3. Science

3.1. Large Scale Structure

Our curiosity about the fundamental nature of the Universe on large scales provides much of the motivation for the SDSS. Here we outline how the size, uniformity, and completeness of this survey will enable us to characterize large scale structure at the present epoch.

3.1.1 *Scientific Questions*

Observations of the cosmic microwave background tell us that the early Universe was homogeneous, but we observe today's Universe to be highly structured over a vast range of scales, from planets to superclusters. The goal of studying large scale structure is to understand the transition from a Universe filled smoothly with hot plasma to one filled with galaxies, clusters, and superclusters. The leading hypothesis is that gravitational instability amplified small fluctuations present at recombination into the structures that we observe today. There are various ways to state the field's "big questions," but any complete list would include the following. Is the gravitational instability explanation basically correct? If so, what were the properties of the primordial fluctuations, and what physical process created them? What is the dark matter? What are the values of the density parameter, Ω, and the cosmological constant, Λ? The answers to these questions are intimately linked to theories of particle physics, and to theories of the origin, early history, and ultimate fate of the Universe. Cosmology thus offers an indirect glimpse of physics at energies that can never be reached in terrestrial experiments.

In addition to these "fundamental" questions, there is a second category of important, "astrophysical" questions, related to galaxy formation. How did galaxies form? What physical processes, besides gravity, played an important role? What processes determine galaxy luminosity, size, color, and morphology? What is the relation between the distribution of galaxies and the underlying distribution of mass? Because galaxies are the markers by which we trace large scale structure, we cannot address the first category of questions without simultaneously addressing the second, especially the final question about the relation between galaxies and mass.

Several factors will make the SDSS a uniquely powerful database for answering these questions. One is the sheer number of galaxies in the redshift sample, a nearly two-order-of-magnitude increase over the largest existing surveys. Another is the quality and uniformity of the photometric data that will be used to select spectroscopic targets and measure Galactic extinction; these will give superb control over systematic effects that might otherwise limit the accuracy of clustering analyses on large scales. The photometric data and high-resolution galaxy spectra also allow one to classify the galaxies of the redshift sample in a variety of ways, making it possible to study the relative clustering of different galaxy types. Finally, while the 1,000,000 galaxy redshift sample is perhaps the most visible element of the SDSS, at least from the point of view of large scale structure, the deeper, multi-color, photometric survey and the spectroscopic survey of quasars and their absorption systems will also make major contributions to our understanding of galaxy and structure formation. Together these features of the SDSS will bring unprecedented precision, dynamic range, and detail to the study of structure in the Universe.

3.1.2 *General Strategy of the Redshift Survey*

The design of the SDSS redshift survey is crucial to its effectiveness as a tool for studying large scale structure. The goal of this survey is not to measure a single statistical property of the distribution of galaxies, such as the power spectrum of density fluctuations, or the

maximum size of non-linear structures. Instead, we desire an accurate description of the galaxy distribution over a wide range of physical scales, in order to address the questions listed above. To meet this goal, we plan to do a wide-angle, deep redshift survey which fully samples the galaxy distribution (*cf.* Chapter). In the remainder of this subsection, we describe the scientific motivation for this approach, and in Sections 3.1.3 and 3.1.4 we discuss ways of using the SDSS data to measure cosmological parameters and test theories of structure formation.

Figure 3.1.1 shows the volume of the redshift survey in schematic form: the upper cone is the Northern survey, while the narrow fan below is the Southern stripe (our current plan is actually to do three separate Southern stripes; see below). Figures 3.1.2 and 3.1.3 display two views of a simulated SDSS spectroscopic sample, drawn from a large N-body simulation of a low-density, cold dark matter Universe. Details of the simulation are given in Chapter 13. For these simulations, the spectroscopic sample consisted of all galaxies with apparent magnitude $r' < 17.55$ and apparent half-light diameter greater than 2 arcseconds, selecting just under one million galaxies over the survey area. Since the time that this simulation was done, the details of the galaxy selection criteria have changed; see below. In any case, the qualitative nature of this figure is not sensitively dependent on these details. Figure 3.1.2 shows the distribution of this Northern spectroscopic sample on the sky. We have plotted only half of the points, selected at random, so as not to blacken the page completely. Figure 3.1.3 shows the distribution of galaxies in redshift space in a 6° by 130° slice along the survey equator, which contains about 6% of the spectroscopic sample. Figure 3.1.4 shows the redshift histogram of the sample galaxies. The median depth of the sample is $\approx 300\ h^{-1}$ Mpc, and there is useful information on the density field to $\approx 600\ h^{-1}$ Mpc, with a few galaxies as far away as $1000\ h^{-1}$ Mpc.

Large Scale Structure and Survey Strategy

A few examples from the recent history of redshift surveys illustrate how the detection of structure depends on both the survey geometry and the sampling rate. Observing in narrow pencil beams in the direction of the constellation Boötes, Kirshner *et al.* (1981) detected a $60\ h^{-1}$ Mpc diameter void in the galaxy distribution. The geometry and sampling of the Kirshner *et al.* survey was well suited to detection of a single void, but it could not answer the question of how common such structures are. De Lapparent, Geller and Huchra (1986, hereafter CfA2) measured redshifts of a magnitude-limited sample of galaxies in a narrow strip on the sky, and found that (1) large voids fill most of the volume of space and (2) these voids are surrounded by thin, dense, coherent structures. Demonstration of the first result was possible because of the geometry and depth of the survey, demonstration of the second because of the dense sampling. There were hints of this structure in the shallower survey of the same region by Huchra *et al.* (1983), but the deeper CfA2 sample was required to reveal the structure clearly. Sparse samples of the galaxy distribution, *e.g.* the QDOT (Saunders *et al.* 1991) survey of one in six IRAS galaxies, have provided useful statistical measures of low-order clustering on large scales, but sparse surveys have less power to detect coherent overdense and underdense regions. Perhaps because of this limitation, members of the QDOT team have elected to follow up their one-in-six survey by obtaining a complete (one-in-one) survey of IRAS galaxies to the same limiting flux, now nearing completion.

In order to encompass a fair sample of the Universe, a redshift survey must sample a very large volume. We do not know *a priori* how large such a volume must be, but existing observations provide the minimum constraints. Our picture of the geometry and size of existing structures became clearer as the CfA2 survey was extended over a larger solid angle (*e.g.* Vogeley *et al.* 1994). The largest of the structures found is a dense region measuring $150\ h^{-1}$ Mpc by $50\ h^{-1}$ Mpc – the "Great Wall" (Geller and Huchra 1989). Giovanelli *et al.* (1986) find a similar coherent structure in a complete redshift survey of the Perseus-Pisces

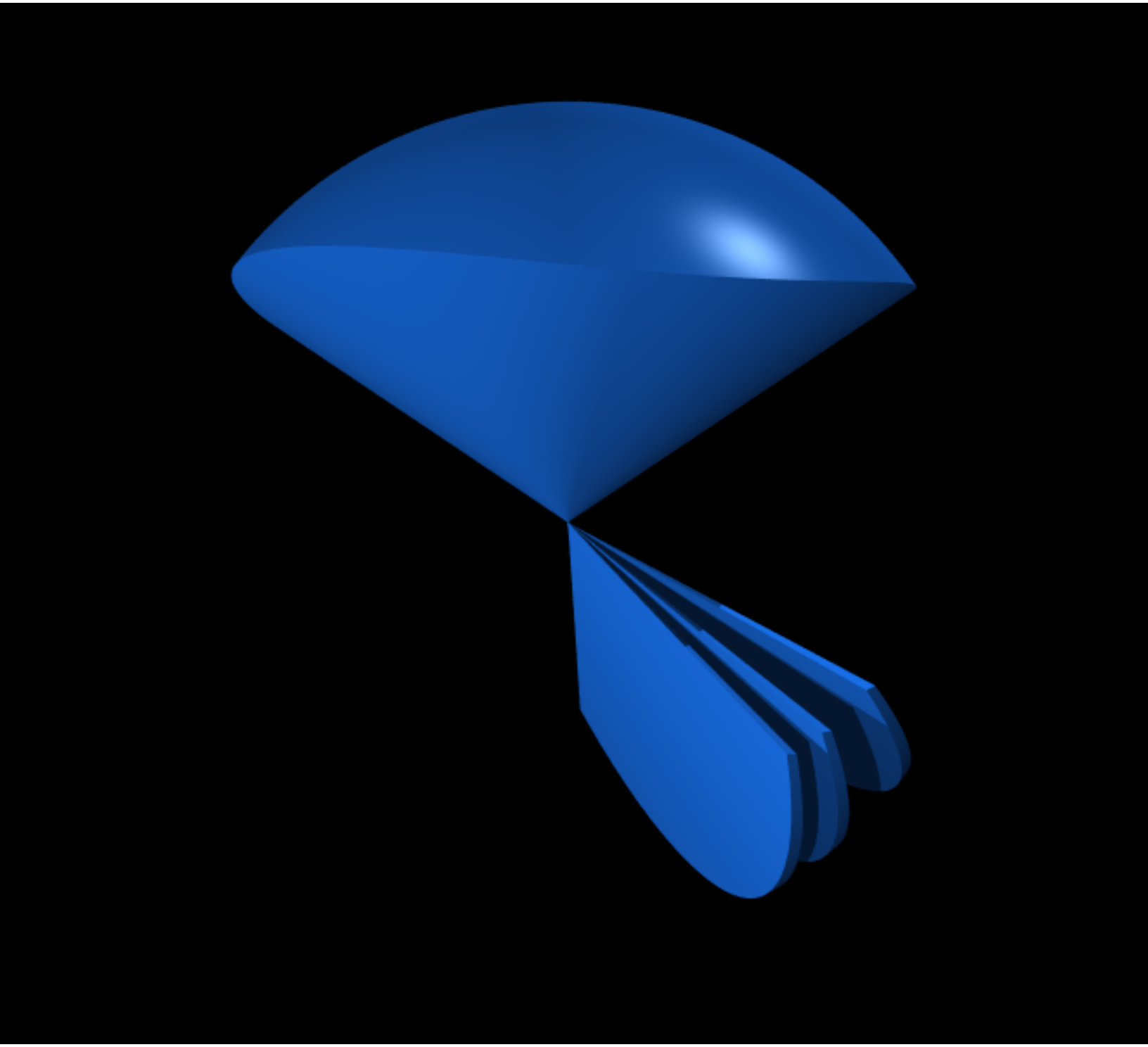

Figure 3.1.1 A schematic representation of the survey geometry. The upper cone is the Northern spectroscopic survey over a large solid angle, while the three thin fans below it are the Southern survey.

Figure 3.1.2 The distribution on the sky of a simulated catalog. This is for the Northern spectroscopic sample, drawn from a large N-body calculation. The survey footprint is a $130^\circ \times 110^\circ$ ellipse, shown in equal-area projection. We have randomly removed half of the points to reduce crowding. Details of the simulation are described in Chapter 13.

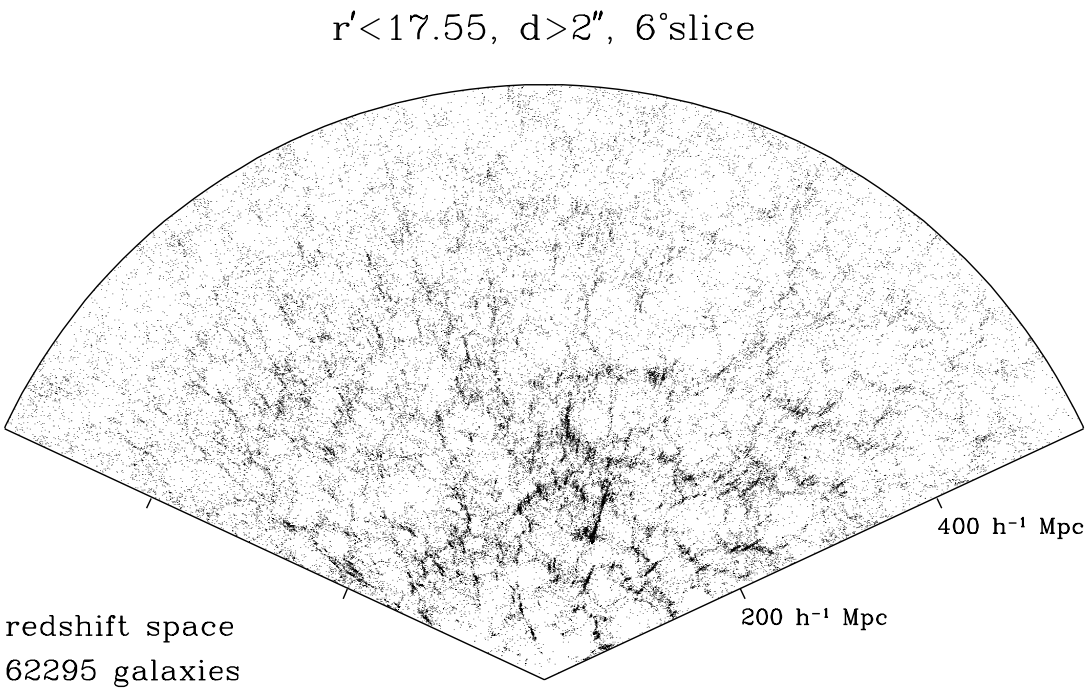

Figure 3.1.3 The redshift-space distribution of galaxies. The objects lie in a slice 6° thick along the survey equator, from the same simulation illustrated in Figure 3.1.2. Galaxies are plotted at the distance indicated by their redshift, hence the appearance of clusters as "fingers-of-God." This slice contains roughly 6% of the galaxy redshift sample.

