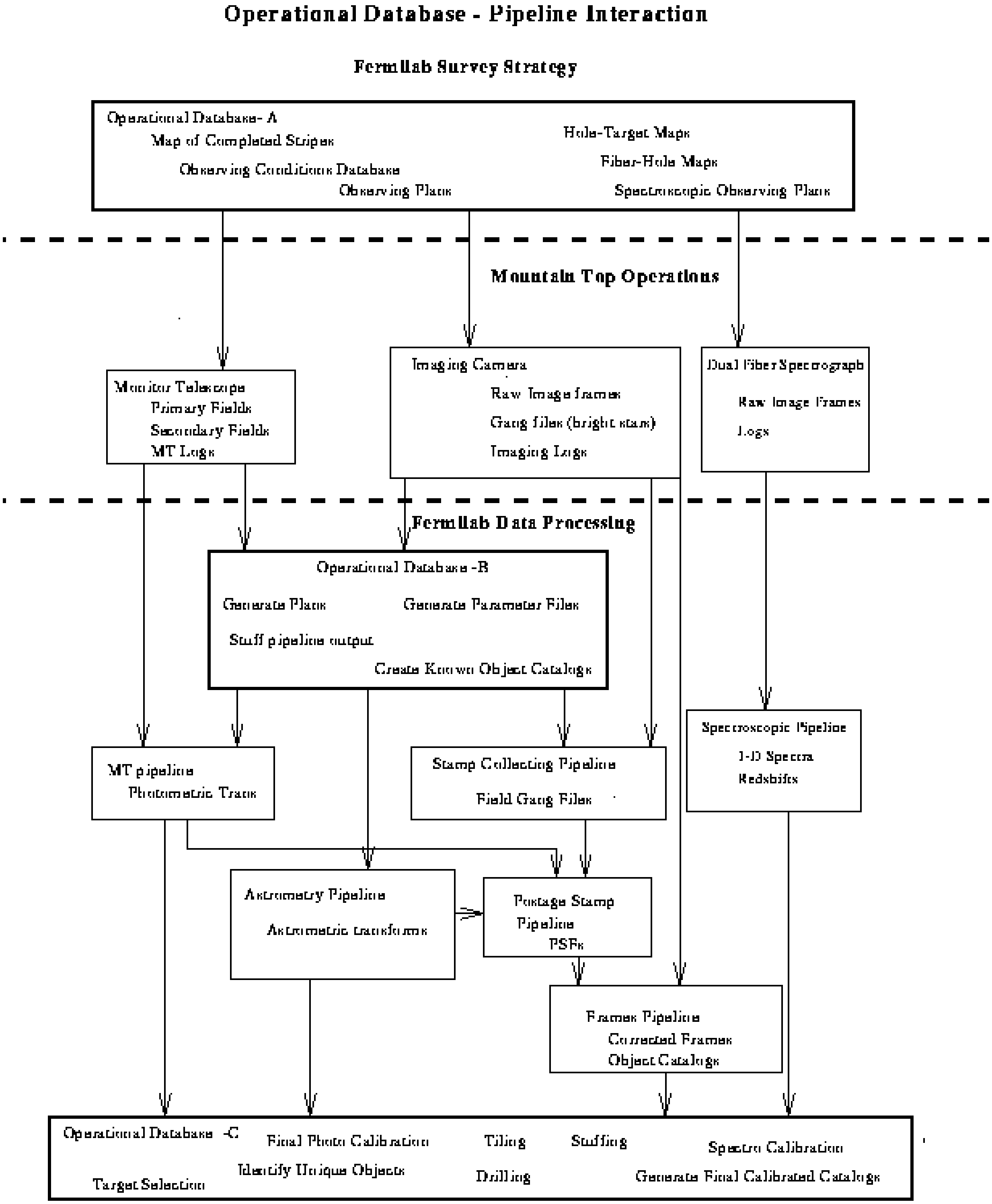
Operational Databases in the SDSS Survey

The duties of the operational database system may be divided into three main subsystems:

1. Guiding the mountain top operations, generating observing plans for the upcoming lunation, and keeping track of which stripes of the survey have been observed successfully so far.
2. Staging raw data from the mountain top tapes to the processing pipelines back at Fermilab. This involves generating processing plans for each pipeline, storing all parameters used to run each pipeline, and generating auxiliary files such as previously known object catalogs to be used by the pipelines.
3. Storing the outputs of the pipelines, applying calibrations, generating final calibrated catalogs of objects to export to the Science Archive Database, and directing target selection for spectroscopic fiber targeting.

These three systems are shown as heavily outlined boxes in Figure 10.17. The flow of information from the operational databases to and from the mountain and to and from the processing pipelines are indicated by arrows. The products of each pipeline are indicated within each pipeline's box.

Figure 10.17



The flow diagram for the Operational Data Base

The operational database system is based on the commercial object-oriented data base system "Objectivity". The interface to the OPDB is supported by the C and TCL-based "DERVISH/SHIVA" interactive programming system developed at Fermilab and Princeton. Currently the system will run on a 6-CPU Silicon Graphics Challenge machine with about 60-100 GBy of spinning disk and access to a hierarchical storage tape robot with 3-TB of on-line secondary storage. Tertiary storage of 12-24 TB is in the form of racks of "DLT" high-density tapes. Data is transferred from the mountain top to Fermilab for processing on DLT tapes.

The completed operational database system has basically been delivered as of Fall 1996. The designing data model is complete, the interfaces to all the pipelines have been defined and the code to stage the data into and out of the database in a systematic fashion is implemented.

Merge Objects and Target Selection

This process includes all operations that occur between the output of the photometric pipeline and the sending of drilling coordinates to the plate vendor. Operations are performed only on the object catalogs. The following steps are involved:

1. Recompute astrometric transformations: This step is included as a placeholder. It will be needed only if the star and galaxy positions measured by the photometric pipeline differ systematically from positions measured in advance by the upstream DA, SSC, and postage stamp pipelines. The output is a set of astrometric transformations for each CCD frame.
2. Recompute photometric transformations: This step is required because the photometric pipeline operated with only an approximate photometric calibration. This step will repeat the calibration by using the same overlap stars observed by the monitor telescope but now with instrumental quantities as measured by PHOTO (see Section 10.3.5.3 above). Full air mass and color term corrections will be included. The output is a photometric solution for each field.
3. Merge object lists: Each piece of sky is uniquely allocated to one run of photometric imaging that passes over that position. For a given run, valid limits are assigned in terms of survey longitude based on an assessment of the data quality. The merging process is a multi-step procedure that involves combining fields, runs, scanlines, strips, and stripes. At the end, each detection of an object is tagged as either a primary or secondary detection. The union of all primary detections will be the final merged object list.

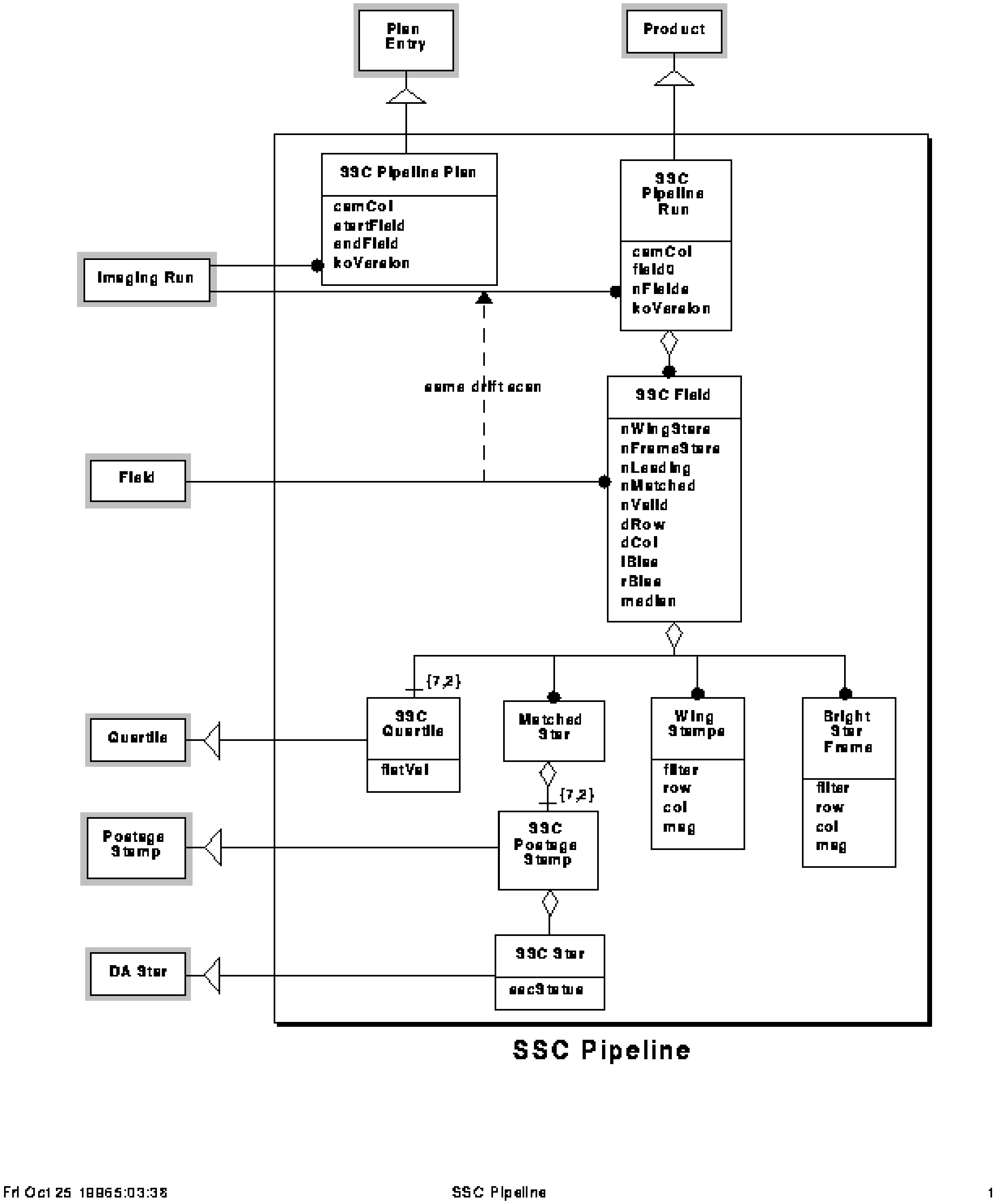
   At this stage, the object lists, flags and other ancillary information are exported to the Science Database.

4. Defined targeting region - Based on the area of sky for which imaging data has been accumulated, a region will be identified by hand for which a targeting run will be performed.
5. Select tilable objects: A list of objects is compiled for which the survey wishes to have a complete sample. This list includes galaxies, QSOs, and very blue stars to be used as reddening standards. Target selection at this stage is done on an object-by-object basis, using the calibrated outputs of PHOTO corrected *a priori* for reddening and extinction, and the target flags are recorded in the Operational and Science Data Bases. Note that an object can have several flags (galaxy, bright red galaxy, radio source etc.).
6. Tile the above list. Tiling is described in more detail in Chapter 8 and is controlled by two parameters: the number of fibers per tile and the number of tiles to be used to cover the region. A tiling flag is set at this stage (i.e. it may not be possible to include a given object because it is too close to another, for example) and sent to the data bases.
7. Select reserved objects - There are several categories of objects for which a fixed number of fibers will be allocated per tile. These include sky fibers, spectrophotometric standards, and guide stars.
8. Select targets of opportunity - The number of fibers that are actually allocated on a give tile after the above processing will vary depending on the local density of galaxies and QSOs and constraints due to the requirement for a minimum separation of fibers. Additional targets will be selected and have fibers allocated based on a quota system. These targets include stars, serendipitous objects, and targets needed for quality analysis. This selection can be done by running automated algorithms as is done for the main targets (and note that these algorithms can be freely changed throughout the duration of the survey because observations of complete samples of these objects are not made). Target lists of interesting objects can also be put in "by hand" at this stage; SDSS scientists, browsing the photometric data, can flag any interesting objects and submit them to a data base containing objects to be observed if possible. The SDSS telescope may only be of modest size, but the rate of acquiring spectra is large enough that decent sized samples of objects in many categories can be acquired this way.
9. Design plates - The sky coordinates for objects on each tile are converted to drilling coordinates (mm) to be used by the plate vendor. The conversion process includes corrections for optical distortions in the telescope, distortions in the plate bending process, and refraction corrections based on a priori guesses at the date, hour angle, and temperature at the time of observation.