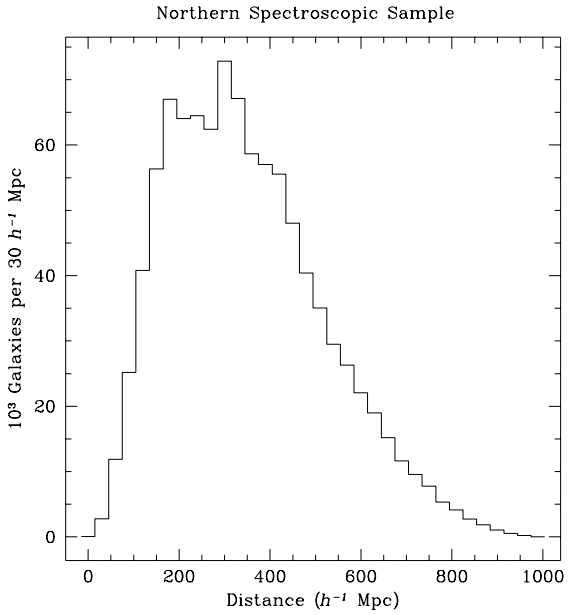

Figure 3.1.4 The simulated redshift histogram of galaxies. The galaxies are those in the simulated spectroscopic sample shown in Figures 3.1.2 and 3.1.3. The histogram is computed in bins of $\Delta z = 0.01$.

region, and da Costa *et al.* (1988) find another such "wall" in the Southern hemisphere (cf. Santiago *et al.* 1996). However, none of these surveys covers a large enough volume to investigate the frequency of these structures. In quantitative terms, the observed power spectrum of galaxy density fluctuations continues to rise on scales up to $\gtrsim 100\ h^{-1}\,$Mpc (Vogeley *et al.* 1992; Loveday *et al.* 1992; Fisher *et al.* 1993; Feldman *et al.* 1994; Park *et al.* 1994; Baugh and Efstathiou 1993; Peacock and Dodds 1994; da Costa *et al.* 1994; Landy *et al.* 1996; cf. the review by Strauss and Willick 1995). The observed power spectrum may be influenced by these few, largest, nearby structures. Thus a fair sample clearly must include many volumes with scale $\sim 100\ h^{-1}\,$Mpc.

The sampling rate and geometry of a survey can strongly influence the largest structures seen. A very deep pencil-beam survey by Broadhurst *et al.* (1990, hereafter BEKS) showed remarkable clustering signatures on a scale of $128\ h^{-1}\,$Mpc. First results from the redshift survey of Shectman *et al.* (1996, hereafter LC for the Las Campanas survey), has found evidence for an excess of power on similar scales (Landy *et al.* 1996). The LC survey presently covers a wedge in the Southern sky; it is deeper than CfA2 but less deep than the BEKS survey. While different survey geometries are appropriate for elucidating different features of large scale structure, a deep survey with a large opening angle and full sampling of the galaxy population is the only way to get a complete picture of the galaxy distribution. Coverage of the whole sky is extremely helpful if one wants to compare velocity and density fields in the nearby Universe, and redshift surveys of infrared-selected galaxies (Strauss *et al.* 1992; Fisher *et al.* 1995; Lawrence *et al.* 1995) are invaluable for this purpose. However, these surveys are much shallower than the SDSS, with almost two orders of magnitude fewer galaxies, so they are not nearly as powerful for statistical analyses of clustering.

Survey Geometry

Our ability to constrain the fluctuation spectrum on scales approaching those examined by CMB anisotropy measurements depends on both the total volume and the shape of the survey. An important feature of the SDSS redshift survey is the "window function" with which it will observe the Universe. A slice or pencil-beam geometry samples many independent volumes with a smaller number of galaxies, but the small extent in one or two dimensions makes the fluctuation spectrum computed for such surveys strongly affected by aliasing (*cf.* Kaiser and Peacock 1991; Tegmark 1995). In other words, strong clustering on small scales may overwhelm the true signal on much larger scales. The SDSS redshift survey volume is large enough to include many independent structures on the scale of the "Great Wall," and has both sufficient angular coverage and depth to ensure that measurement of the fluctuation spectrum on scales of a few hundred $h^{-1}\,$Mpc is not strongly affected by aliasing. The redshift survey in the South Galactic cap will further help by nearly doubling the largest baseline in the survey. Indeed, we plan to do galaxy spectroscopy in three stripes in the South, along great circles centered at $\alpha = 0^h, \delta = 0^\circ$, $\alpha = 0^h, \delta = +15^\circ$, and $\alpha = 0^h, \delta = -10^\circ$. This maximizes the number of independent measures of the longest baselines. In Figure 3.1.5 we provide a schematic representation of the window function in the Fourier domain for the SDSS and several current surveys. Figure 3.1.6 shows the window functions more quantitatively. The quantity plotted is the mean square of the window function as a function of k (averaged over direction), and normalized to unity at $k = 0$. Also plotted is the window function for the CfA2 survey (Park *et al.* 1994), which represents the current state of the art; the tremendous improvement of the SDSS over the CfA2 for measuring the power spectrum on large scales with high resolution in k-space is immediately apparent.

Sampling – Motivation for a Filled Redshift Survey

A key element of the strategy for the SDSS redshift survey is full sampling of the galaxy distribution. For some specific instrumental set-ups, specific models of the galaxy distribution,

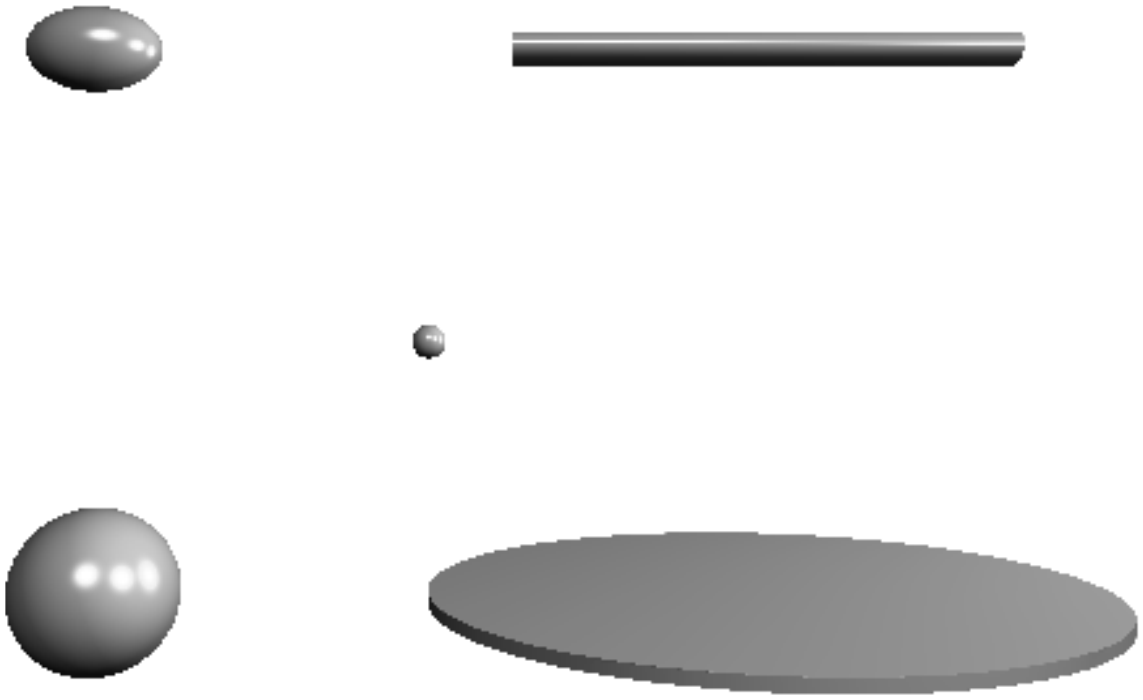

Figure 3.1.5 A schematic three-dimensional comparison of window functions. The window function is the Fourier transform of the survey volume; it can be thought of as the point-spread function for measurements of the galaxy power spectrum. The narrower the window function, the better the sensitivity to large scale clustering. Surveys with highly anisotropic shapes (slices, pencil beams) have highly anisotropic window functions; as a result, estimates of large scale clustering can be polluted by aliased power from small scales. The window functions of five surveys are shown here: lower left is QDOT, upper left is CfA2, upper right is LC, lower right is the BEKS survey, and the small dot in the center is the SDSS.

and specific, low-order measures of large scale clustering, one can gain efficiency by sparse sampling, *i.e.* by observing only a fraction of the galaxies down to some limiting magnitude. However, we do not know the properties of the large scale distribution of galaxies *a priori*, and we want the flexibility to measure a wide range of clustering statistics from the data, including statistics that characterize higher-order clustering. In this case, there is no better strategy than simply observing all galaxies; any attempt to "optimize" in other ways would depend strongly on our presumptions about the the statistical properties and the topology of the large scale galaxy distribution. The design of the SDSS telescope itself (in particular, the field of view and the number of fibers in the spectrograph) has been affected by these considerations.

If fluctuations in the Universe are strictly Gaussian, their full statistical description is contained in the two-point correlation function or in its Fourier transform, the power spectrum. The phases of the individual Fourier components are random for such a process, and all higher-order measures of clustering vanish. In this case, Kaiser (1986) shows that measuring the redshifts of a small fraction of the galaxies is the most efficient way to measure the power spectrum on large scales, for a given amount of telescope time, and *for a single-object*

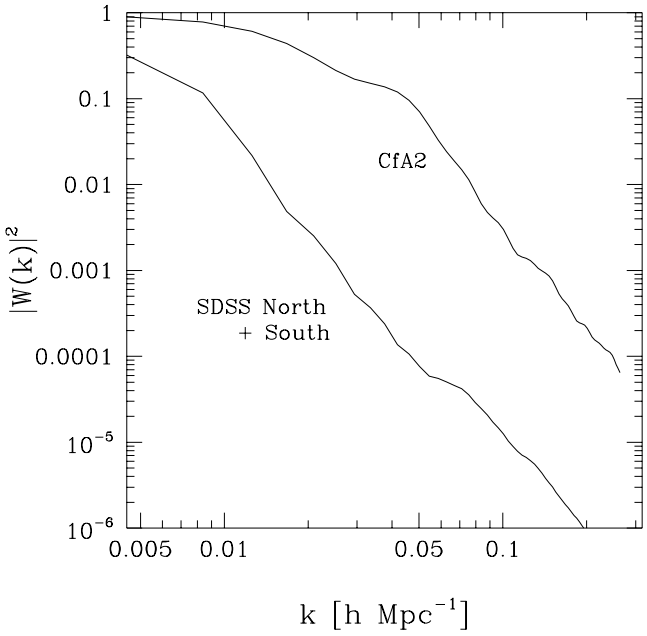

Figure 3.1.6 The Fourier window function of a volume-limited sample. The function is shown for the SDSS galaxy redshift survey ($R_{max} = 500h^{-1}$ Mpc), including the main North Galactic Cap survey and three stripes in the South, compared to the window function for the full CfA2 Redshift Survey ($R_{max} = 130h^{-1}$ Mpc; cf., Park *et al.* 1994). The observed power spectrum is a convolution of the true power spectrum with this function. To reveal the relative sharpness of the window functions, the normalization in this figure is such that $|W(k = 0)|^2 = 1$.

spectrograph. However, given the field of view and number of fibers in our spectrograph, we will be able to observe essentially all the galaxies in a given field to the faintest magnitude for which we can measure redshifts in a reasonable amount of time. With the SDSS instrumentation, therefore, *there would be no gains in efficiency by sparse sampling.* Moreover, if the Universe contains sharp large scale features like the "Great Wall", a sparsely sampled survey may fail to identify them because it is less sensitive to the higher-order correlations (equivalently, the phase correlations) that characterize such structures. A distribution with high-order clustering can be very different from a homogeneous, isotropic, Gaussian random field with the same second-order statistical properties. We demonstrate this with a simple but graphic example in Figure 3.1.7, which shows a pair of two-dimensional distributions with the same two-point correlation functions. One of these has very non-random phases, while the phases in the other have been randomized. A sparse survey designed only to measure the two-point correlation function would miss the very real differences between these two distributions.