After a spectroscopic run is made, the reduced spectra, redshifts etc., are also stored in the data bases, together with yet more flags; those linking them to their photometric objects, and those which describe failures of the spectroscopic observations, due to eventualities such as broken fibers, inadequate signal to noise ratio to obtain a useful spectrum, and so on. There is further discussion of the flags and data base objects in Chapter 11.

The Data Model

Figure 10.18

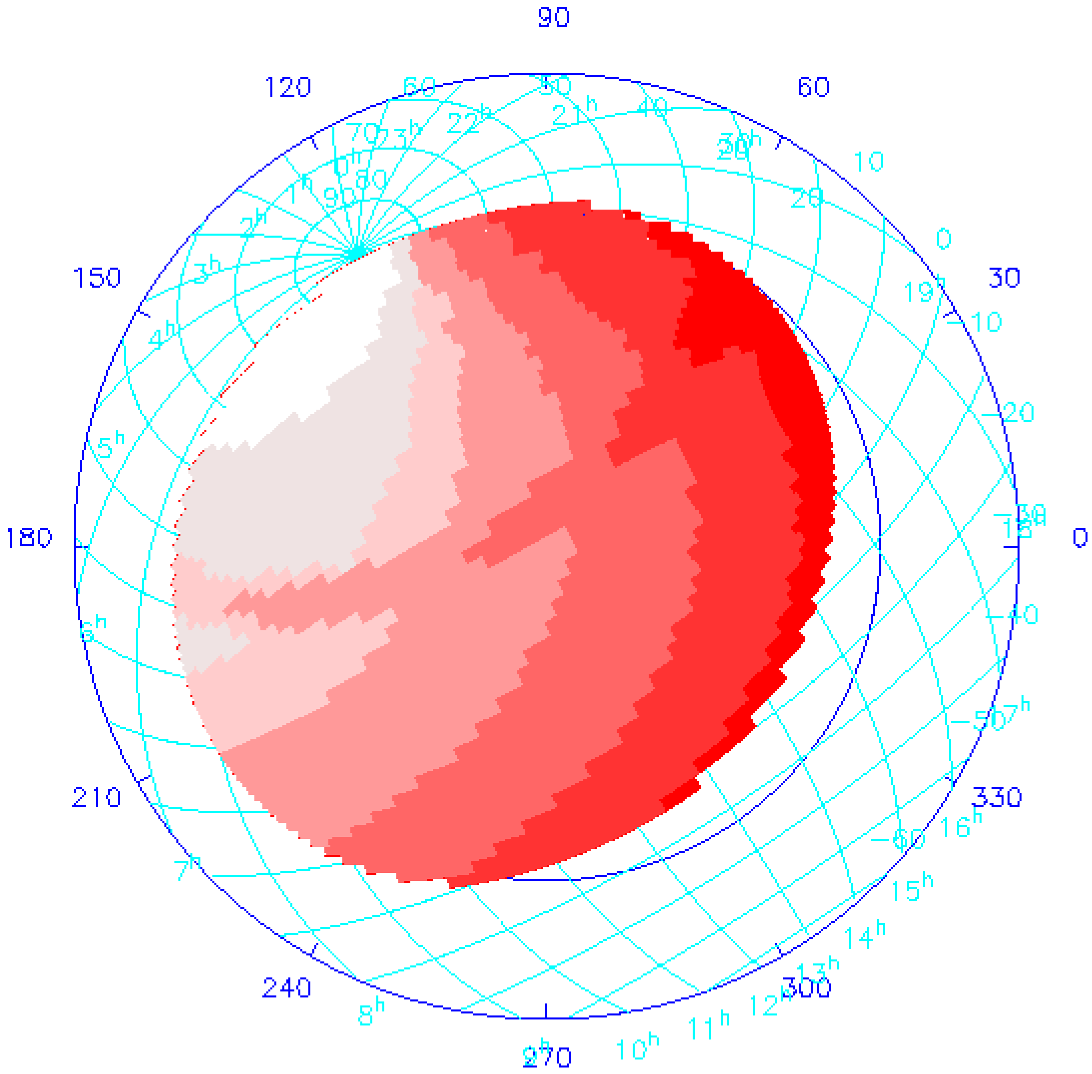


Data model for the SSC pipeline

The software inputs and outputs are tracked with a data model, which is the unique definition of the retained data in the Survey. The data model also defines the interface data between independent components of the Survey (e.g. between separate pipelines, or between pipelines and the operational data base). All pipelines, as well as the operational data base, must be in compliance with the data model. The formal methodology is that of Rumbaugh et al. (1991). The data model is essentially complete for the Monitor Telescope, the DA for the imaging arrays, and the main imaging pipelines. An example, for the SSC, is shown in Figure 10.18. The model is under development for target selection and spectroscopy.

Survey Strategy

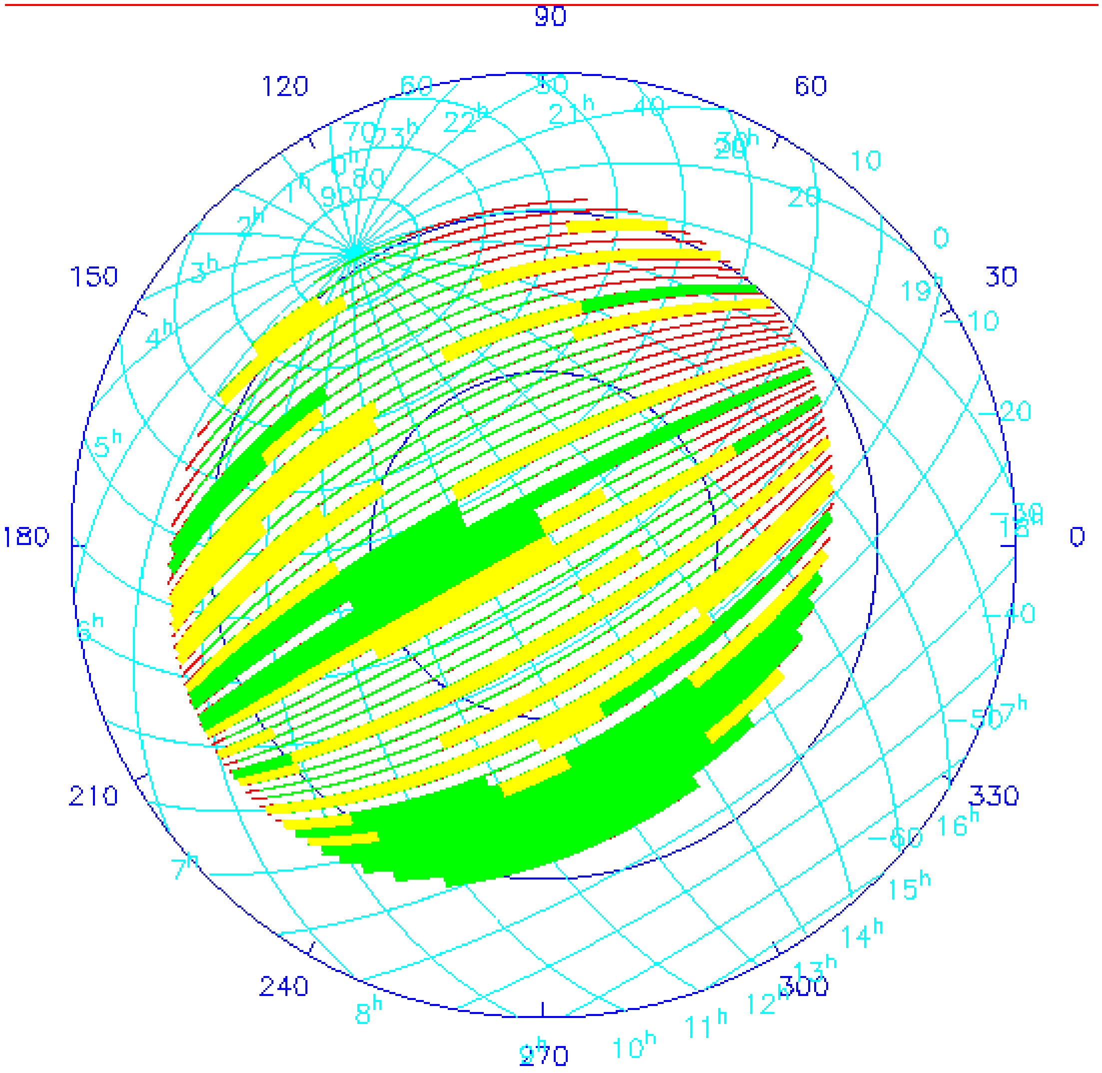
Figure 10.19



Opportunity weights for the SDSS Northern Survey. Red areas are available less often, white areas more often. The feature through the middle of the map is due to avoidance of the zenith. Both equatorial and galactic coordinates are shown.

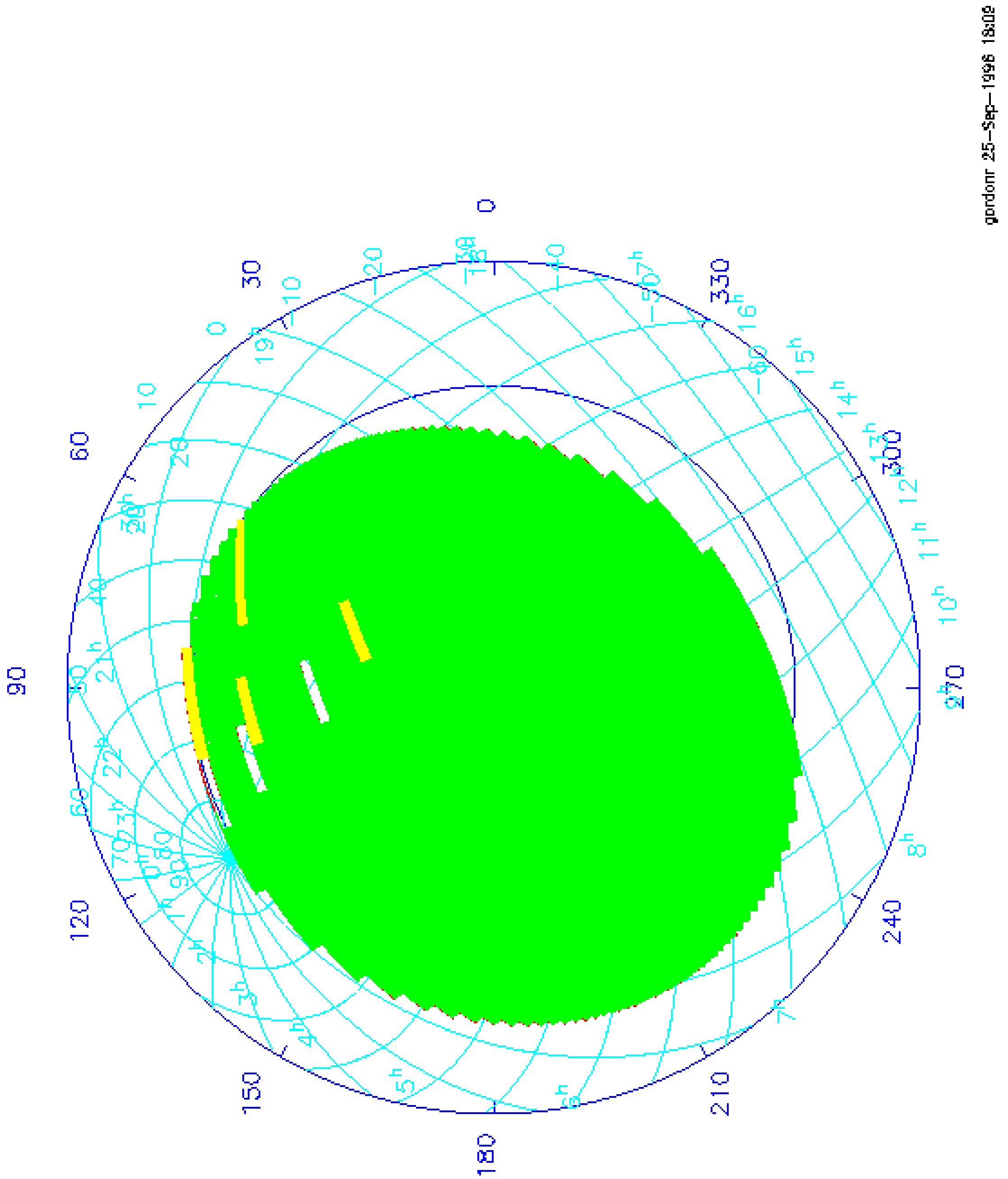
Chapter 1 describes in detail the planning for how the survey will be run. The survey operations will be assisted by a set of sophisticated software tools, loosely called `survey strategy', which: (1) allow the survey to be conducted in an orderly and efficient manner, i.e. ensure that the observing time is efficiently used and that the entire survey time-to-completion is as short as possible. This means that we do not find ourselves, for example, complete except for a few very hard to get pieces of sky, or doing things like always observing at at an airmass of two; (2) monitor the progress of the survey and do the necessary bookkeeping; and (3) allow the total time to completion of the survey to be estimated given assumptions about the weather, etc. At present, we have thoroughly explored the time-to-completion for the Northern imaging survey using these tools (Richards 1996).