Sparse sampling can severely compromise one's ability to recover the underlying galaxy density field (*cf.* de Lapparent *et al.* 1991; Ramella *et al.* 1990). Figure 3.1.8 illustrates this point with a simple example (cf., Szapudi and Szalay 1995). Suppose that we are interested in the distribution of galaxy counts in cells that contain, on average, 10 galaxies above the survey magnitude limit. We assume that the underlying distribution of galaxy counts N is

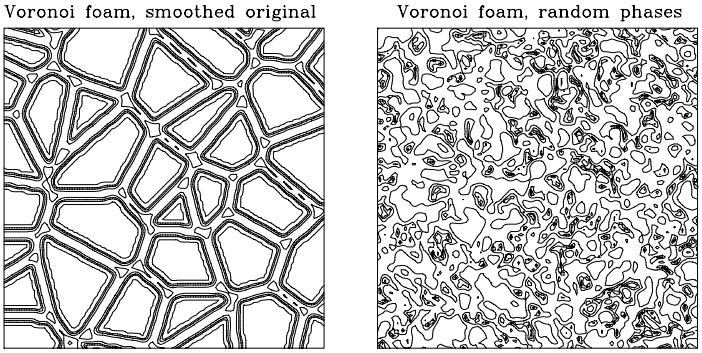

Figure 3.1.7 Two distributions with the same correlation function. The left-hand figure is a two-dimensional Voronoi foam, generated by the median surfaces between Poisson "seeds" at a mean separation of 100 h^{-1} Mpc. In this simple toy model, galaxies reside only on the walls of the foam, smoothed to give the walls a finite thickness. The structure has a well defined second-order statistic, but it also has correlated phases. This picture has been Fourier transformed, all the phases randomized, then transformed back again. The result is another two-dimensional density plot, shown on the right hand side, with the same second-order properties, but with a Gaussian distribution.

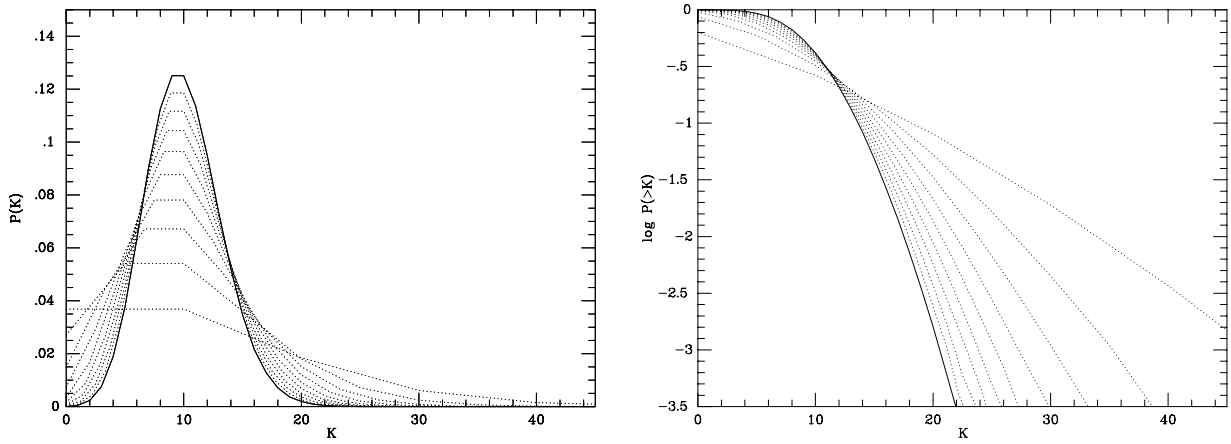

Figure 3.1.8 Effects of low sampling rates on the distribution of galaxy counts. The simple model is described in the text. The left-hand panel shows the differential distribution, the right-hand panel the cumulative distribution (with a logarithmic axis). In each case the solid line is the true distribution of the model, a Poisson distribution of mean 10. Dotted lines show the distribution of inferred counts $K = M/p$, where M is the number of galaxies observed, for sampling rates $p = 0.9, 0.8, 0.7, \cdots, 0.1$. With sparse sampling, the distribution of inferred counts is much broader than the true distribution.

a Poisson distribution with a mean of 10. The solid line in the left-hand panel shows this distribution. Now suppose that we observe the galaxy distribution with a sampling rate $p < 1$. The number of galaxies M detected in a cell that actually contains N galaxies will have a binomial distribution

$$P(M|N,p) = \binom{N}{M} p^M (1-p)^{N-M} \quad .$$

The unbiased estimator for the true number of galaxies in the cell is simply $K = M/p$. The dotted lines in the left-hand panel show the distribution of inferred counts K for sampling rates $p = 0.9, 0.8, 0.7, \ldots, 0.1$. This distribution is

$$P(K|p) = \sum_N P(Kp|N,p)P(N),$$

where $P(Kp|N,p)$ is the binomial distribution given above and $P(N)$ is a Poisson distribution of mean 10. At $p = 0.9$ the distribution of inferred counts is quite close to the true distribution, but as the sampling rate decreases the distribution of inferred counts becomes very broad. The right-hand panel shows a logarithmic plot of the cumulative distribution, the number of cells with inferred counts exceeding a value K. For large values of K, the curves for low sampling rates exceed the true curve by orders of magnitude. To some extent one can correct estimates of statistical quantities for finite sampling (cf., Colombi, Bouchet, and Schaeffer 1994, 1995), but when the sampling rate is low even a corrected estimate can have large random errors and systematic biases that are sensitive to *a priori* assumptions about the clustering. Furthermore, as Figure 3.1.8 makes clear, it can become virtually impossible to infer the actual galaxy density field — the true number of galaxies in each spatial cell — because sparse sampling leaves only a tenuous connection between the number of galaxies present and the number actually observed.

Low sampling rates seriously jeopardize the ability to recover information about the galaxy distribution beyond second-order clustering statistics. By obtaining a full sample, the SDSS will produce a data set that is well suited to a wide range of analyses aimed at a wide range of questions, including ones that have not yet been formulated. The design of the imaging and spectroscopic systems allows this full sampling to be obtained without loss of efficiency.

Galaxy Selection Algorithm

Simply saying that we will select for spectroscopic observation the brightest million galaxies in the survey area is not adequate. There are many ways in which to measure magnitudes for extended galaxies; we must also specify in which band or bands the galaxy selection is to be done. The selection criteria must be simple, and robust. They must target the full range of galaxy types, and must give a high probability of actually being able to measure a redshift. We must have selection criteria such that the efficiency of identifying the right spectroscopic targets (*e.g.* separating galaxies from stars) and obtaining redshifts should not be sensitive to the observing conditions or the Galactic latitude. Finally, in order to do large-scale structure studies, we need to be able to define a selection function, which quantifies how the galaxy population for which we have spectra depends on redshift. This allows us to make quantitative comparisons of the derived galaxy properties (such as the density field) at different distances in the sample. This in turn requires that the sample be selected on physically meaningful quantities. For this reason, we cannot simply select on the flux within the $3''$ aperture of the fibers, even though this would maximize the efficiency with which redshifts are measured.

With all these desiderata in mind, we plan to select our galaxy sample in the r' band using Petrosian (1976) magnitudes. We correct the photometry for the effects of Galactic extinction, using the *a priori* reddening map discussed in Chapter . We considered a joint selection in r' and g', fearing that r' alone would systematically exclude blue (i.e., spiral), nearby (because of the steepness of the K-correction) galaxies. However, tests with realistic mock catalogs (Chapter 13) showed that the redshift range and morphological mix of samples selected in these two bands was quite similar; the $g' - r'$ colors of galaxies are not a strong function of redshift (at least to the depth of our sample) or Hubble type. The u' band is of course too shallow to be useful for target selection (Chapter 8).

Define the Petrosian *radius* R_P as that at which the ratio of azimuthally-averaged surface brightness in an annulus to the mean surface brightness within R_P, falls to some specified value:

$$\frac{\int_{0.8R_P}^{1.25R_P} I(r)2\pi r dr/[\pi(1.25^2 - 0.8^2)R_P^2]}{\int_0^{R_P} I(r)2\pi r dr/[\pi R_P^2]} = f_1.$$

The Petrosian flux is that within a fixed number of Petrosian radii:

$$f_P = \int_0^{f_2 R_P} I(r)2\pi r dr.$$

The Petrosian magnitude is of course given by $C - 2.5\log f_P$. Typical parameters are $f_1 = 1/4$ and $f_2 = 3$. Thus the Petrosian radius is tied to the photometric properties of the galaxy, and defines a fixed aperture on the galaxy that is independent of redshift.

The Petrosian ratio as defined above has the drawback of being non-monotonic for some galaxies with multiple components (such as disk galaxies with prominent bulges, or active galaxies with stellar nuclei); for cD galaxies with power-law surface-brightness profiles, the Petrosian ratio becomes a constant with scale. We are still exploring algorithms to deal with these problematic cases. For f_1 as small as 1/4, the multiple component issue is rarely a problem, but we may be forced to measuring a magnitude through an aperture based on a model fit for galaxies with power-law profiles.

We need a secondary cut to minimize the number of very low-surface brightness galaxies, for which the total light down the fiber aperture is very small. We wish to avoid having a cut on the light within the $3''$ aperture, but rather we continue to define quantities based on the galaxy profile itself, by making a cut on the surface brightness within the circular aperture which contains half the Petrosian flux. We find that limiting in Petrosian magnitude at $r' = 18.15$ and Petrosian surface brightness at $\mu_{r'} = 22$ mag arc-sec^{-2} yields a sample of one million galaxies over the survey area, with a sharp cutoff in the distribution of $3''$ aperture magnitudes at $r'_{3''} = 19.5$. Moreover, the redshift distributions of spirals and ellipticals are very similar. There are some small corrections to the Petrosian magnitude and surface brightnesses, which depend on the compactness of the galaxy and the seeing, but these are straightforward to apply.

Because some fraction of the large galaxies will overlap the edges of the CCD's (Section 3.8.2), photometry of these galaxies will be unreliable, and thus we intend to augment the spectroscopic sample with *all* galaxies with diameters greater than $1'$ (the overlap between strips), independent of their Petrosian properties. The total number of galaxies affected will be quite small.

The faintest $3''$ limiting magnitude of 19.5 is near the point at which the contribution from the galaxy and sky in the aperture are equal; integration times would have to become substantially larger if we were to go much fainter (cf. the discussion in Section 5.3). However, there is a class of galaxies, the most luminous metal-rich red ellipticals, which have very strong absorption lines, and for which simulations show that we can go 1.5 magnitudes fainter and still measure redshift-quality spectra. We thus plan to select such galaxies by their colors and their photometric redshifts (cf., Figure 3.1.11), and target them 1.5 mag fainter than the main galaxy sample. The details are described in Section 3.5. Because these objects are luminous, they will be seen to large redshifts; we will have essentially a volume-limited sample of these objects to $z \approx 0.5$ with which to do large-scale structure studies on the largest scales, and to study the evolution of clusters with redshift. Moreover, because these photometric properties are those of the Brightest Cluster Galaxies (e.g., Postman and Lauer 1995, and references therein), this automatically includes all clusters with a "normal" BCG out to those redshifts (cf. the discussion in section 3.5).

3.1.3 *Statistical Measures of Large Scale Structure*

Power Spectrum and Correlation Function

The most basic statistical measures of galaxy clustering are the power spectrum and the two-point correlation function. The two quantities are a Fourier Transform pair, but estimators of the two can have quite different noise properties when applied to a finite sample, so in practice it is useful to measure them independently. With the SDSS, we can measure the angular correlation function, $w(\theta)$, from the photometric survey, and we can measure the redshift-space power spectrum and correlation function directly from the redshift survey. Because galaxy fluctuations are very small on the largest scales probed by the SDSS, the selection of galaxies must be exquisitely accurate to ensure that measurements are limited by statistical rather than systematic errors. *Thus, the high-quality imaging survey is an absolutely essential prerequisite for the redshift survey.* The photometric survey will allow galaxy selection to be precise and uniform across the survey region, much more so than for any other large scale redshift survey in existence or planned for the near future. The photometric data will also yield accurate measurements of Galactic extinction (*cf.* Section and Chapter), which would otherwise be a debilitating source of systematic errors and limit the accuracy with which the large-scale structure could be measured. The determination of the extinction is discussed in more detail in Section .

The state of the art in measurements of the angular correlation function is the analysis of $\sim 10^6$ galaxies in the digitized, photographic, APM survey by Maddox *et al.* (1990; cf., Baugh and Efstathiou 1993). At large angles, the accuracy of the estimates is limited by systematic uncertainties about Galactic extinction and photometric uniformity rather than by statistical uncertainties. The SDSS photometric survey will have 100 times as many galaxies as APM, much smaller photometric errors, and much tighter control of extinction and uniformity. It should therefore yield much more accurate measures of $w(\theta)$. The additional leverage of color-redshift relations (*cf.* Section 3.1.4) will make it possible to study the evolution of clustering in the photometric sample, especially in the Southern survey, which will go about 2 magnitudes deeper than the Northern survey.

The SDSS redshift sample, itself the size of the APM catalog, will yield precisely calibrated measures of the present-day correlation function, $\xi(s)$, and power spectrum, $P(k)$, in redshift space. The SDSS will probably be the first survey with the sensitivity and freedom from systematic errors to determine unambiguously the scale at which $\xi(s)$ goes negative — this scale provides an important constraint on theoretical models.