Figure 10.20



Completed part of the Northern imaging survey after one year. The survey stripes are outlined in red, and those which can be done at sun/moon rise/set "tonight" are overplotted in green. Complete strips are yellow and complete stripes green. Both equatorial and galactic coordinates are plotted.

Figure 10.21



Completed part of the Northern imaging survey after four years. See caption for Figure 10.20.

The software has been used to carry out a series of long-term strategy simulations based on the assignment of opportunity weights to the SDSS Northern area. This map, shown in Figure 10.19, is based on airmass, availability for observation throughout the year, and assumptions about the amount of time lost to bad weather.

The code is then used to investigate how much of the survey will be completed after a given elapsed time under further assumptions about the minimum scan length which is to be observed, night-to-night weather correlations, maximum allowed airmass during a scan (this is a function of the stripe declination, but the goal is to observe as close to the meridian as possible), and so on. The algorithm works roughly as follows. For a given night, the code checks to see which strips have not yet been observed. It then determines how long a scan can be made for each of these strips, given the sidereal time through the night. "Points" are then awarded to each possible scan according to: length (the longer the better); whether the stripe is next to one already done; whether the other strip of the stripe has been done; the stripe's opportunity weight; and whether the entire length of a strip can be done "tonight". It then selects the strip with the best score. If several strips have similar scores, the strip with the lowest opportunity weight is selected to be observed "tonight". The survey time to completion can then be computed by running the code until all observations have been made, on the assumption that all observations are successful. Two examples are shown in Figures 10.20 and 10.21. These figures show the amount of the imaging survey which has been completed in one year and in four years if the APO weather statistics are as described in Chapter 3 and the weather is uncorrelated from night to night. The simulations show that, depending on how the weights are varied and on assumptions about the weather, the times to completion of the imaging survey (not much shorter than the entire time to completion) range from 3.5 to 5.2 years.

Data Distribution

The intent of this project is to make the survey data available to the astronomical community in a timely fashion. We currently plan to distribute the data from the first two years of the survey no later than two years after it is taken, and the full survey no later than two years after it is finished. The first partial release may or may not be in its final form, depending on our ability to calibrate it fully at the time of the release. The same remarks apply to the release of the full data set, but we expect the calibration effort to be finished before that release.

Products

Basically, the full set of data in the operational databases will be made available for scientific use by members of the collaboration and for public distribution. The data will likely be reformatted, restructured, and compressed from the operational databases. In addition, the imaging data will be further processed to provide the following products:

1. Merged pixel map (8.1 TBy, ~ 3 TBy compressed). There will be tiny regions in which a given point in the sky is covered as many as 6 times, somewhat more at 4 times, more at 3, and quite large areas (nearly 50 percent of the sky, counting all overlaps) covered twice.
2. Atlas images (~ 250 GBy, ~ 80 GBy compressed) This is a subset of the data from the merged pixel map. Note that this set will include objects that span two or more strips. This will require a processing effort which is smaller than, but comparable to, the on-line (or nearly on-line) construction of the strip atlas. It can, however, be done in a leisurely fashion and may be the distribution product of choice.
3. Catalog for objects found in the merged pixel map (25 GBy). Parameters, calibrated sky position, and classification.
4. Variable object list (2 GBy (?)). This is a catalog of the objects that have changed in brightness and/or shape significantly between two strips. We list the times of measurement, and the relevant sets of information from the strip catalog; atlas pictures might be included as well, though that will increase the size considerably.

Distribution Format

The two largest databases, the corrected pixel map and the merged pixel map, are candidates for distribution using wavelet transform compression, whereby the photometric integrity and resolution are preserved at high compressions at the price of not properly recording the noise. Compressions of order 20 are to be had by this technique, which for the merged map (the one of probable interest) reduces to 400 GBy. This, at 700 CD-ROMs, still seems a bit excessive, but there may well be more capacious inexpensive media by the time it exists. The compression technique also has the advantage that there is an accompanying *residual file*, or a set thereof, successive ones of which add a bit to the reconstruction of the original data and all of which reconstruct it with no loss, but the total is, of course, about the same size as the losslessly compressed database. An alternative to this is to publish a `sky map' with the objects (all of which are fully represented in the atlas) excised or subtracted out, at significantly lower resolution. If we bin 4 x 4 and rescale (for better intensity resolution), we have 1.6 arcsecond resolution, and achieve a reduction of a factor of 16 in data volume. In addition, the data are almost all indistinguishable from sky, and should compress by our normal techniques a factor of very nearly 4, so that the compressed data occupies only 120 GBy or 200 CD-ROMS.

A word is in order about the atlases, their sizes and contents. We have done experiments on faint stars and galaxies in other data and on our Hercules cluster simulations for the survey, and have arrived at a subframe size which typically contains all the statistically significant parts of a galaxy; one typically needs a square region

$$S = 100 \times 10^{0.17(16.5-g')}$$

pixels on a side. This is only a guideline, of course, and individual objects will have this dimension individually determined. With the observation that galaxies typically cross the 5:1 signal-to-noise threshold at about g' = 22.5 , we have 5×10^7 galaxies and 7×10^7 stars in the survey region, at that S/N cutoff. There are 1.7×10^{10} pixels associated with these. There are 2 bytes per pixel and five colors, which comes to 170 GBy. It would be prudent to include a surrounding area at lower resolution to allow following any faint features which might exist, and the overhead is only 19% to include the periphery in a box twice as big but averaged over 4 pixels, i.e. 1.6 arcseconds. This brings the total to 200 GBy, about 2.5% of the whole merged pixel map. Another set of estimates have been made recently from the simulation catalogs (Chapter 9) which also give a number which is 3 to 4% of the sky. In all these cases, the criterion is that the region should include an annulus of width about half its radius which is large enough that the ratio of total object signal in it to sky noise is about unity; any larger one clearly does not, on average, contain significant object data. We have rather arbitrarily chosen a value of 3% areal coverage to arrive at the figures in Table 10.2.

The vast majority of these data are, at the one pixel level, indistinguishable from sky, so the lossless compression factor should be between 3 and 4. Thus the compressed total should be about 80 GBy. The atlas of objects found in the scans is about 1.5 times bigger because of the overlaps, and includes important consistency checks in that a large fraction of the objects will have been observed twice.

There is some question about how to handle objects for the atlas which our classifier thinks are stars; one way which seems satisfactory is to store the parameters of the fitted PSF, subtract it, and make a (highly compressible) image of the *residuals* at lower resolution. Since there are more stars than galaxies in the sample, doing this in an efficient way is quite important.

The *catalogs* must be treated very carefully if the amount of data is not to be overwhelmingly large. For all objects, we will want to record positions, at least crude radial profile information in 5 bands from which aperture and isophotal fluxes can be constructed, shape parameters, and pointers to atlas image and calibration information. All told we expect about 250 bytes of information per object, but this may well change as we learn more. One may want to store more for the brighter objects, but their numbers are small and should not increase the total catalog size significantly. The strip scan catalog will be 50% bigger.

It is premature to make firm plans for the form of the public distribution, but we can envisage a multi-level distribution based upon CD-ROMs, which might look something like Table 10.2. In this table, the higher levels correspond to wider anticipated distribution, perhaps with levels <= 0 stored only in the master archive at Fermilab.

Table 10.2: Hierarchy of Data Distribution

Level	Content	Size	Compressed Size	No. CD-ROMs
3	Parameter list of Objects with spectra	750 MBy	400 MBy	1
2	Parameter list of all objects	25 GBy	15 GBy	25
1	Atlas Images	250 GBy	~80 GBy	~130
1	Spectra	50 GBy	25 GBy	40
1	Sky Map	500 GBy	120 GBy	200
0	Flattened 2D Spectroscopic Frames	70 GBy	35 GBy	
-1	Merged Pixel Map	8 TBy	<3 TBy	
-2	Raw data, strips, and scans	12 TBy	<5 TBy	

It must be kept in mind that it may be possible and most convenient by the time the data are ready to maintain a public-access archive over the successor to the Internet, which circumvents the distribution problem completely and allows easy maintenance of software, though we would probably still want to make some software available for investigators who wish to copy some part of the archive to their own systems. Examples of such software might be routines to uncompress an atlas image into a FITS tape image file or other standard image format, to do the same for some area of the merged pixel map, etc. We are planning to look carefully at the database aspects of the survey catalog data (sorting, keys, etc.), but have not decided at this time what level of support we will offer. At the very least we will supply software to allow *reading* the catalogs, which will likely be stored in highly compressed binary form.

The Production System

The production system consists of the individual data processing pipelines integrated together into a cohesive data processing system plus the hardware needed to run that system in routine operation. It also provides the infrastructure needed to allocate resources, schedule and track data processing jobs, store output data sets, and provide mechanisms for dealing with abnormal conditions.

The data flow between all of the major data processing pipelines was shown in Figure 10.3.

The Compute Hardware

The hardware requirements were determined by conducting benchmark tests on the time-intensive sections of pipeline code and by requiring that we be able to turn around a night of imaging or spectroscopic data in 24 hours.

We have purchased three major machines. Two are DEC Alphaserver 8200 5/300 systems. Each has 5 300 MHz processors plus 1 GBy of memory. The third system is an SGI Challenge with 6 150 MHz processors and 512 MBy of memory. We expect to have 300 GBy or more of disk storage distributed among the three systems for the first year of operations, plus several DLT 4000 tape drives. We shall use the two DEC machines for production processing of the imaging and spectroscopic data. The SGI Challenge will hold the object catalog databases and will be used for post-pipeline activities such as quality analysis, merging of object lists, target selection, and distribution of data to collaboration members. In addition, we have access to a hierarchical tape storage system at Fermilab. This system consists of a tape library (maximum capacity of 30 TBy), disk caches, and a file management system. For the first year of operations we plan to utilize 1 - 3 TBy of storage capacity in this system.

All computing systems and the hierarchical tape system will be connected by an FDDI network, which can transfer data at approximately 12 MBy per second.

The major output from the pipelines will consist of corrected frames, the object catalogs, the atlas images, and the 1-D spectra. We expect to hold the object catalogs and 1-D spectra spinning on disk at all times. It is unclear if the atlas images will be kept spinning as well. The corrected frames will be archived to DLT 4000 tapes and will also be transferred to the tape library.

Software

The software needed to operate the production system is not yet developed, since it requires that pipelines be close to their final form. However, the architecture will mimic the standard Fermilab "farms" architecture for processing high energy physics data. The spectroscopic and imaging pipelines, which are the most CPU intensive by far, are designed to process data in discrete, independent "events". For spectroscopy, an "event" consists of all frames pertaining to a single spectroscopic field of 640 fibers. The imaging data from the photometric CCDs is already divided into 6 independent streams, one for each column of CCDs. Each stream is further divided into "fields", each one consisting of 5 frames in the 5 colors of the same area of sky. As an imaging tape is read, each field of 5 frames is assembled and appended with overlapping data from the next field so that each field can be processed as an independent entity. By organizing the data in this fashion, it is possible to have multiple copies of each pipeline running in parallel, with events "farmed" to the next pipeline that becomes idle. At present we expect to stage all outputs from a processing run to disk, so there is no need to require that pipelines deliver their output synchronously (such as would be required if we were writing directly to tape), although such capability could be added if needed.

Operations - Imaging

The following steps will be followed to process imaging data.

Eight tapes are created at APO - 6 Photometric Data tapes from the photometric cameras, 1 2.5M tape with all files produced by the online system (quartiles, star parameters, etc) plus observing log, and 1 MT data tape with all data from the Monitor Telescope.

Tapes are shipped to Fermilab via commercial express carrier

Unpack and label tapes and record in a log

Untar the 2.5M and MT tapes

Create operational database entries for each SDSS imaging run

Create processing plans for the following steps

Process MT data. This step can be done in parallel with the next three

Mount 6 photometric data tapes. Run the SSC pipeline.

Run primary and secondary astrometric calibrations

Review outputs. Create processing plan for photometric data processing

Run postage stamp pipeline

Run photometric frames pipeline on photometric data tapes

Write output corrected images to two sets of tapes and to tape robot.

Write object catalogs to database.

Operations - Target Selection

Select a set of imaging runs. For each run, determine acceptable start and stop limits

Run Merge Objects operation

Distribute objects to Science Database

Select targeting area of sky

Set target selection parameters

Run target candidate selection code

Run tiling

Run remainder of target selection code

Define observing conditions

Run plate design code

Operations - Spectroscopy

One tape will be created at APO per night that contains all data.

Tapes are shipped to Fermilab via commercial express carrier

Unpack and label tapes and record in a log

Untar the tape

Create operational database entries for each spectroscopic field observation

Create processing plans for the following steps

Process data through IRAF to perform 2-D extractions and wavelength calibrations

Process 1-D spectra through spectroscopic pipeline

Write parameters to operational database

Write 2-D frames and 1-D spectra to tape robot

Distribute parameters to Science Database

Monitoring Progress

The software effort is large, and is spread among widely geographically separated institutions; it is a challenge to track the progress of the software on a routine basis and to coordinate all of the activities and people involved. Further, as noted above, it is via this software development that much of the astronomical expertise of the collaboration scientists is incorporated into the survey design. Accordingly, we have set up a series of regular meetings and reviews to track and plan the effort.

Pipeline Coordinators

The development of each major pipeline is coordinated and managed by a responsible individual at the institution. Generally, the pipeline developers hold regular weekly meetings.

The "Working Group" Group

This group consists of the pipeline coordinators and the scientific working group chairs, and meets at Fermilab for two days at intervals of 4-6 weeks.

The Weekly Software Conference

This is held every week at Fermilab with the pipeline coordinators participating by telephone hookup, and focuses on Fermilab software issues, coding standards, framework development etc.

Reviews

The individual software systems, the integrated software, and the plan for the data acquisition hardware have all received major reviews involving outside participants.

Electronic Archives

The minutes of all meetings are archived, as are the software requirements documents. In addition, the SDSS has set up a large number of email exploders, maintained and archived at Princeton. These were originally set up to make general announcements and to track progress and milestones in the development of the individual pipelines. Since then further mailing lists were set up for the discussion of scientific issues and progress, and they have expanded yet further to discuss work on several pieces of hardware.

References

Collins, C.A., Guzzo, L., Nichol, R.C., and Lumsden, S.L. 1995, MNRAS 274, 1071.

Connolly, A.J., Szalay, A.S., Bershady, M.A., Kinney, A.L., and Calzetti, D. 1995, AJ 110, 1071.

de Vaucouleurs, G. 1948, *Ann. d'Astroph.* **11**, 247.

Fukugita, M., Ichikawa, T., Gunn, J.E., Doi, M., and Shimasaku, K. 1996, AJ 111, 1748.

Heavens, A.F. 1993, MNRAS 263, 735.

Kepner, J.V., Fan, X., Bahcall, N.A., Gunn, J.E., and Lupton, R.H. 1997, in preparation.

Lupton, R.H. 1993-1996, "Image Processing in the SDSS: Algorithms", SDSS Internal Documents.

Press, W.H., and Rybicki, G.B. 1993, in `Time Series Prediction: Forecasting the the Future and Understanding the Past', eds. A.S. Weigand and N.A. Gershenfield, SFI Studies in the Science of Complexity, Proc. Vol. XV (Addison-Wesley), 493.

Richards, G. 1996, SDSS Internal Memorandum.

Rumbaugh, J., Blaha, M., Premerlani, W., Eddy, F., and Lorensen, W. 1991, `Object Oriented Modeling and Design', (Prentice-Hall, NJ).

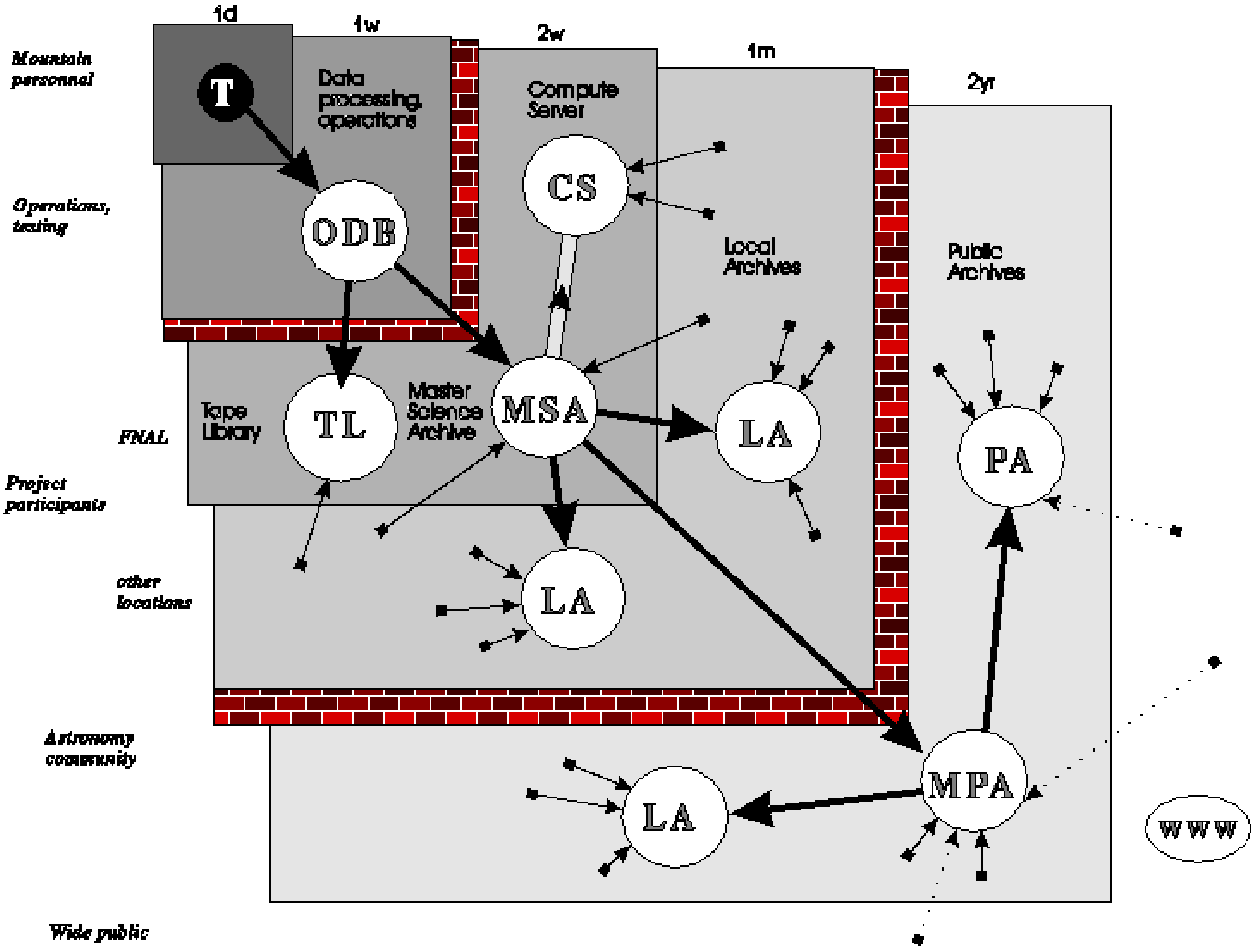
Tonry, J., and Davis, M. 1979, AJ 84, 1511.

General Objectives for The Science Archive

Survey Data Flow -- the Archive Perspective

As discussed in Chapter 10, the SDSS will generate about 40 TB of raw data at the telescope site, which will be shipped to Fermilab for reduction. After processing, the reduced data will be stored in the Operational Archive (OA) which supports the mission critical part of the survey. Its main functionality is to store the processed data in instrumental form, perform calibrations, and provide information for target selection and quality assurance. It will contain positions, colors, sizes, profiles, classification, calibration information, targeting information, housekeeping, logs, history, "atlas images" in five colors, spectra, spectral features and redshift.

Figure 11.1



Data-flow diagram showing the various levels of data access and timescales. The labels on the left hand side describe the typical user, those on the top the timescale on which data will propagate to that level. The 40TB of raw data will be processed by a workstation farm at Fermilab, saved in the Operational Data Base, and can only be accessed by the operations personnel. To provide timely access for science testing and analysis, the data will be moved over to the Master Science Archive every one or two weeks.

What is the Science Archive?

For subsequent science queries the requirements are quite different: scientists expect to see a final calibrated data set, where they want to perform statistical queries of quite complex nature. Given the huge volume and complex nature of the data set, this is quite a demanding task. We are currently designing the Science Archive (SA) to support these activities. The Science Archive consists of the Master Science Archive, which resides at Fermilab, and derived replications at the collaboration member sites. The Science Archive will also serve as a pilot model for the subsequent public distribution system. The Science Archive is read-only and contains calibrated quantities only, but all housekeeping information, including calibration coefficients, is retained. One of the major design principles has been to build a system with maximum modularity and flexibility, since the data are expected to have a useful lifetime much beyond that of the people currently creating the archive. We also expect that this data set will become a de facto standard for the next century, with all other astronomical catalogs, both existing and future, cross-referenced against it.

The primary purpose of the Science Archive is to provide convenient, high speed access to the data. It can be split into multiple data products:

1. the object catalog, containing positions, magnitudes, colors, sizes, radial profiles, classifications, etc. for over 100 million objects
2. housekeeping data, containing calibrations and log information
3. atlas images in 5 colors for all identified objects
4. one-dimensional spectra of all spectroscopic targets

The Object Catalog will have a built-in multidimensional index, which enables a very quick search on all linear combinations of the 5 colors, a very flexible set of constraints involving the different angular coordinate systems on the sky, and angular proximity. Objects with similar properties will be stored together on the disk, achieving a very high level of cache efficiency when executing queries.

The Science Archive will use a three-tiered approach for maximum flexibility and portability. At the lowest level (Data Warehouse), the system will be built upon a commercially available object-oriented database engine, Objectivity, supported on a wide range of hardware platforms, and complying with several existing (SQL) and emerging (OQL, ODMG93) standards. The middle-ware will contain all the proprietary indexing and query optimizing software, written in strict C++, while the top level component provides a high level user interface, protecting the user from the details of how the data are organized in the lower layers. Our approach will also enable us to use a highly parallel database engine, should it become available.

The Nature of the Problem

Although substantial progress has been made in combining multi-frequency astronomical data sets (Helou et al. 1992, Cordova 1993), these approaches provide only limited scientific inquiry into the data; the size of data bases and the number of queries is causing substantial loads on database servers. Moreover, searches are generally restricted to simple positional queries, while with the multicolor nature of the SDSS dataset one might really wish to ask more complicated questions, e.g. "find all blue ($g' - r' < 0.7^m$) galaxies fainter than $g' = 22^m$ and within 3 arcseconds of a quasar brighter than $r'=17$." Such a query can be modeled quite succinctly as a geometric query to find all data points within a given distance of a specified simplex or hyperplane in the multidimensional space of galaxy colors. Such search techniques represent the cutting edge of research in computer science are not commercially available yet.