The power spectrum of the galaxy distribution has been measured from existing redshift surveys out to scales of about $100\,h^{-1}$ Mpc with approximately a factor of two precision (Vogeley *et al.* 1992; Fisher *et al.* 1993; Feldman *et al.* 1994; Park *et al.* 1994; Peacock and Dodds 1994; cf., Strauss and Willick 1995 for a review). These scales are roughly one order of magnitude smaller than the smallest scales for which the COBE satellite has measured fluctuations in the CMB (Smoot *et al.* 1992; Bennett *et al.* 1994). In the simplest interpretation, the CMB anisotropies are directly related to fluctuations in the gravitational potential (Sachs and Wolfe 1967). With the SDSS redshift survey, we will measure the galaxy power spectrum on scales where the Sachs-Wolfe effect is measurable in the CMB, and thus directly compare the amplitude of galaxy fluctuations to that of gravitational potential fluctuations.

After one year of operation, the SDSS spectroscopic sample will be an order of magnitude larger than the largest redshift surveys that exist today, and a factor of several larger than the largest redshift surveys expected to exist by that time. At the end of five years, we will measure the power spectrum over an enormous range of scales with unprecedented precision.

Figure 3.1.9 shows the model power spectra for the SDSS galaxy sample and for the luminous red galaxies (cf. Section 3.5). The power spectra in Figure 3.1.9 have been computed for an $\Omega h = 0.25$ CDM model, normalized to $\sigma_8 = 1$, with the error bars calculated for a volume limited sample in the Northern survey region. Both cosmic variance and shot noise are included in the calculation of the error bars. The power spectrum for the luminous red galaxies assumes a volume limited sample to a depth of $z = 0.45$. These galaxies, found preferentially in the densest regions of the Universe, are assumed to be biased by a factor of 2 with respect to the complete galaxy sample, so that the amplitude of the power spectrum is a factor of four higher than that of the galaxy sample. Both cosmic variance and shot noise are included in the estimation of error bars, but other effects, e.g. clustering evolution, are likely to dominate the uncertainties in this sample. If the model power spectrum shown in Figure 3.1.9 is a reasonable representation of the truth, *we will be able to measure it beyond the turnover scale*, which corresponds to the radius of the horizon at matter-radiation equality.

Also plotted in Figure 3.1.9 are the smallest comoving wavelength scales that can be observed by COBE and by the upcoming MAP (Microwave Anisotropy Probe) mission. Only SDSS can access the COBE scale; but comparison of the power spectra obtained by SDSS and other redshift surveys with the MAP data should allow the examination of structure on a very wide range of scales. The possibility of making this comparison is very exciting, for it will reveal the evolution of structure across the lifetime of the Universe.

The detailed shape of the power spectrum reflects the physical source of the primordial fluctuations (*e.g.* inflation, topological defects) and the matter content of the Universe (*e.g.* the relative mass fractions of radiation, other relativistic components, hot dark matter, cold dark matter, baryons, and vacuum energy). Current observations are sufficient to rule out some theoretical models, such as the simplest and most attractive versions of the cold dark matter scenario. The high-precision measurements afforded by the SDSS, especially near the peak of the power spectrum, will provide much more stringent tests of theoretical models, and the ability to detect subtle features in the spectrum should allow us to ask detailed physical questions. For instance, if baryons contribute a non-negligible fraction of the cosmic mass density, then oscillations of coupled baryon-photon fluctuations produce oscillations in the power spectrum at wavelengths near the horizon radius at recombination (*cf.* Holtzman 1989). Measuring such oscillations would allow us to constrain Ω_{baryon} and to distinguish between adiabatic and isocurvature fluctuations. The high resolution we will get in k-space due to our large filled volume (cf., Figure 3.1.5) is absolutely necessary for looking for such subtle effects.

The primordial power spectrum of mass fluctuations differs from the present-day power spectrum of the galaxy distribution because of (1) non-linear gravitational effects (which are strongest on small scales), (2) distortions by peculiar velocities in redshift space (which enhance the spectrum on large scales and depress it on small scales), and (3) possible "biasing" between the galaxy and mass distributions (with unknown scale dependence). These effects complicate the comparison between theory and observation, but they are extremely interesting in their own right, and they can be explored and constrained by other statistical methods, as we discuss below.

Redshift-Space Distortions

Peculiar velocities shift galaxies along the line of sight in redshift space, creating anisotropic distortions in the correlation function and the power spectrum. On small scales, velocity dispersions stretch dense structures along the line of sight. Thus clusters of galaxies appear as "fingers-of-God" in redshift space. On large scales, coherent outflows from voids and inflows to superclusters amplify density contrasts along the line of sight (Sargent and Turner 1977; Kaiser 1987). One can recover the correlation function or power spectrum in real space

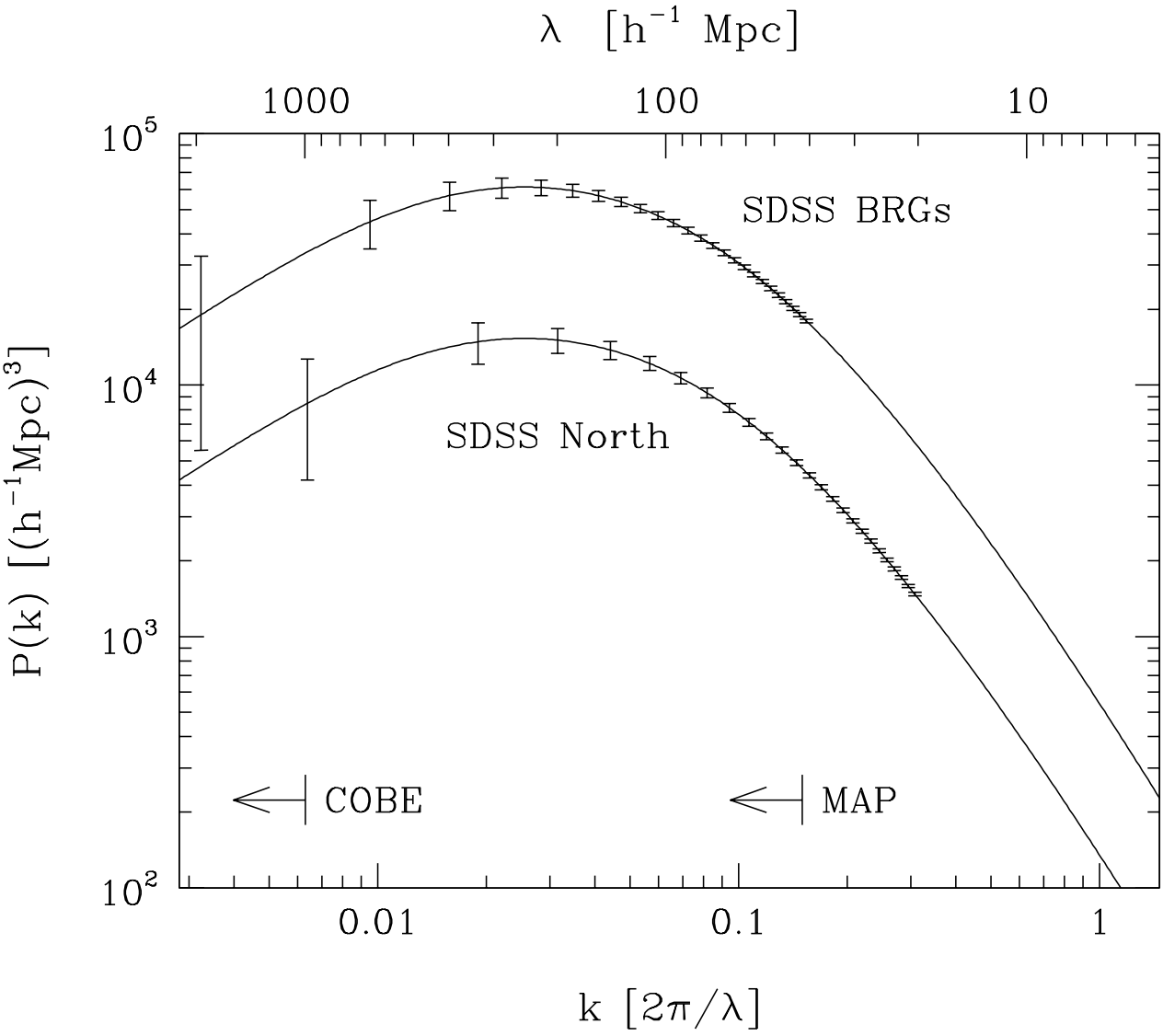

Figure 3.1.9 Model power spectra. The power spectra for the SDSS Northern galaxy survey and for the luminous red galaxies (BRGs) are shown. The latter are assumed to be biased by a factor of two with respect to the galaxies. The individual points and their error bars are statistically independent. On small scales, the errors are smaller than the smallest bars. Also shown are the smallest comoving wavelength scales accessible to COBE and to the upcoming Microwave Anisotropy Probe (MAP).

by projecting over the redshift direction, but the distortions themselves are interesting, and can be quantified by measuring the correlation function or power spectrum as a function

of *direction* with respect to the line of sight, at fixed values of the pair distance or the wavelength. From the anisotropy of the correlation function, one can extract moments of the galaxy pairwise velocity distribution (Davis and Peebles 1983; Fisher *et al.* 1993). These provide powerful constraints on theoretical models, particularly on the values of Ω and the "bias parameter" b, which describes the relative amplitude of galaxy and mass fluctuations. On scales in the linear regime, distortions of the correlation function or the power spectrum depend directly on the combination $\Omega^{0.6}/b$ (Kaiser 1987; Hamilton 1992).

With present data one can estimate $\Omega^{0.6}/b$ with $\sim 50\%$ uncertainties (Hamilton 1993; Fisher *et al.* 1994; Cole *et al.* 1994, 1995; Loveday *et al.* 1995). The large uncertainties reflect the limited galaxy numbers and volumes of existing surveys; these add noise to the statistical estimates, and they prevent one from reaching scales that are completely in the linear regime. In particular, this technique depends on averaging over enough coherent structures of supercluster size that the sample can be considered isotropic in real space. The SDSS will be ideal for this sort of analysis — it will provide the large number of galaxies, large survey volume, and freedom from systematic errors that are needed to measure angular modulations of low-amplitude clustering on large angular scales. From the largest, *i.e.*, linear, scales we will obtain precise measurements of $\Omega^{0.6}/b$. By examining the transition from the linear regime to the non-linear regime, we can break the degeneracy between Ω and b (Cole *et al.* 1995). Obtaining separate constraints on these parameters requires an enormous data set, and we expect that the SDSS will be the first redshift survey to allow an unambiguous distinction between a high density Universe with biased galaxy formation and a low-density Universe in which galaxies trace mass. We will be able to test the robustness of this distinction by applying the same analysis to galaxies of different types — these may have different bias factors, but they should imply the same value of Ω (see Figure 3.1.9 and the discussion below on clustering of different galaxy types).

Higher-Order Correlations and Count Probabilities

The correlation function and power spectrum measure only the *rms* amplitude of fluctuations as a function of scale. To obtain a complete statistical description of clustering, one must also specify the higher-order correlation functions. The careful selection criteria and dense sampling of the SDSS spectroscopic survey are designed to allow accurate measurement of higher-order clustering.

Many statistical measures are sensitive to high-order moments of the galaxy distribution, *e.g.* void statistics and percolation methods. Here we concentrate on the probability distribution of counts in cells (of which void statistics are a special case). White (1979) shows that the count probability distribution is given by an infinite sum over *all* the higher-order correlation functions; thus these two descriptions are, in principle, mathematically equivalent. Estimates of higher-order correlations and count probability distributions have been obtained from a variety of angular and redshift catalogs (*e.g.* Peebles and Groth 1976; Alimi *et al.* 1990; Meiksin *et al.* 1992; Maurogordato *et al.* 1992; Gaztañaga 1994; Bouchet *et al.* 1993; Szapudi, Meiksin, & Nichol 1996; cf., Strauss and Willick 1995 for a review).

On large scales, count probabilities reveal important clues about the origin of primordial density fluctuations. Simple models of inflationary cosmology predict Gaussian probability distributions in linear theory, while topological defect models (*e.g.* cosmic strings, texture) predict non-Gaussian fluctuations. Non-linear evolution distorts the probability distribution, but models with Gaussian initial conditions predict a specific hierarchy of relations between moments of the mass density field, at least on scales that can be treated by perturbation theory (Peebles 1980; Fry 1984; Bernardeau 1992). If the smoothed galaxy density is a local function of the smoothed mass density, then the galaxy distribution retains this hierarchy of moment relations (Fry and Gaztañaga 1993). Existing data (Gaztañaga 1992; Bouchet *et al.*

1993) show the expected relations for the low-order moments, but these observations probe only a modest range of length scales. The redshift and photometric samples of the SDSS will allow much more stringent tests over a large dynamic range. Even small departures from the predictions of Gaussian models would be highly significant, implying the existence of non-Gaussian primordial fluctuations, or a partial decoupling between galaxy density and mass density, which could be caused by large scale modulations in the efficiency of galaxy formation (Babul and White 1991; Bower *et al.* 1993; see the discussion by Frieman and Gaztañaga 1994). One could distinguish these possibilities by studying a range of physical scales and by examining galaxies of different types.