Geometric View of the Data

We have therefore undertaken a research program directed at the development and implementation of advanced data organization and querying techniques using the methodologies of object-oriented design and computational geometry. Our data organization methods are based upon a novel use of space partitioning schemes from computational geometry, such as *k-d* trees, to provide a mechanism that is effectively equivalent to a general multi-dimensional index, with typical access times growing only logarithmically with the number of data points. In addition, by adding a simple interactive feedback loop which utilizes the coarse grained density map that space partitioning provides, our approach allows simple and accurate prediction of query times and output size. A lot of queries -- especially those from novice users -- are unnecessarily broad, wasting substantial resources. By predicting query times and output volume (in a matter of seconds) users can adjust their search criteria or acknowledge and confirm before an extensive query is performed. This substantially reduces the load on servers, especially in the exploratory stage, while keeping the interaction between the user and the database at a maximum.

Current datasets are either big or complex, but seldom both. The next generation datasets are emerging as both big (many Terabytes) AND complex (spatial, multiwavelength). This poses several bottlenecks: the database size presents I/O problems, and the database complexity makes it hard to extract general features to detect similarities, i.e. associative relations. The approach we propose and describe in detail below enables us to solve both problems simultaneously: we can execute queries in parallel, and provide the database with the ability to search for similarities (spatial proximity) and to learn from examples.

Here we present a detailed description of these ideas and how we envisage their potential use and application with an object oriented database management system to solve the problem that the size and complex nature of the SDSS data set represents. We discuss the envisaged usage patterns, the underlying design principles of our system, the geometric organization of the data in the archive, the geometric query capabilities, and the system architecture.

Envisaged Usage Patterns

Typical Access Modes

Let us first discuss the phases of a typical search project that most scientists accessing the Archive will undertake. Even though this may seem trivial, we will see later how the support requirements of these activities naturally translate into requirements for the Science Archive.

1. Make plan
2. Explore
3. Test queries on small scale
4. Do large scale search
5. Process output
6. Draw conclusions

Such activities involve the use of several search techniques, all of which have to be supported by the Science Archive.

Object Browsing

Efficient Object Browsing requires access to parameters, the atlas images, the spectra, but only one field at a time. The image of a field can be reconstructed from the cut-out atlas images. During this phase of exploration the researcher will need access to various simple tools like xy plot, 1D-2D histograms, and calculating aggregates. A typical load time of 3-5 seconds to access a frame is acceptable. Since the images are stored on an object-by-object basis, this enables us to have an especially useful image display. Objects can be switched on/off individually, and we can thus show only the galaxies, stars, or QSOs. We can tag objects which have spectra, or satisfy certain conditions, and have their properties piped into external tools, for example IRAF, SM, IDL, and DERVISH/SHIVA. The log file for such manual browsing can be saved, and converted into a query for subsequent use.

Sweeping searches

These form the hardest problem, since the data set we are dealing with is close to a TB in size. It requires a high bandwidth to the data, efficient output handling, and good multi-dimensional indexing. The queries will use constraints on not just attributes but algorithmic methods, as well as linear combinations of attributes, and they have to be able to follow links to housekeeping or to spectral information. The implementation has to be able to make a lot of trial-and-error possible, by providing rapid feedback. It has to be highly customizable, and to use a standard query grammar (SQL/OQL). The system should be able to execute queries remotely in parallel, which can also be interrupted.

Proximity of Two Classes

These searches are also non-trivial to execute in a traditional relational system. They consist of two queries and a distance, like our example of quasars near faint galaxies. The representation of the output can be quite complex, since there may be a lot of many-to-many associations. Also, the query optimization can be non-trivial.

Cross-identification

We anticipate that the SDSS data set will be very heavily used for cross identification against all existing and future catalogs. Even though the schema of the database (class design) is still under construction, we have tried to address the issue of cross-references, and/or cross identifications rather early. Reserving a slot in the schema for all other cross-identifications is not feasible, because one cannot foresee all the applications so far in advance.

Creating Personal Subsets

Almost every scientist working with the archive will want to save one or more customized subsets. Since these can be quite large, in the 100 GB+ range, this application needs to be addressed early on. If the data sets are small, flat output should be easy, and the data can be used off-line. If the data size is medium, data should not be copied out of the Archive, but one should save links to the objects, leave the data in place, and use the data within the archive. If the subset is large, i.e. comparable to the whole data set, one cannot even save the links to the objects, and the only way we currently envisage to retain information is to save the query script itself.

Creating New Data Products

Eventually, some of these personalized subsets will evolve into new data products, which will become standards, within the project and beyond. Examples of these are clusters of galaxies and their members, a catalogue of galaxy and quasar absorption lines, variable stars, etc.

Design Principles

The data archiving scheme that we propose to implement is based on four design tenets: Object Oriented Programming (OOP), Distributed Processing, Spatial Data Structures, and Hierarchical Memory. Before we discuss the basic algorithmic components of our system, these four principles should be explored.

Object Oriented Philosophy

Historically, astronomical databases have relied upon relational database technology. A straightforward mapping existed between the `flat' data and the tabular records of the relational database. These records consisted of fields, which are fundamental data types, such as INT or CHAR. The querying mechanism was provided by a standard query language (SQL) augmented by index tables generated according to predefined key fields (e.g. position or magnitude). More recently these data types have been extended to include more complex data structures (e.g. multiwavelength images, spectra, dynamic length arrays) which are less suited to a tabular form.

In direct contrast to this is the Object Oriented Paradigm. This approach utilizes the fundamental relationships within the data, treating them as individual units rather than collections of attributes (e.g. a galaxy would be considered as an object whose schema might include coordinates, magnitudes and associated bitmap images and spectra). An object is, therefore, simply accessed as another datatype (equivalent to an INT or CHAR). How it is actually stored on a disk is hidden from the user. This allows efficient storage of complex data (positions can be stored as integers and returned as floats).

Implementing this paradigm within a database system, an OODBMS, adds the benefit of persistence. Accessing an object as a whole rather than as a collection of tabular attributes increases the speed of many queries by factors exceeding 10. A further benefit of using OODBMS is the ability to execute code (methods) on the server side. This is advantageous since the attributes of an object are extracted via method functions, allowing for a dynamic update of data without rebuilding the database. Thus, if the photometry or astrometry is recalibrated for a released catalog, only the access functions must be modified, saving a costly reissue of hundreds of CD-ROMs. In addition, the development of an Object Query Language (OQL), a superset of SQL, provides browsing tools and enhanced query capabilities. Given the potential of this approach for handling large multi-dimensional databases its evaluation and development in an astronomical environment is clearly necessary for the efficient handling of current and future data sets.

Distributed Processing

A client-server architecture provides the ability to distribute the processing load over multiple computers. A well defined protocol (TCP/IP) exists for the interprocess communication that enables the processes to be distributed in distance. The major design benefits of such a distributed processing system are flexibility, component isolation, and future expandability.

The flexibility of a client-server system arises from the separation of processing tasks. Processes can be optimized for a given CPU or operating environment. In addition, the processing load can be distributed in a redundant fashion that can be optimized to avoid congested archive sites when processing queries. Since each process is isolated from the rest of the system, commercial products can be easily integrated with minimal impact on the rest of the system while maintaining cross platform connectivity.

The client-server approach is easily expandable through the addition of more processing sites or additional archive databases through a notification action to the appropriate component processes in the system. The data retrieval can be expanded to a parallel I/O operation through the addition of multiple archive servers and media striping. This approach is easily adapted to interface to a massively parallel I/O machine, eliminating any code rewriting.

Spatial Data Structures

With the increase in the volumes of data sets how we partition these multidimensional data on the storage medium can severely affect our ability to conduct efficient searches. Our particular algorithm for splitting up a `k' dimensional volume is based upon the heuristic technique of *k-d* trees in which search times scale logarithmically with the number of data points (Friedman et al. 1977). This approach is becoming increasingly popular in computer science, especially in databases related to spatial structures (Samet 1989a,b). They find a special niche in the geographical information systems (GIS), a popular, rapidly growing area, quite similar in the nature of its queries to astrophysics.

The *k-d* tree is a multidimensional space partitioning scheme in which `$d$' dimensional data is split in `k' dimensions. Each tree node represents a subvolume of the parameter space, with the root node containing the entire `$k$' dimensional volume spanned by the data. A balanced binary tree is constructed by splitting each node into two sections along the median of the `k' dimension which will maximize the clustering of data in the two child nodes. The `k' dimensions can be actual attributes or the principal components of the parameter data.

Hierarchical Memory

Something that is often overlooked when algorithms are designed is that when they are implemented in practice they must contend with the fact that real computers do not have unlimited memory, as is the case in the theoretical RAM model. This is especially important for scientific computations that use very large data sets, where the number of objects in the data set can greatly exceed the space available in main memory on the computer (even a supercomputer). Thus, algorithms should be designed for fundamental problems that arise in our applications, so as to take into account the fact that the structures used by the algorithm may not completely fit in main memory. Instead, the memory space occupied by these structures must be partitioned into *blocks*, and these blocks must be periodically swapped in and out of main memory as the computation proceeds. One can, of course, rely on the operating system to decide how to partition the data and how to perform the swapping of blocks. However, one should be able to do better by a more explicit specification. Moreover, we believe that one can modify the algorithm so that it takes the blocking into account and performs better than would otherwise be possible when the blocking and swapping procedures are ignored (which is currently the case in most algorithm design). We have already done some of the foundational research for fundamental computational geometry problems, and we intend to next examine new blocking techniques for geometric databases and space-partitioning algorithms.

In the case of our space partitioning scheme, the number of levels in the tree is determined by maximizing the number of objects in a leaf node (container) with the constraint that the *k-d* tree must be small enough to reside in memory. The maximum cell size depends on the amount of available memory, disk partitions, total database size, and I/O bandwidth. The desire is to minimize I/O operations by pulling over a large enough set of data in one operation to minimize search times. Opposing this is the increased speed of a larger *k-d* tree. One method of circumventing this is to introduce the concept of a *container*, which could hold multiple cells. Thus a search request would be applied to the appropriate containers and the appropriate cells could then be extracted. This allows the *k-d* tree to be deeper, while minimizing I/O. This has the additional benefit of allowing a queuing approach to multiple searches of the same volume within the database.

A Geometric Approach

Geometric Storage and Query Prediction

Most large astrophysical data sets consist of points in a multidimensional space, linked with additional objects (images, spectra). Traditional indexing techniques are very efficient if the queries are by predetermined keys. However, with many data sets in astrophysics the nature of the queries is different, based upon a complex set of criteria, some of which will be similarity of properties, or spatial proximity of certain types of objects. Therefore, the spatial relations between the points in the multidimensional space have a very important role in all of these queries. If most of our searches are localized in `*k*' dimensional space, one can use these relations very effectively -- there is no need to search the regions that one can easily reject.

Here we propose that the data be organized hierarchically, split into `cells' (or `buckets') covering the `*k*' dimensional space, such that points near to one another (i.e. of similar properties) are stored together even on the physical media (in `containers') Optimally, the cells should be balanced, i.e. they should contain roughly an equal number of objects.

Since we store with each node its actual boundaries in the `*k*' dimensions, the *k-d* tree can also serve as a coarse grained density map of the actual data parameters, as well as an efficient multidimensional index. Queries are first performed on the *k-d* tree, limiting the query volume to those leaf nodes that fall within the query, prior to the entire database, resulting in predictions for the search time and estimated number of objects satisfying the search. This provides a feedback feature for the user, which can minimize costly queries. Since the *k-d* tree resides completely within the memory of the user's system, the feedback should be instantaneous. In addition, a list of the cells which intersect the query volume is also generated.

Geometric Partitioning

The actual dimensions chosen for partitioning could be the natural dimensions of the dataset, i.e. the fields of the database, or they could be a derived basis set that takes advantage of intrinsic correlations within the data. Choosing one of the natural dimensions on which to partition would be done by calculating an appropriate statistic, such as the variance, for all of the `*k*' dimensions, and splitting on the dimension with the greatest value. The partition value is then chosen to be the median value along the partition dimension. On the other hand, a derived basis might be calculated from a principal component method, which, though CPU intensive, may pay off in subsequent searches. This type of approach would provide an additional tool for serendipitous searches, since strange objects would reside away from the bulk of the data.

A more generalized method of partitioning the data can be determined algorithmically in terms of a cost function. At every node to be split, the field with the lowest associated cost will be chosen as the dimension to partition on. In this way our *a priori* knowledge of the most common search criteria is used to weight the partitioning algorithm (e.g. the primary source of queries may be positional, therefore we may wish to force an initial cut in declination). To ensure an even split in the available dimensions the cost function will be dynamical (i.e. when we split in a given dimension we can increase its cost). The first subdivision along any dimension will be to split off those data with indefinite values along that axis. This INDEF subset should not be divided further in that dimension: its cost will be infinite.

Maintenance of the *k-d* tree would be necessary as new data are added into the data base, whether as new fields (such as observations at different wavelengths) or improved data, or if cell boundaries need to be moved due to a major recalibration of the data (e.g. new astrometry coming from Hipparcos). This should be accomplished by modifying the *k-d* tree, either as a whole or, preferentially, via some of the sub-branches.

A further refinement comes in the storage of the data within the buckets. The most naive approach would be a linked list of data points sorted in spatial coordinates. Increased search efficiencies can be accomplished by an intelligent organization of the data within each bucket (Samet 1990a,b). Several methods under consideration are bit interleaf, hashing, standard indexing, and even a continued *k-d* tree.

Geometric Querying

Algorithms for searching can be visualized in a geometric manner as finding all data points within some search volume. If the number of dimensions in the search is less than the dimensionality of the database, the extra dimensions are projected down onto the hyperplane of the search request. Geometric searches, such as nearest neighbor or all objects within a metric distance from a given object, can be visualized as finding all points within the search ellipsoid or polyhedron.

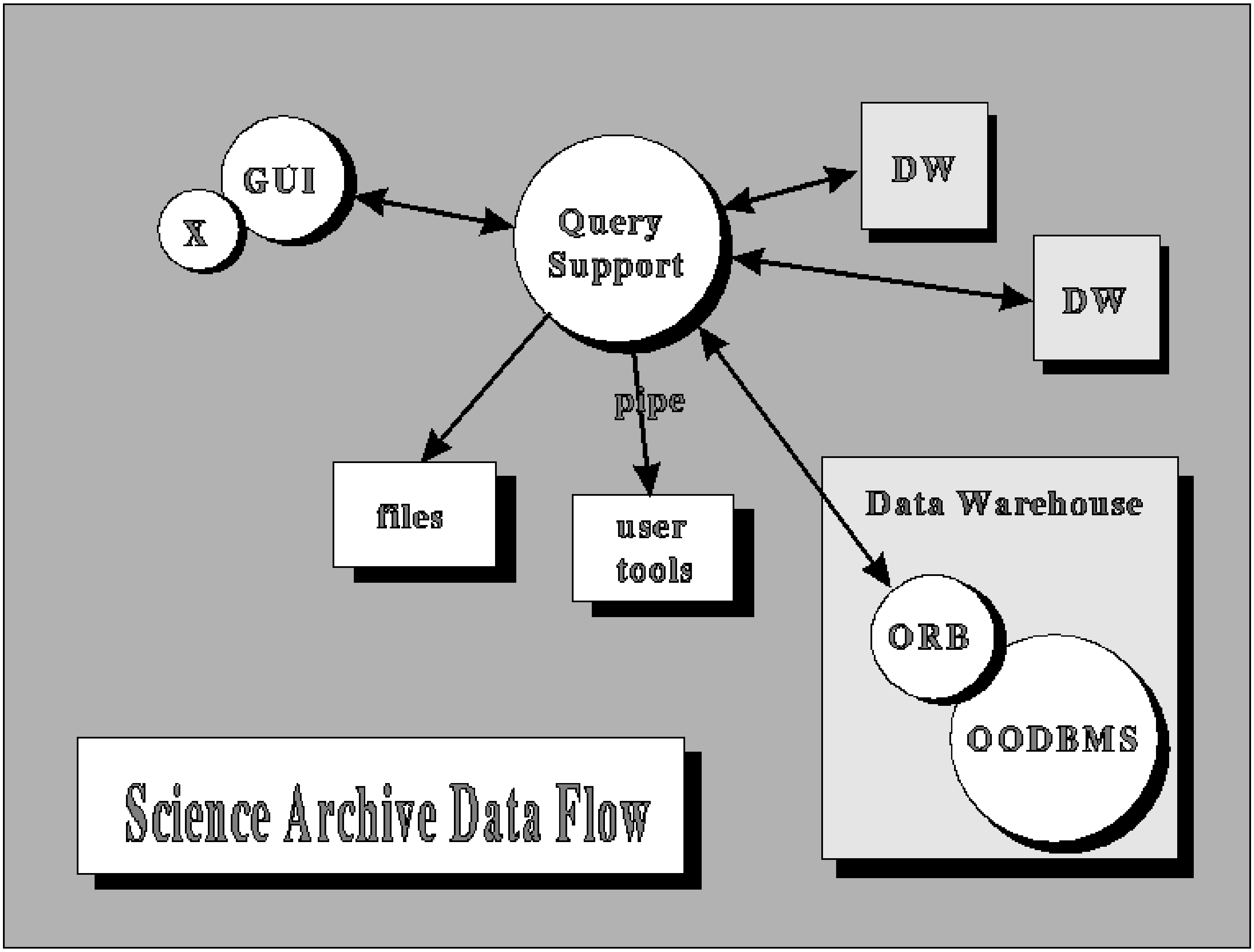
Simple database queries can be modeled as cuts -- hyperplanes -- within the parameter space of the database. These parameter cuts can be combined using Boolean operators to create complicated query volumes. Geometric queries can be constructed through linear combinations of the parameters, and proximity cuts. Curved decision surfaces, separating different classes of objects, can be approximated by Boolean combinations of hyperplanes i.e. multidimensional polyhedra. This approach also allows associative queries to be performed, by first constructing a convex hull from a given sample of objects (training set), and then finding all other objects in the database that lie within this hull.

These geometric queries utilize the nodal boundaries to determine the containers that intersect the query volume and must be searched in more detail. One of the basic operations required is a quick rejection test of the rectangular cells to eliminate those cells that do not intersect the query volume. For those cells not rejected, the volume fraction of the intersection between the cell and the query volume must be estimated. These operations are typical examples of advances in computational geometry that are directly related to this work.

System Architecture

The system we have implemented can be subdivided into three different tiers: the User Interface, the Query Support Layer (QSL), and the Data Warehouse. Communication between the different tiers is accomplished through TCP/IP socket calls, enabling each tier to operate independently on optimized systems in different locations.

Figure 11.2



Data flow through the Science Archive

User Interface

The topmost tier is the User Interface (both Graphic and Command Line Interface are supportable) where the query is constructed, the desired attributes selected, and the returned data processed. The user interface, which generates queries in a predefined syntax, would pass the request to the QSL, where it would be captured, and the search time and output volume evaluated. A small window would open, showing these quantities. A typical exploratory trial-and-error user would reduce the search volume by modifying the query, until an acceptable response time is reached (e.g. a few minutes), then explore the objects that the search provided. If satisfied, the limits would be slowly increased, and eventually a full search be done, but only after it has been ascertained that the output is along the lines of the user's intentions.

Several more powerful querying methods can be applied by utilizing the geometric querying capabilities of the system with the complexities primarily hidden from the user. One such method to access data and search by similarity would be by displaying a bitmap or a graph, where by clicking on certain objects, their properties would be loaded, through a hypertext link, into the center of the search cuts with some error bounds ("*select all similar objects*"). In a given area of the sky one would immediately see how many objects of similar properties (e.g. *colors*...) would there be, and one can again shrink or expand the search easily at wish. An example of associative querying would be to find the `*k*' nearest neighbors in a subspace (i.e. colors only). An example of supervised learning would involve the construction of a training set from a given sample of objects, which can then be applied to the entire dataset "*Select all objects which match the pattern generated from the training set* ". Finally, learning can be unsupervised, producing a complex decision surface which is then automatically translated into our query syntax.

Query Support Layer

The query support layer contains the coarse grained density map of the points in `*k*' dimensions, represented as a binary tree. For every cell we store the boundaries of the cells, the number of data points within it, and a reference to the location of the data. This cell-map is small enough to fit entirely in the memory of a typical workstation. Searches are first performed on this database, generating a linked list of cells that intersect with the search volume. At this point, without opening any of the containers (or cells), one can predict the volume and the time of the query. The search time can be predicted from the I/O bandwidth and seek times of the various storage media. The number of returned objects can be computed by scaling the total number of objects in a cell by the fraction of the cell that lies within the query volume. This information can then be returned to the user interface layer, where the query can be modified until an acceptable compromise is found. Each iteration can be accomplished within a few seconds.

Once the query is finalized, the QSL sends the query tree and list of cells that need to be searched to the appropriate data warehouse. This architecture can be easily parallelized if the data warehouse is composed of multiple servers: the list of cells is sent in subsets to the appropriate servers and then processed concurrently. The QSL can also utilize multiple data warehouses in different physical locations to minimize the query times: large complex queries may be best executed on a remote parallel server even with the network penalty, while simpler queries may be better done on a local server. Such resource allocation and sharing can be transparent to the user.

Data Warehouse

In our scheme, the data warehouse includes the actual data stored in the OODBMS, and an Object Request Broker (ORB) that then isolates the commercial software from the rest of the system. The ORB receives a query tree and a list of containers that intersect the query volume. The ORB translates the query tree into an OQL request to the database, which then returns an iterator containing the objects satisfying the OQL request. Any remaining query functions that could not be represented in the OQL syntax are then used to further pare down the object list. Once the final list is produced, the necessary attributes are extracted and passed back to the user via the QSL.

The extraction of the data from the OODBMS can be done in a highly optimized fashion; while we are searching through the contents of one cell in memory, the next cell is prefetched, utilizing the available I/O. The other advantage is that of customization: frequently accessed cells can be moved to faster physical storage, statically, or even dynamically, implementing a simple cacheing scheme by copying the contents of a cell to hard disk from a CD-ROM or optical disk, for example, and updating the pointer in the cell-database. One can also implement a queuing scheme: if the access time of containers is long, it may be worth queuing requests until several searches are performed on the contents of cells in the same container.

Cross-identifications with the SDSS Science Archive

Creating a cross reference

We propose to build into the structure of every object in the catalog a `cross-reference hook', a link (pointer) to a variable length object, set to NULL in the beginning, thus not taking up much space. If there is a new cross reference, we create a new cross- identification (XID) object in the database, which has a link to the SDSS object, but also has a unique pointer (catalog name, identification, etc.) to the external object. All these XID objects will be derived from the same base class, therefore their customization will require very little modification and extra programming of the original database. All these objects can be kept in a well defined separate part of the archive, which can be updated more frequently, as new cross-references are added. Through this indirection, the original catalogs can remain static, with only the XID objects modified.

Cross-identification tools

What are the steps of creating cross-identifications between various catalogs, with very different errors, wavelength coverage, or positional accuracy? Clearly, every external catalog will have a lot of different systematics, and certain parts of the cross-identification process will be unique to those. On the other hand there will be a lot of common steps as well, inherently tied to the SDSS Archive. Here we would like to identify these, and we propose to create the necessary tools as part of our database and archive development.

1. It clearly requires a rapid search of the angular vicinity of every object to be cross-referenced. Our archival system will have a highly efficient proximity-search mechanism. We will create a tool for the first phase of the cross referencing, which will take a list of objects, with positions, search radii and constraints on the SDSS objects (star, galaxy or QSO, or a color or magnitude range), and for each of the external objects will create a list of pointers (references) into the SDSS catalog, with all the possible candidates.