The observed galaxy distribution also exhibits hierarchical relations between higher-order correlation functions in the strongly non-linear regime. Similar behavior is seen in some numerical simulations (*e.g.* Bouchet *et al.* 1991). These relations have recently been established up to eighth order in the Lick and APM catalogs (Szapudi and Szalay 1993; Gaztañaga 1994). These scalings not only tell us about an important symmetry of the galaxy clustering, namely the validity of the $(N-1)$th order tree hierarchy, but also provide a quantitative measure of higher-order organization in the galaxies. The non-linear dynamics of galaxy clustering are not fully understood from first principles, and this remains an active area of research (*e.g.* Balian and Schaeffer 1989a; 1989b). The SDSS will significantly improve the accuracy of these measures and extend them to much higher orders, both in redshift space and projected onto the plane of the sky. It will allow a detailed examination of the transition from the perturbative to the fully non-linear regime, and it may reveal places where the hierarchical description breaks down; this information will help us understand the physical significance of the observed scaling.

Topology of the Galaxy Distribution

A particularly interesting way to measure high-order clustering in the galaxy distribution is to examine the topology of high- and low-density regions. At the qualitative level, one would like to know whether the galaxy distribution is best described by a sponge topology, in which high- and low-density regions have equivalent connectivity, a bubble topology, in which voids are separated from each other by high-density walls, or a cluster topology, with isolated high-density regions residing in a low-density sea. A quantitative measure of topology is the genus curve of Gott *et al.* (1987), who smooth the galaxy density field and measure the genus of isodensity contour surfaces (the number of "holes" or "handles" minus the number of isolated regions), as a function of the fractional volume enclosed by the contour. The use of fractional volumes makes the topology statistic independent of the rms value and one-point distribution of the smoothed density fluctuations, so it complements the power spectrum and count probability statistics described above.

On large scales the topology is closely related to the physics of the primordial fluctuations; in particular, positive and negative fluctuations are statistically equivalent in Gaussian fields, and Gaussian models predict a specific, universal form for the genus curve. Measurements of the genus curve by Gott *et al.* (1989), Moore *et al.* (1992), and Vogeley *et al.* (1994) show that, to within the statistical errors, the galaxy distribution is consistent with Gaussian fluctuations on linear scales. However, none of the existing surveys includes a large enough volume to detect small deviations from the genus curve predicted for a Gaussian model. On smaller scales, Gott *et al.* (1989) and Vogeley *et al.* (1994) find departures from the Gaussian prediction, but the statistical significance of these departures is not overwhelming ($\sim 2\sigma$ for Vogeley *et al.* 1994). It is also unclear whether these these results indicate non-Gaussian initial conditions or residual effects of non-linearity and biased galaxy formation.

Because of its large volume, the SDSS redshift survey will provide a measurement of the topology on linear scales with sufficient accuracy to measure small deviations from the

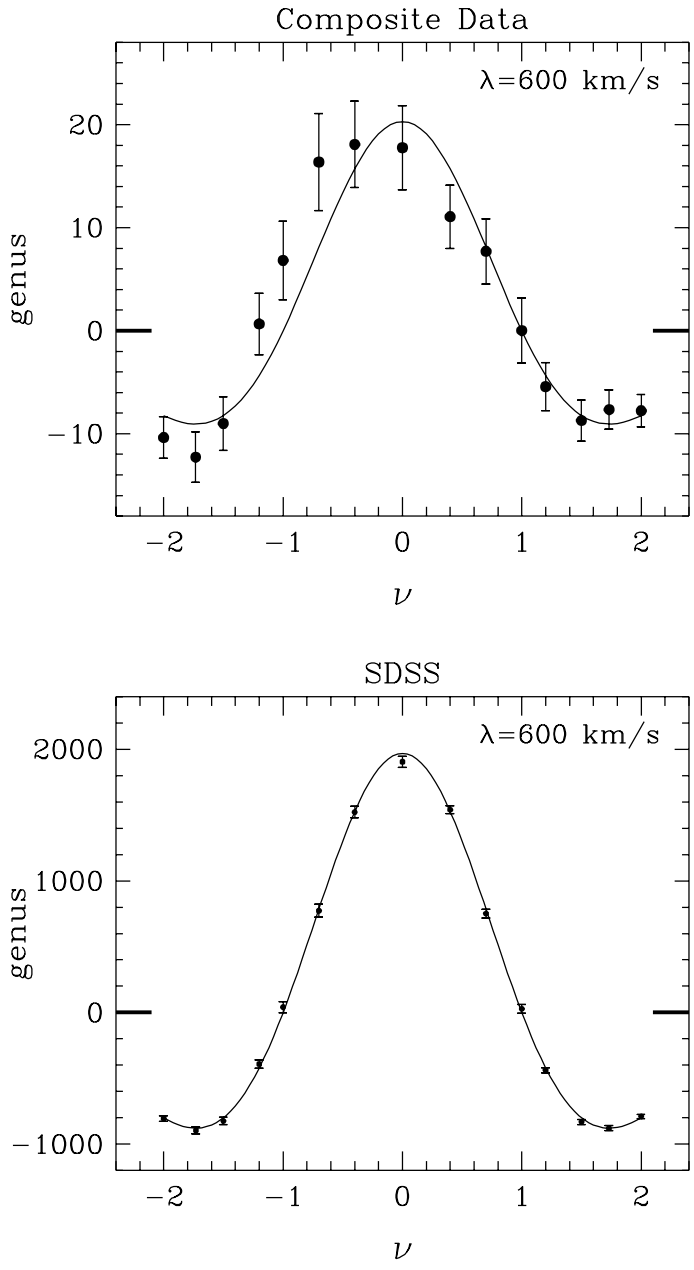

Figure 3.1.10 Quantitative measurement of the topology of the galaxy distribution. The upper panel, adapted from Figure 12 of Gott *et al.* (1989), shows the combined genus curve measured from four large, complete redshift surveys, at a smoothing length $\lambda = 600$ km s^{-1}. Error bars are computed by bootstrap methods. The lower panel illustrates the expected precision of genus measurements from the SDSS redshift sample. The solid curve has the universal form predicted for Gaussian density fields, with an amplitude appropriate to the SDSS survey volume assuming a standard CDM power spectrum. Points are distributed about this curve in accordance with the expected error bars, which we extrapolate from the Gott *et al.* sample assuming that they scale with the square-root of the number of detected structures. High-precision topology measurements will allow clear detection of small deviations from the predictions of Gaussian models.

Gaussian prediction. The combination of this large volume with dense sampling will yield a measurement of the small scale genus curve which has a factor of ten higher signal-to-noise ratio than previous measurements (see Figure 3.1.10). Thus these topological analyses will provide tight constraints on primordial fluctuations, independent of those obtained from count probability statistics.

Other Statistical Techniques

The above discussion touches on only a few of the possible methods to quantify galaxy clustering and test cosmological models. Many other statistics, designed to address a variety of physical questions, have been proposed for describing large scale structure. These include statistics that measure the alignment of galaxies and clusters with their parent superclusters, techniques that quantify the presence of voids, sheets, and filaments, wavelet decompositions of the galaxy density field, and measurements of the fractal properties of the galaxy distribution. Reconstruction methods, wherein one works backward from the current galaxy distribution to the initial density fluctuations, provide an alternative way of addressing many of the questions posed in Section 3.1.1, complementing the purely statistical approach (Weinberg 1992; Nusser and Dekel 1993; Gramman 1993). The SDSS will provide a superb database for applying many of these techniques. We anticipate that other powerful methods will be developed over the next few years, particularly as people think about how to exploit data of the scope and quality provided by the SDSS.

Clustering of Different Galaxy Types

One of the most exciting features of the SDSS is that we can study clustering in detail for different classes of galaxies. It is well known that the fraction of elliptical and S0 galaxies rises dramatically in the cores of rich clusters (Dressler 1980; Postman and Geller 1984; Giovanelli *et al.* 1986). However, it is unclear whether this morphological segregation extends to large scales, and whether it occurs outside of the densest regions. One major obstacle in addressing these issues is the small size of existing samples. Random and systematic errors in the morphological types listed by existing catalogs also pose serious problems (Santiago and Strauss 1992).

With the SDSS we can study small and large scale differences in the clustering of galaxies of different morphological types (not just spiral *vs.* elliptical but Sa *vs.* Sb, barred spiral *vs.* unbarred spiral, etc.), different colors, different luminosities, different surface brightnesses, and different spectroscopic properties (comparing galaxies with starburst spectra to galaxies with old-population spectra, AGN galaxies to "normal" galaxies, and so on). The SDSS photometric sample will be cross-correlated with samples from large–angle surveys in other bands, *e.g.* X–rays from the ROSAT all–sky survey, near–infrared from the Two-Micron All Sky Survey (2MASS; cf., Kleinmann 1992), far–infrared from IRAS, and radio from the VLA B Array survey at 20 cm (Becker, White, and Helfand 1995) (*cf.* Section). We can therefore compare the clustering of garden-variety optical galaxies to that of galaxies that are bright in these various bands. In addition, the SDSS spectroscopic sample will go appreciably deeper than the 2MASS survey, allowing us to define the 2MASS galaxy selection function exquisitely well over the Northern Galactic Cap. This is a necessary ingredient in calculating the angular dipole moment of the (full-sky) 2MASS galaxy sample (cf. Villumsen and Strauss 1987). This dipole moment can be compared with the 600 km/s motion of the Local Group relative to the rest frame of the Cosmic Microwave Background in order to determine Ω_0 (cf., Strauss and Willick 1995).

Two features of the SDSS make it uniquely suited to such studies: the accurate photometric fluxes, high-resolution images, and high signal-to-noise spectra allow us to divide galaxies into classes objectively and with small errors, and the full galaxy sample is large enough

that we can take even rather specialized subdivisions and still measure clustering accurately over a wide range of scales. No other planned redshift survey will match these capabilities.

Clustering studies divided by galaxy type will be particularly valuable for addressing the second category of questions introduced in Section 3.1.1. They will provide important clues about the physical processes that influence galaxy formation and determine galaxy properties. They will also provide crucial information about the relation between the galaxy and mass distributions, thereby helping to separate effects of biased galaxy formation from signatures of primordial fluctuations and cosmological parameters (particularly Ω). As a single example, suppose that we measure correlation functions for several different galaxy types at large scales. If all of these correlation functions have the same shape (perhaps with different amplitudes), we will have excellent reason to believe that this is also the shape of the mass correlation function, and hence a direct constraint on the primordial fluctuations. If the shape varies from one type to another, on the other hand, we might conclude that galaxy formation has been influenced by some long-range process (*e.g.* Bower *et al.* 1993), and that the true shape of the mass correlation function remains uncertain.

3.1.4 *Further Applications*

Distance Indicators and Large Scale Velocities

The high-resolution galaxy spectra provide more than just redshifts for 10^6 galaxies. One can measure other properties of the galaxies from their spectra, including linewidths and velocity dispersions. Combined with the photometric data, this information will allow us to derive distance indicators and therefore study the peculiar velocity field out to very large scales. Two fundamental questions can be addressed using this data set: the rate at which flows, averaged over progressively larger scales, converge to the CMB rest-frame, and the relation between the mass-density field (inferred from the flow pattern within the survey region) and the galaxy-density field obtained from the redshift survey. The answers to these questions provide important information about the amplitude of mass fluctuations on very large scales and about the values of the density parameter and the bias parameter (Strauss and Willick 1995).

Given the resolution of the spectrograph, to what depth and accuracy can peculiar velocities be measured? With a resolving power of ~ 2000 the velocity resolution is ~ 150 km sec^{-1}. As each resolution element is sampled by 3 pixels, velocity dispersions and linewidths can be fitted to an accuracy of 10–20 km sec^{-1} for high signal–to–noise ratio data. With these data, distances can be derived from the $D_n-\sigma$ (Dressler *et al.* 1987), Tully–Fisher (*e.g.*, Pierce and Tully 1988), and Brightest Cluster Galaxy (BCG; Hoessel 1980; Lauer and Postman 1994) relations. The simulations of galaxy spectra in Figure 3.2 indicate that velocity dispersions can be measured for elliptical galaxies to $r' = 17$ and linewidths for spiral galaxies to $r' = 17.5$ (assuming an equivalent width of $\gtrsim 30$Å for late type spirals, [Kennicutt 1992]). A further limitation applies to the Tully–Fisher relation. While Hα rotation curves give Tully-Fisher distances as accurate as those from HI linewidths (Courteau 1992), the data we obtain will be from fiber, not slit spectroscopy. For an Hα linewidth to be meaningful the rotation curve must turn over within the diameter of the fiber. Typically this occurs at 2-3 kpc (from the data of Courteau 1992). Therefore, we will be able to apply the Tully–Fisher relation only for $z > 0.05$, which is the outer limit of existing velocity-field surveys. For more nearby spirals, one can supplement the SDSS multi-color photometry with independent measures of optical or HI linewidths. There exists a distance indicator for BCGs which depends on photometric measurements only (Lauer and Postman 1994), so if one can compensate for the effects of seeing one can measure their distances even without spectroscopic information.