2. There will be multiple candidates within the possible error-box. We need a mechanism to assign a probability for the match to each candidate, based upon our prior astrophysical knowledge, then to choose the most likely candidate. The colors and profile parameters will be extremely useful for such a purpose, although one can easily envisage cases where even the atlas images or the spectra play a role. Such an assignment can be in the form of a likelihood function, computed from the colors and from the properties of the external object itself (correlation of IR fluxes with our i' band, etc.). One can also envisage a Bayesian classifier. This tool has to be created for every external catalog separately; we would provide a template.

3. Once these two steps have been performed, the results must be verified, then the XID objects can be checked into the Master Science Archive. This involves an update of the schema (since XID objects with possibly new properties will have to be created), the check-in of the new objects into the XID segment of the Archive, and creation of the appropriate bi-directional references back to the SDSS objects.

References

Cordova, F.A.-D. "MultiWaveLink: an interactive data base for the coordination of multiwavelength and multifacility observations" Adv. Space Res., 1993.

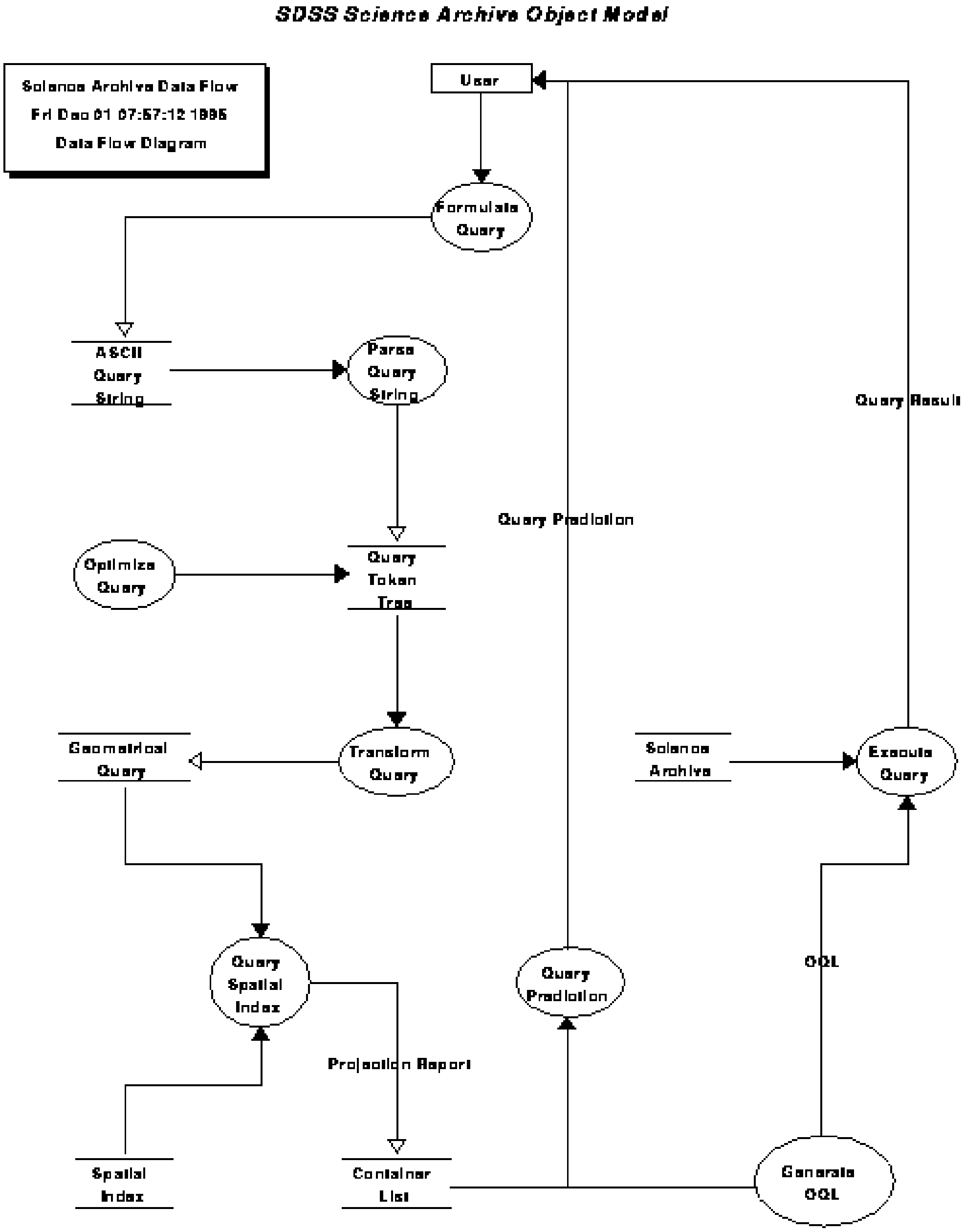
Friedman, J.H., Bentley, J.L. and Finkel, R.A., 1977, *ACM Transactions on Mathematical Software*, **3**, 209.

Helou, G., Madore, B., Schmitz, M., Corwin, H.G., Jr., Wu, X., Bennett, J., Lague, C. 1992, "NASA/IPAC Extragalactic Database", in *Morphology of Galaxies: Nature or Nurture?, 1992 Moriond Meeting*.

Samet, H. 1990a. *Applications of Spatial Data Structures, Addison-Wesley.*

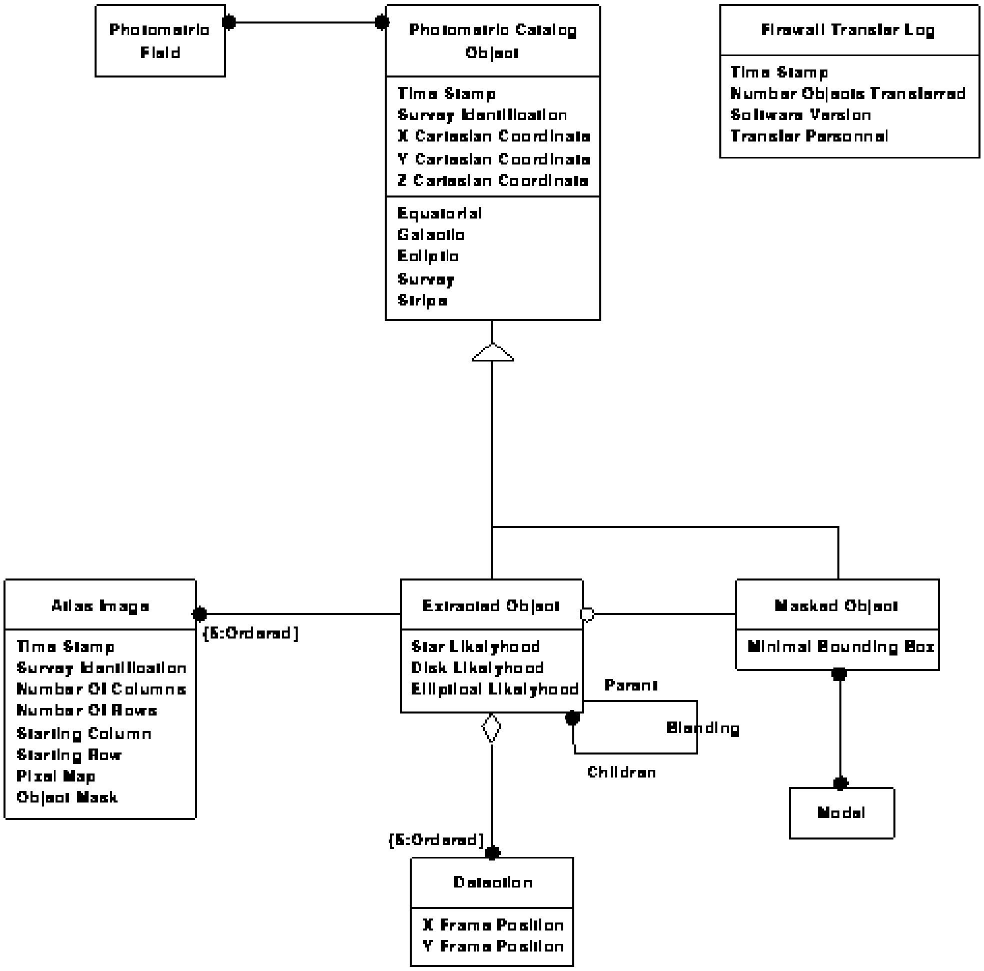
Samet, H. 1990b. *The Design and Analysis of Spatial Data Structures, Addison-Wesley.*