Distance errors of 15% are typical for each of these distance estimators. Assuming a median

redshift of $z = 0.1$ (the sample for which we can measure $D_n - \sigma$ distances will be shallower than this, and the Tully–Fisher and BCG samples deeper), the velocity error on an individual galaxy is 4500 km sec^{-1}. To reduce the error on these velocities and, therefore, detect the signature of the velocity field, we can bin the data in 50 h^{-1} Mpc cells, each containing ~ 400 galaxies suitable for Tully-Fisher measurements, allowing us to reduce the errors to 200 km sec^{-1} per volume. In overdense regions and at lower redshift (as the error per galaxy decreases and the number density of galaxies in the sample increases) we can decrease the cell size dramatically. We estimate that we will obtain over 300 independent, statistically significant measures of the velocity field across the full SDSS survey volume.

Because the SDSS has a large opening angle, the measured peculiar velocity field will be sensitive to the bulk motions detected by existing surveys (Mathewson *et al.* 1992; Courteau *et al.* 1993; Lauer and Postman 1994), and to the convergence of these flows towards the CMB rest-frame. Analyses in radial shells will be able to detect the scale at which convergence occurs. The rate of convergence and the amplitude of the velocity field binned on large scales provide direct constraints on long-wavelength modes of the *mass* power spectrum.

Using the peculiar velocity field within the survey volume, we can apply the reconstruction techniques of Dekel *et al.* (1990) and Ganon and Hoffman (1993) to determine the mass-density field for the volume surveyed. Comparison between the galaxy number densities and mass density derived from the velocities yields a direct measure of the cosmological density parameter and the bias, in the combination $\Omega^{0.6}/b$ (*cf.* Dekel *et al.* 1993), a measure that is independent of that obtained from the anisotropy of clustering in redshift space discussed in Section 3.1.3. The peculiar velocity approach will also require measuring the galaxy density fluctuations on scales comparable to the smoothing of the velocity field, but the uniformity and high sampling density of the redshift survey will make this possible.

A major strength of this data set is the existence of a fully sampled redshift survey containing the same galaxies from which the peculiar velocity data are obtained. Existing studies of peculiar velocities are limited by an incomplete understanding of the biases inherent in proper distance measurements, such as Malmquist bias. As reviewed by Strauss and Willick (1995), convolution of the dispersion in the distance indicator with the real distribution of galaxies leads to spurious components in the velocity field. From the complete redshift survey we will know the number density distribution of galaxies and the selection function used in deriving the velocity field samples. Thus we can correct the peculiar velocity data for these biases in a self-consistent way.

We may even be able to use the redshift-distance relation to put constraints on the curvature of the Universe q_0. The expected effects will be small, but no smaller than those we are trying to measure to get a handle on peculiar velocities. However, as Fisher *et al.* (1992) show, curvature effects and those due to galaxy evolution are expected to be comparable even at low redshifts, and detailed modeling of the colors and spectra will be necessary to disentangle these effects. Evolutionary effects will also have to be taken into account when measuring the peculiar velocities at $z \sim 0.1$.

The discussion here is based on our current understanding of the properties of distance indicators. In the local Universe, the effects of environment, density, and morphology on distance indicators are poorly understood (Bottinelli *et al.* 1982; Silk 1989). The SDSS will be a superb data set with which to study these questions, since it provides spectroscopy and multi-color photometry for a large sample of galaxies. The results will have important implications for peculiar velocity measurements with the SDSS and for existing peculiar velocity surveys. Correlating with the distance-independent parameters available from the photometric data may lead to improvements in the distance indicators (Connolly 1993), or to completely new redshift-independent measures of distance.

While the SDSS galaxy redshift survey will be complete to $r' \approx 18.1$, the photometric data extend 5 magnitudes deeper for the Northern sample and 7 magnitudes deeper for the Southern strip. We can substantially increase the power of the SDSS by deriving approximate galaxy redshifts from the broad-band colors. Photometric redshifts are accurate enough for basic measures of the evolution of the galaxy population and of galaxy clustering. Deriving approximate redshifts to the limit of the photometric survey will increase the number of objects available for study of the spatial distribution of galaxies from 10^6 to $5{\times}10^7$.

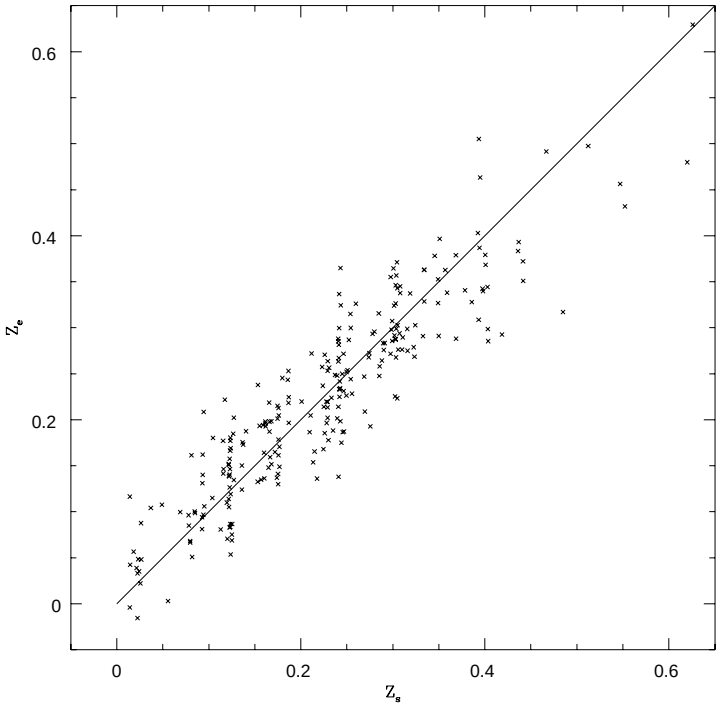

Figure 3.1.11 Estimating redshifts from photometry. The figure shows the estimated redshift *vs.* the spectroscopic redshift for the Koo and Kron UJFN photographic data, to $B_J = 22.5$. The formal dispersion is $\Delta z = 0.047$, largely dominated by photometric errors. The underestimate of galaxy redshifts at faint magnitudes is due to Malmquist bias. From Connolly *et al.* (1995).

Deep multi–color photographic data in the selected areas SA57 and SA68 (*cf.* Koo and Kron 1992) have shown that the distribution of galaxies in the four-color space U, B_J, R_F, I_N is almost planar (Connolly *et al.* 1995). In this plane, lines of constant redshift and lines that trace the luminosity evolution of galaxy spectral types are almost orthogonal. The dispersion in the color-redshift relation is $\Delta z = 0.047$ to $B_J = 22.5$ (Figure 3.1.11), where the scatter is already dominated by photometric errors. We have modeled the relation given the signal-to-noise ratio of the SDSS and find that the dispersion reduces to $\Delta z = 0.02$ for $z < 1$, especially for the reddest galaxies (cf. Section 3.5). The photometric survey of the Southern strip will yield many galaxies at $z \gtrsim 1$.

Calibration of the color–redshift relation requires a homogeneous sample of galaxies with

spectroscopic redshifts that covers the magnitude and redshift range of the photometric survey. For $r' < 18.1$ the complete distribution of galaxies in color space is determined from the SDSS spectroscopic survey. One can extrapolate to higher redshifts by simply applying K–corrections to the galaxy colors, but it is far better to calibrate the color–redshift relation directly with faint galaxy redshift surveys, which currently extend to $B = 24$ (*e.g.* Glazebrook *et al.* 1995). These data will be superseded by the Keck DEEP redshift survey (Mould 1993; *cf.* Appendix), which will obtain 20,000 redshifts to a depth of $B = 25.5$.

Given a sample size of $N \approx 5 \times 10^7$ with distances accurate to $\Delta z = 0.02$, we can study the evolution of the galaxy luminosity function and the galaxy correlation function with unprecedented accuracy. The latter will be particularly interesting given recent results (Efstathiou *et al.* 1991) implying that the faint blue population shows anomalously weak clustering, perhaps implying that this is a quite distinct population. The Southern survey, in particular, will go deep enough to address these issues. The results of these analyses will allow us to test predictions of a wide range of galaxy formation theories, including detailed scenarios like standard cold dark matter and more phenomenological models involving mergers or interactions.

Identifying Clusters and Superclusters

In an optical survey like the SDSS, galaxies provide the basic markers of large scale structure. Nonetheless, it is useful to group galaxies into clusters and superclusters, and to consider these objects as entities in themselves, for several reasons. First, clusters allow one to trace structure to larger scales and higher redshifts, even without a complete galaxy redshift sample, via the cluster-cluster correlation function and similar measures. Second, it is interesting to study the properties of galaxies in cluster and supercluster environments, especially at high redshifts. Finally, the properties of clusters and superclusters, and the evolution of these properties, themselves provide important constraints on theories of structure formation. For example, cluster mass-to-light ratios and the distribution of cluster velocity dispersions depend strongly on the amplitude of mass fluctuations and the values of Ω and b. Substructure and evolution of cluster properties are expected to be strong in $\Omega = 1$ cosmologies and weaker in open cosmologies.

For several decades, studies of clusters and superclusters have relied heavily on Abell's (1958) catalog, based on visual examination of photographic plates. Abell identified clusters by counting galaxies within a specified magnitude range and metric radius. Recent cluster catalogs based on visual examination (Abell, Corwin, and Olowin 1989) and digitized Schmidt plates (Dodd and MacGillivray 1986; Dalton *et al.* 1992) have used variants of Abell's criterion.

The SDSS galaxy redshift catalog will have a median depth similar to that of the Abell cluster catalog. Within this volume, we can identify clusters using complete redshift-space information, greatly reducing contamination effects. We will obtain dozens or even hundreds of galaxy redshifts in these clusters, allowing accurate measures of cluster properties. State-of-the-art algorithms for identifying clusters from photometric data use matched filters in position-magnitude space (Postman *et al.* 1995). The SDSS photometric survey will provide a superb database for application of such techniques. Multi-color photometry will be extremely useful here, as it provides redshift information that can be used to reduce contamination from foreground and background galaxies (cf. the previous subsection). We can improve cluster identification in both the redshift and photometric catalogs by using color and morphology data to select early-type galaxies, which are much more common in cluster cores than in the field. This approach will be especially important at high redshifts, where every reduction in contamination greatly increases sensitivity. As discussed earlier and in Section 3.5, we plan to identify the most luminous red ellipticals, and target them for spec-

troscopy a magnitude fainter than the main galaxy survey, taking advantage of the fact that they tend to have complex spectra with strong absorption lines. This sample will include the majority of brightest cluster galaxies, and in this way, we will obtain a cluster redshift survey that extends to much greater depth than the galaxy redshift survey, at almost no cost in fibers. The high-redshift clusters will be interesting targets for follow-up observations with other telescopes, and we will want to correlate the SDSS cluster catalog with the Abell catalog and ROSAT and other X-ray surveys. Further aspects of cluster science are discussed in Section 3.5.

To date, most catalogs of superclusters have been created by applying some sort of percolation algorithm to a cluster catalog (Huchra and Geller 1982; Bahcall and Burgett 1986). The depth and high sampling density of the SDSS will allow us to identify superclusters directly from the galaxy redshift survey, using objective techniques. We can, for instance, apply Bayesian reasoning to the problem of identifying groups of cells that lie above overdensity thresholds in redshift space. Initially, we wish to classify cells as lying above or below the threshold, and given the local sampling rate (a function of distance), we can find the optimal threshold so that the ratio of erroneous classifications to correct classifications is minimized. We can then account for correlations between nearby cells using an iterative algorithm in which neighboring cells "interact" with a strength proportional to the correlation, pulling anomalous low-density cells in a high-density environment up, and vice versa. Small fluctuations are suppressed, and when the procedure reaches a steady state one is left with coherent groups of cells that represent superclusters. The whole operation can be repeated for various thresholds. Thus we obtain an optimal reconstruction of the largest scale coherent structures.

QSOs and Absorption Systems – Relation to Large Scale Structure

The QSO redshift survey will allow us to probe clustering over enormous length scales and at high redshifts. The formation epoch of the first quasars, the evolution of the quasar luminosity function, and the clustering of quasars provide important tests for theories of structure formation (*e.g.* Efstathiou and Rees 1988; Turner 1991). These and other aspects of quasar science are discussed in Section 3.7.

From the point of view of large scale structure, one of the most exciting by-products of the QSO survey will be an enormous catalog of absorption line systems, both metal-line absorbers and Lyman α clouds. In effect, each QSO provides its own pencil-beam redshift survey of absorbers. Unlike a flux-limited galaxy catalog, the selection function for finding absorbers does not drop rapidly with redshift; in fact it increases, making absorbers excellent markers for the analysis of structure at high redshift (Crotts *et al.* 1995; Quashnock *et al.* 1996).

The interpretation of absorption line clustering is somewhat more complicated than that of galaxy clustering, since the number density of absorbers can be affected by the local state of the intergalactic medium (IGM) and by fluctuations in the metagalactic radiation field as well as by underlying fluctuations in the mass density. Despite the wealth of observational data on Lyman α absorbers, their precise nature and formation mechanism remain uncertain; several models have been proposed (*e.g.* Ikeuchi and Ostriker 1986; York *et al.* 1986; Bond *et al.* 1988; Rees 1986; Hernquist *et al.* 1996). The continuity of HI column densities up to and beyond $10^{17}\,\mathrm{cm}^{-2}$ suggests that the clouds extend in mass all the way up to Lyman-limit systems (Tytler 1987) and damped systems (Wolfe 1983), which are often thought to be related to forming galaxies. However, clustering in the Lyman α forest appears to be quite weak at high redshifts. It is not clear whether this reflects strong evolution of clustering in the mass distribution, or whether it just indicates that the clouds are less strongly clustered than the mass. Recent results find weak clustering signals in the Lyman α forest (Kulkarni *et*

al. 1995; Songaila and Cowie 1996), favoring the former explanation (Lanzetta *et al.* 1995; Zhang *et al.* 1995, 1996).

Improved data on absorbers at high and low redshifts, and the increasing power of cosmological simulations with gas dynamics, should help resolve these questions over the next few years. We anticipate that clustering of absorption lines will become one of the testable predictions of theories of structure formation and models of the IGM, a particularly important prediction because it applies to the high-redshift Universe. The SDSS sample of absorption line systems will be unique, 1000 times larger than anything in existence today (see Section 3.7). The quasar survey will probe over 100,000 pencil beams, whose lengths are determined by the redshifts of the quasars and by the locations of Lyman-limit systems that block part of the continuum. These pencil beams fill, sparsely but quite uniformly, a volume whose linear dimension is measured in Gpc. The statistical noise will be extremely small, enabling us to detect even weak clustering. The high density of beams will allow us to measure the power spectrum using pairs of objects on separate lines of sight, thus obtaining a well-behaved window function (Szalay *et al.* 1991); this approach offers substantial advantages over clustering analyses that only consider absorbers along the same line of sight. The resolution of the SDSS spectrographs will not be as high as that of some other absorption line studies, making identification of weaker Lyman α lines difficult. Press *et al.* (1993) and Zuo and Bond (1993) propose correlating the quasar spectra in the flux regime before any lines are identified, which will largely circumvent this problem. Cosmological simulations like those of Cen *et al.* (1994), Zhang *et al.* (1995) and Hernquist *et al.* (1996) can yield theoretical predictions for flux distributions and correlations incorporating instrumental resolution and noise, thus allowing direct comparison to the immense data base that the SDSS quasar spectra will provide. This type of analysis will result in an accurate picture of large scale structure at high redshift ($z = 1.5$ to 5), complementing photometric redshifts in this régime (Lanzetta *et al.* 1996).

3.1.5 *General Remarks*

Galaxy redshift surveys have provided spectacular maps of the large scale galaxy distribution and taught us a great deal about structure in the Universe. They have stimulated, and begun to answer, the questions with which we introduced this chapter. New redshift surveys, some ongoing, some planned for the future, will extend our knowledge to larger scales, and they will complement the SDSS redshift survey in important ways, by probing structure at higher redshifts and by providing maps of the nearby galaxy distribution that cover the whole sky.

The SDSS will make an exceptionally powerful contribution to the study of large scale structure because of its size, its uniformity, and the high quality of its photometric and spectroscopic data. These will allow galaxy clustering in the present-day Universe to be measured with unprecedented precision and detail. As we have emphasized throughout this section, precision and detail are the keys to addressing important physical questions, whether constraining the matter content from subtle features near the peak of the power spectrum, determining Ω from the anisotropy of clustering in redshift space, searching for small departures from Gaussian initial conditions to identify the physical origin of primordial fluctuations, or comparing clustering of different galaxy types to test models of biased galaxy formation. The design of the SDSS has to a large extent been motivated by these questions, and specifically by the goal of measuring galaxy clustering via the statistical methods described in Section 3.1.3. Other investigations, like measurements of large scale peculiar velocities or photometric redshifts, are possible because of the resulting, general purpose design. The breadth and quality of the SDSS data make it capable of addressing a wide range of questions about large scale structure, including questions we have not yet thought to ask. They also make it a powerful database for addressing many other astronomical issues, as discussed in the remainder of this chapter.

References

Abell, G.O. 1958, *Ap. J. Suppl.*, **3**, 211.

Abell, G.O., Corwin, H.G., & Olowin, R.P. 1989, *Ap. J. Suppl.*, **70**, 1.

Alimi, J.-M., Blanchard, A., & Schaeffer, R. 1990, *Ap. J. (letters)*, **349**, L5.

Babul, A., & White, S.D. M. 1991, *M. N. R. A. S.*, **253**, 31P.

Bahcall, N.A., & Burgett, W.S. 1986, *Ap. J. (letters)*, **300**, L35.

Balian, R., & Schaeffer, R. 1989a, *Astr. Ap.*, **220**, 1.

Balian, R., & Schaeffer, R. 1989b, *Astr. Ap.*, **226**, 373.

Baugh, C. M., & Efstathiou, G. 1993, *M. N. R. A. S.*, **265**, 145.

Becker, R.H., White, R.L., & Helfand, D.J. 1995, *Ap. J.*, **450**, 559.

Bennett, C.L. *et al.* 1994, *Ap. J.*, **436**, 423.

Bernardeau, F. 1992, *Ap. J.*, **392**, 1.

Bond, J.R., Szalay, A.S., & Silk, J. 1988, *Ap. J.*, **324**, 627.

Bottinelli, L., Gouguenheim, L., & Paturel, G., 1982, *Astr. Ap.*, **113**, 61.

Bouchet, F.R., Schaeffer, R., & Davis, M. 1991, *Ap. J.*, **383**, 19.

Bouchet, F.R., Strauss, M.A., Davis, M., Fisher, K.B., Yahil, A., & Huchra, J.P. 1993, *Ap. J.*, **417**, 36.

Bower, R. G., Coles, P., Frenk, C.S., White, S.D.M. 1993, *Ap. J.*, **405**, 403.

Broadhurst, T.J., Ellis, R.S., Koo, D.C., & Szalay, A.S. 1990, *Nature*, **343**, 726.

Cen, R., Miralda-Escudé, J., Ostriker, J.P., & Rauch, M. 1994, *Ap. J. (letters)*, **437**, L9.

Cole, S., Fisher, K.B., & Weinberg, D.H. 1994, *M. N. R. A. S.*, **267**, 785.

Cole, S., Fisher, K.B., & Weinberg, D.H. 1995, *M. N. R. A. S.*, **275**, 515.

Colombi, S., Bouchet, F. R., & Schaeffer, R. 1994, *Astr. Ap.*, **281**, 301.

Colombi, S., Bouchet, F. R., & Schaeffer, R. 1995, *Ap. J. Suppl.*, **96**, 401.

Connolly, A.J. 1993, Ph.D. Thesis, University of London.

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

Courteau, S. 1992, PhD. Thesis, University of California, Santa Cruz.

Courteau, S., Faber, S.M., Dressler, A., & Willick, J. 1993, *Ap. J. (letters)*, **412**, L51.

da Costa, L. N., Pellegrini, P., Sargent, W., Tonry, J., Davis, M., Meiksin, A., Latham D., Menzies, J., & Coulson, I. 1988, *Ap. J.*, **327**, 544.

da Costa, L. N., Vogeley, M. S., Geller, M. J., Huchra, J. P., & Park, C. 1994, *Ap. J. (letters)*, **437**, L1.

Crotts, A.P.S., Melott, A.L., York, D.G., & Fry, J.N. 1995, *Phys. Lett. B*, **155B**, 251.

Dalton, G.B., Efstathiou, G., Maddox, S.J., & Sutherland, W.J. 1992, *Ap. J. (letters)*, **390**, L1.

Davis, M., & Peebles, P.J.E. 1983, *Ap. J.*, **267**, 465.

Dekel, A., Bertschinger, E., & Faber, S. M. 1990, *Ap. J.*, **364**, 349.

Dekel, A., Bertschinger, E., Yahil, A., Strauss, M., Davis, M., & Huchra, J. 1993, *Ap. J.*, **412**, 1.

de Lapparent, V., Geller, M.J., & Huchra, J.P. 1986, *Ap. J. (letters)*, **302**, L1.

de Lapparent, V., Geller, M.J., & Huchra, J.P. 1991, *Ap. J.*, **369**, 273.

Dodd, R.J., & MacGillivray, H.T. 1986, *A. J.*, **92**, 706.

Dressler, A. 1980, *Ap. J.*, **236**, 351.

Dressler, A., Faber, S.M., Burstein, D., Davies, R.L., Lynden-Bell, D., Terlevich, R.J., & Wegner, G. 1987, *Ap. J.*, **313**, L37.

Efstathiou, G., Bernstein, G., Katz, N., Tyson, J.A., and Guhathakurta, P. 1991, *Ap. J. (letters)*, **380**, L47.

Efstathiou, G. & Rees, M.J. 1988, *M. N. R. A. S.*, **230**, 5P.

Feldman, H., Kaiser, N., & Peacock, J. 1994, *Ap. J.*, **426**, 23.

Fisher, K.B., Strauss, M.A., Davis, M., Yahil, A., and Huchra, J.P. 1992, *Ap. J.*, **389**, 188.

Fisher, K.B., Davis, M., Strauss, M.A., Yahil, A., and Huchra, J.P. 1993, *Ap. J.*, **402**, 42.

Fisher, K.B., Davis, M., Strauss, M.A., Yahil, A., and Huchra, J.P. 1994, *M. N. R. A. S.*, **267**, 927.

Fisher, K. B., Huchra, J. P., Davis, M., Strauss, M. A., Yahil, A., & Schlegel, D. 1995, *Ap. J. Suppl.*, , 100, 69.

Frieman, J., & Gaztañaga, E. 1994, *Ap. J.*, **425**, 392.

Fry, J.N. 1984, *Ap. J.*, **279**, 499.

Fry, J.N., & Gaztañaga, E. 1993, *Ap. J.*, **413**, 447.

Ganon, G. & Hoffman, Y. 1993, *Ap. J. (letters)*, **415**, L5.

Gaztañaga, E. 1992, *Ap. J. (letters)*, **398**, L17.

Gaztañaga, E. 1994, *M. N. R. A. S.*, **269**, 913.

Geller, M.J., & Huchra, J.P. 1989, *Science*, **246**, 897.

Giovanelli, R., Haynes, M.P., & Chincarini, G. 1986, *Ap. J.*, **300**, 77.

Glazebrook, K., Ellis, R., Colless, M., Broadhurst, T., Allington-Smith, J., & Tanvir, N. 1995, *M. N. R. A. S.*, **273**, 157.

Gott, J.R., Weinberg, D.H., & Melott, A.L. 1987, *Ap. J.*, **306**, 341.

Gott, J.R., *et al.* 1989, *Ap. J.*, **340**, 625.

Gramann, M. 1993, *Ap. J.*, , 405, 449.

Hamilton, A.J.S. 1992, *Ap. J. (letters)*, **385**, L5.

Hamilton, A.J.S. 1993, *Ap. J. (letters)*, **406**, L47.

Hernquist, L., Katz, N., Weinberg, D.H., & Miralda-Escudé, J. 1996, *Ap. J. (letters)*, **457**, L51.

Hoessel, J.G. 1980, *Ap. J.*, **241**, 493.

Holtzman, J.A. 1989, *Ap. J. Suppl.*, **71**, 1.

Huchra, J., Davis, M., Latham, D., & Tonry, J. 1983, *Ap. J. Suppl.*, **52**, 89.

Huchra, J. and Geller, M.J. 1982, *Ap. J.*, **257**, 423.

Ikeuchi, S., & Ostriker, J.P. 1986, *Ap. J.*, **301**, 522.

Kaiser, N. 1986, *M. N. R. A. S.*, **219**, 785.

Kaiser, N. 1987, *M. N. R. A. S.*, **227**, 1.

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

Kennicutt, R.C. 1992, *Ap. J.*, **388**, 310.

Kirshner, R.P., Oemler, A., Schechter, P.L., & Shectman, S.A. 1981, *Ap. J. (letters)*, **248**, L57.

Kleinmann, S. G. 1992, in Robotic Telescopes for the 1990's, edited by A. Filippenko, ASP Conference Series, 34 (San Francisco: Astronomical Society of the Pacific), 203.

Koo, D., & Kron, R. 1992, *Ann. Rev. Astron. Astrophys.*, **30**, 613.

Kulkarni, V., Huang, K.-L., Green, R.F., Bechtold, J., Welty, D., & York, D.G. 1996, *M. N. R. A. S.*, **279**, 218.