Figure 11.3



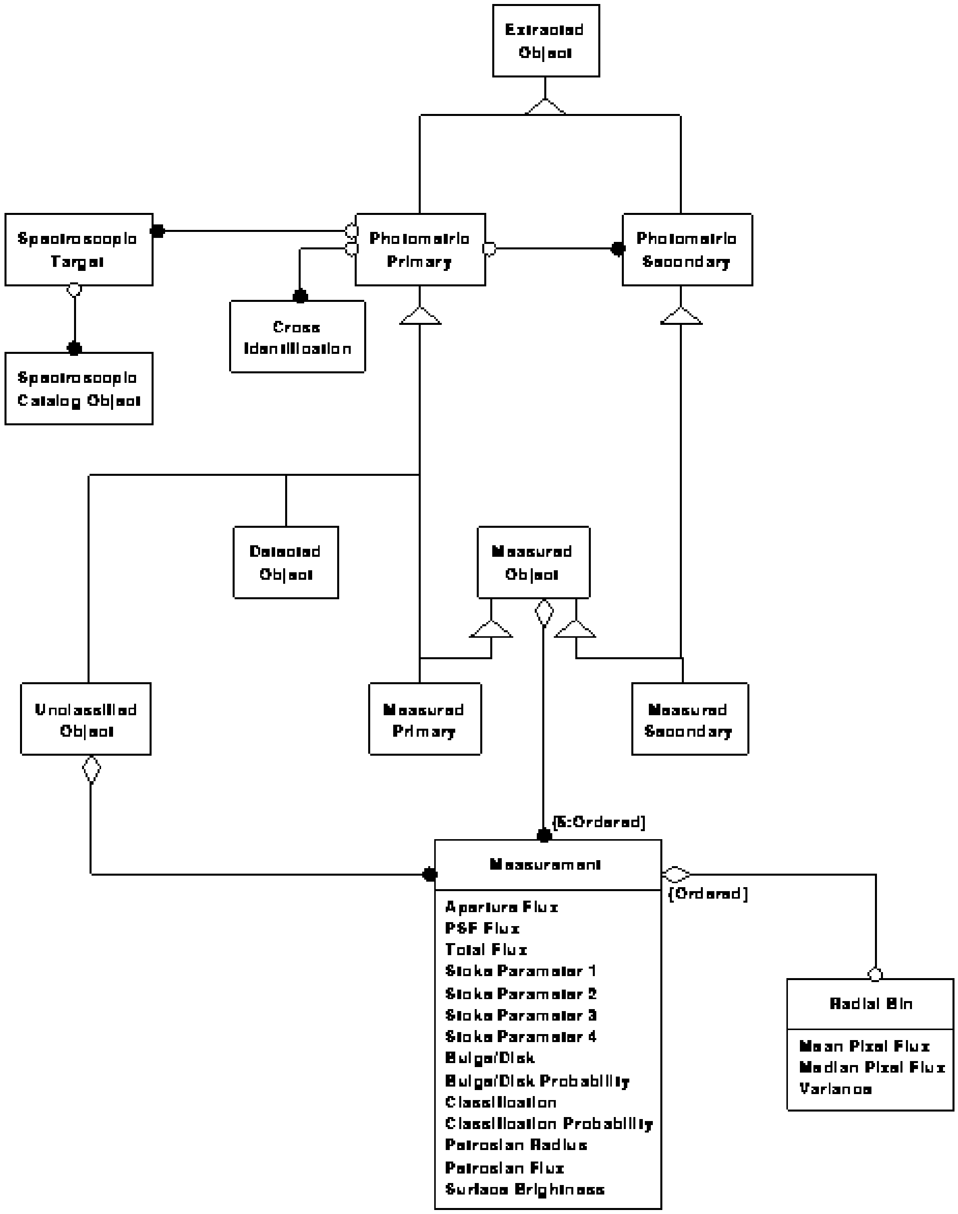
Object model for the science archive

Figure 11.4



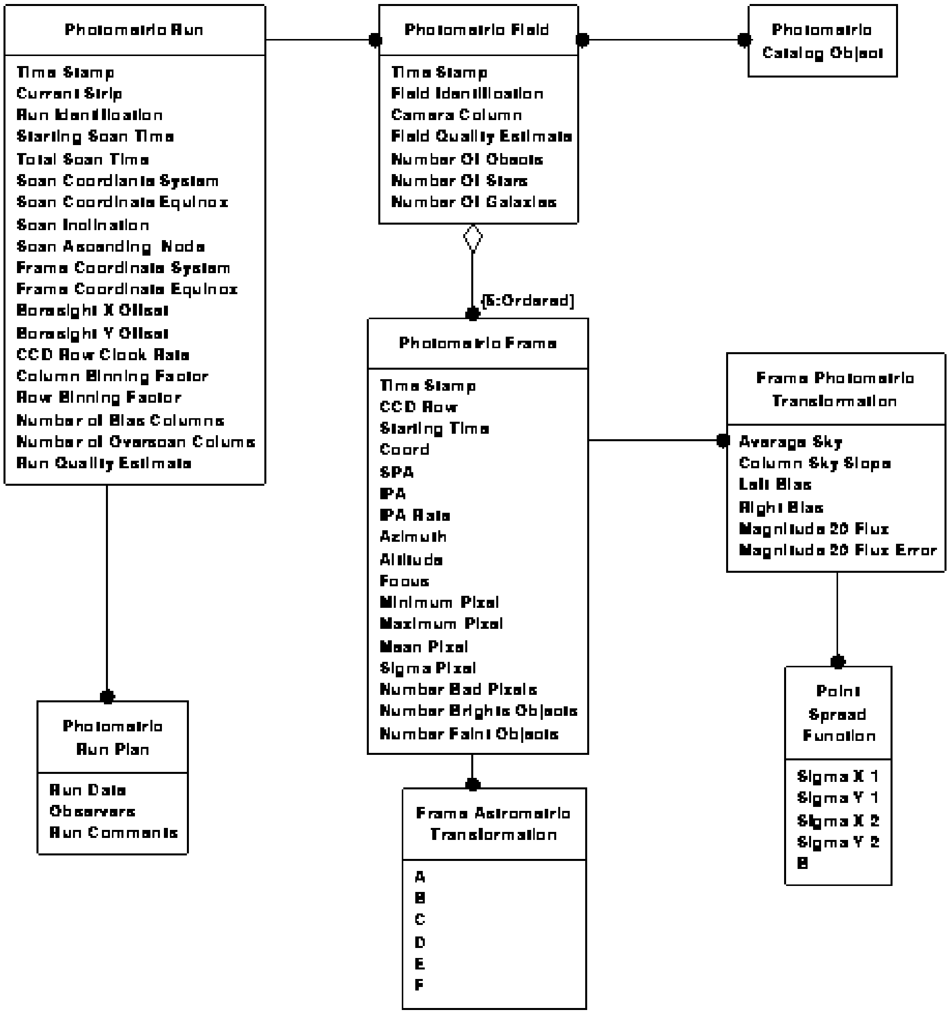
Photometric data model 1

Figure 11.5



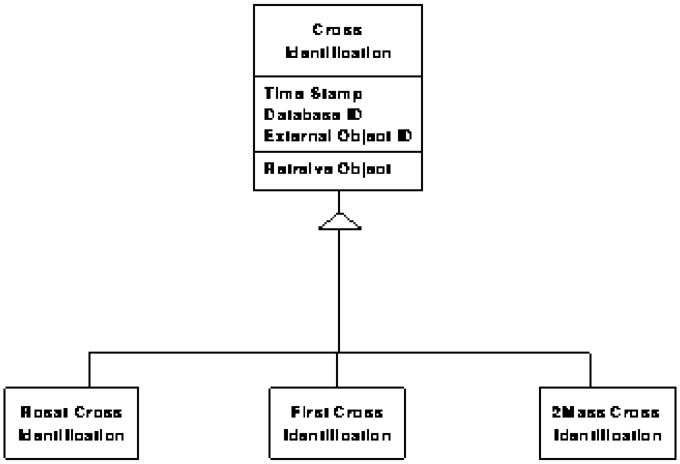
Photometric data model 2

Figure 11.6



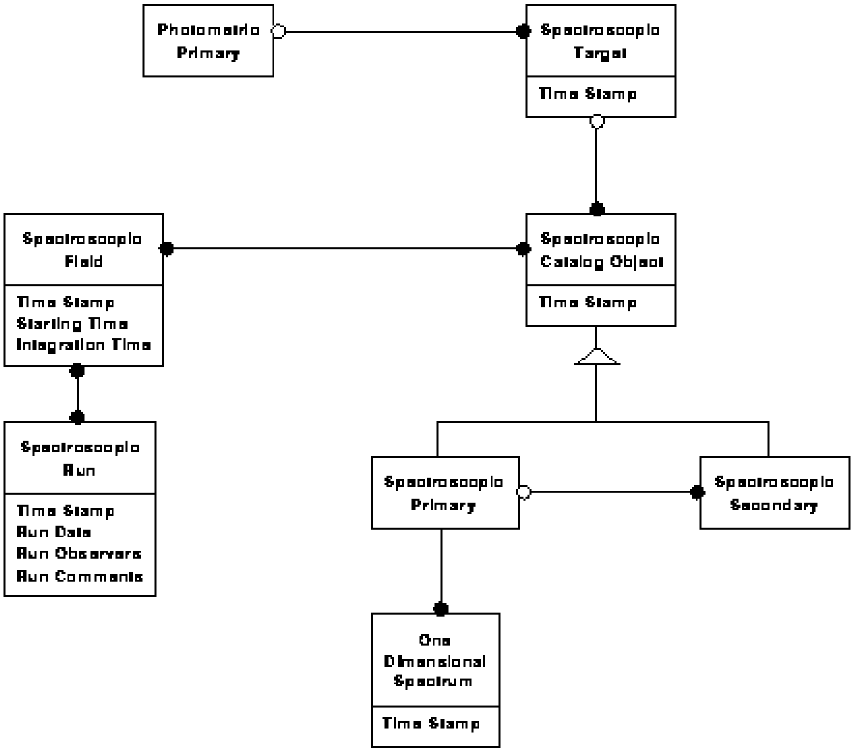
Photometric data model 3

Figure 11.7



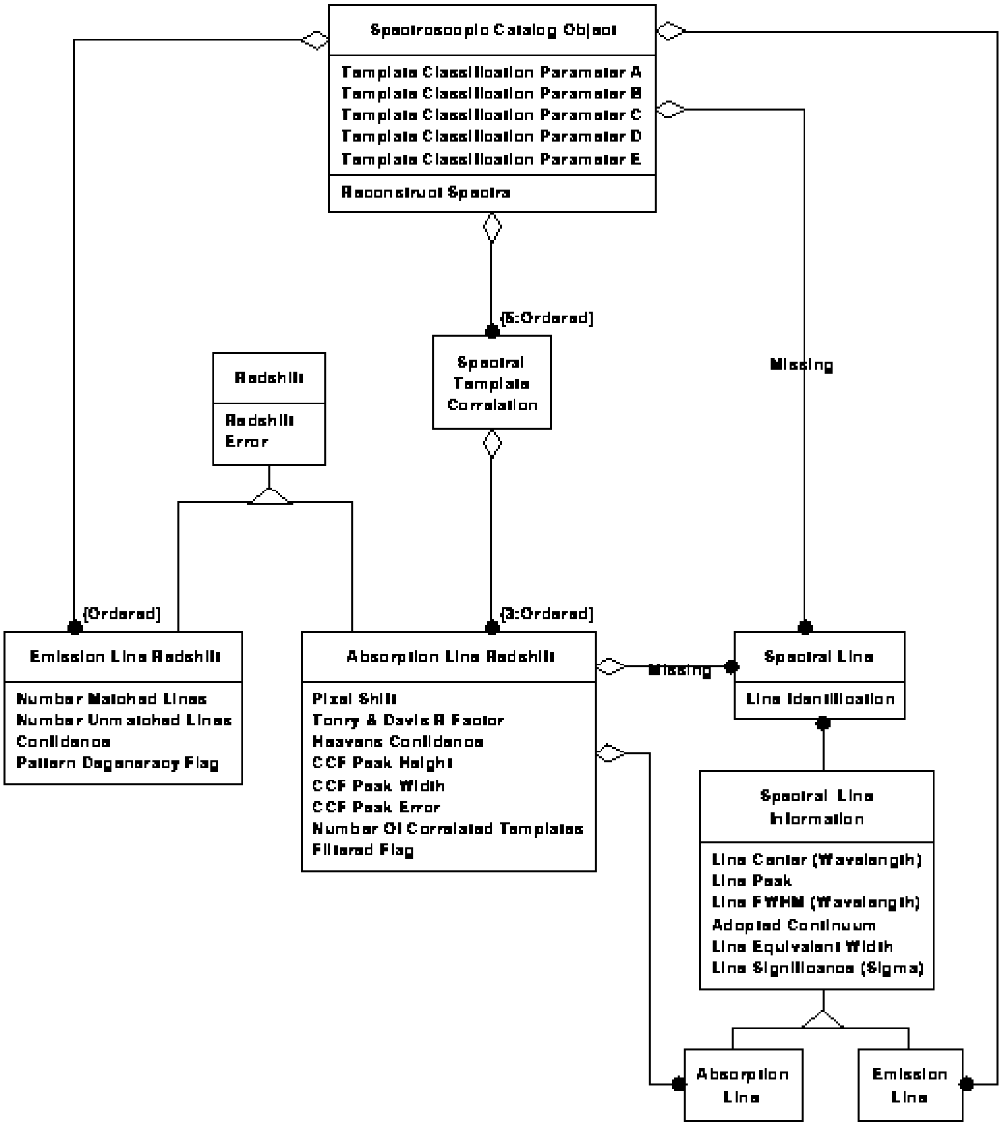
Photometric data model 4

Figure 11.8



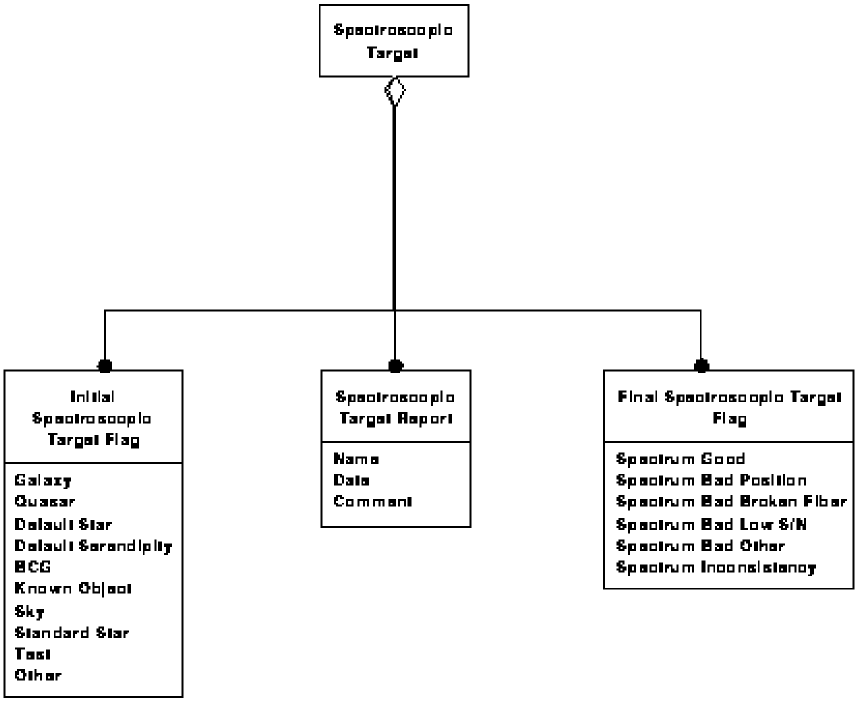
Spectroscopic data model 1

Figure 11.9



Spectroscopic data model 2

Figure 11.10



Spectroscopic data model 3