Landy, D.S., Shectman, S.A., Lin, H., Kirshner, R.P., Oemler, A.A., & Tucker, D. 1996, *Ap. J. (letters)*, **456**, L1.

Lanzetta, K.M., Bowen, D.V., Tytler, D., & Webb, J.K. 1995, *Ap. J.*, **442**, 538.

Lanzetta, K.M., Yahil, A., & Fernandez-Soto, A. 1996, *Nature*, **381**, 97.

Lauer, T., & Postman, M. 1994, *Ap. J.*, **425**, 418.

Lawrence, A. *et al.* 1995, in preparation.

Loveday, J., Efstathiou, G., Peterson, B., A., & Maddox, S. J. 1992, *Ap. J.*, **400**, 43.

Loveday, J., Maddox, S. J., Efstathiou, G., & Peterson, P. A. 1995, *Ap. J.*, **442**, 457.

Maddox, S.J., Efstathiou, G., Sutherland, W.J., & Loveday, J. 1990, *M. N. R. A. S.*, **242**, 43P.

Mathewson, D.S., Ford, V.L., & Buchhorn, M. 1992, *Ap. J. (letters)*, **389**, L5.

Maurogordato, S., Schaeffer, R., & da Costa, L.N. 1992, *Ap. J.*, **390**, 17.

Meiksin, A., Szapudi, I., & Szalay, A. 1992, *Ap. J.*, **394**, 87.

Moore, B., Frenk, C., Weinberg, D., Saunders, W., Lawrence, A., Rowan-Robinson, M., Kaiser, N., Efstathiou, G., & Ellis, R.S. 1992, *M. N. R. A. S.*, **256**, 477.

Mould, J.R. 1993, ASP Conference Series 43, 281.

Nusser, A., & Dekel, A. 1993, *Ap. J.*, **405**, 437.

Park, C., Vogeley, M.S., Geller, M.J., & Huchra, J.P. 1994, *Ap. J.*, **431**, 569.

Peacock, J. A., & Dodds, S. J. 1994, *M. N. R. A. S.*, **267**, 1020.

Peebles, P.J.E. 1980, The Large Scale Structure of the Universe (Princeton: Princeton University Press).

Peebles, P.J.E., & Groth, E.J. 1976, *Astr. Ap.*, **531**, 31.

Petrosian, V. 1976, *Ap. J. (letters)*, **209**, L1.

Pierce, M.J., & Tully, R.B. 1988, *Ap. J.*, **330**, 579.

Postman, M.P. & Geller, M.J. 1984, *Ap. J.*, **281**, 95.

Postman, M.P., Lubin, L., Gunn, J.E., Oke, J.B., Hoessel, J.G., Schneider, D.P., & Christensen, J.A. 1995, *A. J.*, **111**, 615.

Postman, M.P., & Lauer, T.R. 1995, *Ap. J.*, **440**, 28.

Press, W.H., Rybicki, G.B., & Schneider, D.P. 1993, *Ap. J.*, **414**, 64.

Quashnock, J., VandenBerk, D.E., & York, D.G. 1996, *Ap.J.(letters)*, in press.

Ramella, M., Geller, M.J., & Huchra, J.P. 1990, *Ap. J.*, **353**, 51.

Rees, M.J. 1986, *M. N. R. A. S.*, **218**, 25p.

Sachs, R.K., & Wolfe, A.M. 1967, *Ap. J.*, **147**, 73.

Santiago, B.X., & Strauss, M.A. 1992, *Ap. J.*, **387**, 9.

Santiago, B. X., Strauss, M. A., Lahav, O., Davis, M., Dressler, A., & Huchra, J. P. 1996, *Ap. J.*, **461**, 38.

Sargent, W.W., & Turner, E.L. 1977, *Ap. J. (letters)*, **212**, L3.

Saunders, W., Frenk, C. S., Rowan-Robinson, M., Efstathiou, G., Lawrence, A., Kaiser, N., Ellis, R. S., Crawford, J., Xia, X.-Y., & Parry, I. 1991, *Nature*, **349**, 32.

Shectman, S.A., Landy, S.D., Oemler, A., Tucker, D.L., Lin, H., Kirshner, R.P., & Schechter, P.L. 1996, *Ap. J.*, **470**, 172.

Silk, J. 1989, *Ap. J. (letters)*, **345**, L1.

Smoot, G.F., Bennett, C.L., Kogut, A., Wright, E.L., Aymon, J., Boggess, N.W., Cheng, E.S., De Amici, G., Gulkis, S., & Hauser, M.G. 1992, *Ap. J. (letters)*, **396**, L1.

Songaila, A., & Cowie, L. 1996, *A. J.*, **112**, 335.

Strauss, M.A., Huchra, J.P., Davis, M., Yahil, A., Fisher, K.B., & Tonry, J.P. 1992, *Ap. J. Suppl.*, **83**, 29.

Strauss, M.A. & Willick, J.A. 1995, Physics Reports, **261**, 271.

Szalay, A.S., Broadhurst, T.J., Ellman, N., Koo, D.C., & Ellis, R. 1991, *Proc. Natl. Acad. Sci.*, **90**, 4858.

Szapudi, I., Meiksin, A., & Nichol, R.C. 1996, *Ap. J.*, **473**, 15.

Szapudi, I., & Szalay, A. 1993, *Ap. J.*, **408**, 43.

Szapudi, I., & Szalay, A. 1996, *Ap. J.*, **459**, 504.

Tegmark, M. 1995, *Ap. J.*, **455**, 429.

Turner, E.L. 1991, *A. J.*, **101**, 5.

Tytler, D. 1987, *Ap. J.*, **321**, 49.

Villumsen, J. V., & Strauss, M. A. 1987, *Ap. J.*, **322**, 37.

Vogeley, M.S., Park, C., Geller, M.J., Huchra, J.P., 1992, *Ap. J. (letters)*, **391**, L5.

Vogeley, M.S., Park, C., Geller, M.J., Huchra, J.P., & Gott, J.R. 1994, *Ap. J.*, , 420, 525.

Weinberg, D.H. 1992, *M. N. R. A. S.*, **254**, 315.

White, S.D.M. 1979, *M. N. R. A. S.*, **186**, 145.

Wolfe, A.M, 1983, *Ap. J. (letters)*, **268**, L1.

York, D.G., Dopita, M., Green, R.F., & Bechtold, J. 1986, *Ap. J.*, **311**, 610.

Zhang, Y., Anninos, P., & Norman, M.L. 1995, *Ap. J. (letters)*, **453**, L57.

Zhang, Y., Meiksin, A., & Norman, M.L. 1996, *Ap. J.* in press.

Zuo, L., & Bond, J.R. 1994, *Ap. J.*, **423**, 73.

3.6. Clusters of Galaxies

3.6.1 *Introduction*

Clusters of galaxies, the largest virialized systems known, are key tracers of the large–scale structure in the Universe and are critical tools in constraining cosmological models. By

studying their mass, distribution and evolution, we obtain vital insights into the formation history of the underlying mass field: a fundamental goal of cosmology. Theoretically, clusters provide an opportunity to study the universe in the linear clustering regime, thereby enabling an analytical approach to the calculation of structure evolution (*e.g.* Mann *et al.* 1993). Observationally, galaxy clusters can be identified and classified to large lookback times ($z \sim 1$), thus allowing comparison studies with the nearby Universe. At present, some of the evidence for a rapidly evolving universe comes from clusters of galaxies (*e.g* Henry *et al.* 1992).

There is little doubt that the SDSS survey will revolutionize cluster research. It will provide much needed advances in systematic studies of clusters, currently limited by the unavailability of modern, accurate, complete, and objectively selected catalogs, and by the limited photometric and redshift information for those that do exist. The three main catalogs of rich clusters, those of Abell (1958), Zwicky *et al.* (1961–1968), and Abell *et al.* (1989) were obtained from visual inspection of the Palomar Sky Survey and the ESO Survey. Even with this inaccurate procedure, and with only a relatively small number of cluster redshifts (*cf.* Struble and Rood 1991), the catalogs have contributed significantly to our understanding of large scale structure (Bahcall and Soneira 1983, Postman *et al.* 1992). Recent advances include wide-angle catalogs selected by objective algorithms from digitized photographic plates (Dodd and MacGillivray 1986, Picard 1991, Dalton *et al.* 1992, Lumsden *et al.* 1992) and from deep CCD images (Postman *et al.* 1995). These efforts have raised some questions about the integrity of the previous visually–constructed catalogues, but the new catalogues are not yet extensive, and we cannot make conclusive statements about completeness (Nichol *et al.* 1992, Nichol & Connolly 1995). The SDSS photometric survey will solve this by covering, in total, an area 3 times greater that the APM (Dalton *et al.* 1992) survey and being nearly as deep as the Postman *et al.* (1995) CCD survey (which covered only five square degrees). Further, it will be in five colors and be the foundation for over 10^6 spectra! The SDSS will also allow the systematic study of the relationship between the X-ray emitting gas and cluster properties, and of the evolution of the cluster population, through the spectroscopy of $\sim 10^5$ luminous red galaxies to z ~ 0.5. In the following section, we outline the expected scope of the SDSS with respect to clusters of galaxies.

3.6.2 *The SDSS Cluster Catalogs*

The digital photometric and redshift catalogs produced by the main SDSS survey will permit the selection of galaxy clusters using well-defined, automated algorithms. The SDSS can be thought of as producing four different, yet parallel, catalogs of overdensities from the $\sim 10^6$ galaxy redshifts and 5×10^7 galaxies in the photometric data. The digital nature, and the completeness, of the data will allow us to catalogue a continuous range of galaxy overdensities from rich clusters down to poor groups. We ask the reader to bear this in mind while reading the following sections even though we may neglect to explicitly discuss this tremendous dynamic range. Below, we summarize the four SDSS cluster catalogues:

a. 3D Selection

Using the $\sim 10^6$ galaxies with measured redshifts, we will be able to construct a catalogue of overdensities seen in 3D. This will transcend all present concerns of projection effects in optical catalogues and allow us to confidently map the local large–scale structure in clusters. The number of clusters we expect to detect in 3D can be predicted using the Abell catalogue as our template. This catalog is generally thought to be complete for richness class $R \geq 1$ clusters out to $z \sim 0.08$ (Huchra *et al.* 1990). In this volume, the SDSS spectroscopic galaxy survey will produce redshifts for galaxies a magnitude fainter than L_* in nearby clusters (Colless 1989), so there will be many galaxies (≥ 100) per $R \geq 1$ cluster in the spectroscopic sample. The number of redshifts per cluster decreases with cluster redshift, down to a few

redshifts per $R \geq 1$ cluster at z $\sim$ 0.2. We can thus expect this shallower SDSS catalog, identified *in galaxy redshift space*, to be nearly as deep as Abell's catalog. The surface density of rich clusters ($R \geq 1$) in the survey is expected to be ~ 0.1 per square degree for $z \lesssim 0.2$. This translates to $\sim$ 1000 clusters some of which have $\geq$ 100 redshift measurements per cluster, which will be an unparalleled data set of great scientific merit.

A potential problem concerning the 3D selection of clusters is the 55 arcsecond minimum spacing of the spectroscopic fibers (Chapter 11), which will allow only one galaxy of a pair closer than this to be spectroscopically observed per plug plate. However, as we discuss in Chapter 12, we propose to distribute the spectroscopic fields to allow more than one observation of dense regions, a scheme called 'adaptive tiling', and in most cases, the second member of a close pair in a dense region, i.e. a cluster, could be targeted in the second pass.

b. 2D selection

The 5 color photometry is predicted to be complete for galaxies to a magnitude limit of $r' = 22$ (Chapter 5). This is substantially deeper than current photographic plate surveys like the APM and EDSGC (Collins *et al.* 1992), and represents a wonderful opportunity to statistically catalogue galaxy overdensities out to intermediate redshifts ($z < 0.4$). Much current research shows that this may be a dynamic epoch of cluster, and galaxy, evolution.

We will implement automated cluster–finding algorithms to identify clusters in the photometric survey. The wide wavelength coverage of the SDSS photometric data will allow us to search for overdensities in position-magnitude-color space. This will greatly improve the efficiency of cluster identification by reducing foreground and background contamination, which is especially important at high redshift. Furthermore, we can tie these photometric cluster catalogs to those identified solely from the spectroscopic data, thus providing a valuable check on the validity of the photometric identification procedures.

Any cluster-finding algorithm is of necessity biased towards galaxy distributions of a certain shape and profile, and we are working on developing techniques that can efficiently find clusters with strong deviations from circular symmetry and/or pronounced substructure, properties which low-density, "unrelaxed" clusters like Hercules often exhibit. We will certainly want to apply a variety of cluster-finding algorithms to the photometric data. At present, we are investigating two well-motivated algorithms that could be run simultaneously, thereby reducing the overall biases in the final selection. These are the matched–filter approach similar to that of Postman *et al.* (1995) and further developed by Kepner *et al.* (1996), and a wavelet–based method (Slezak *et al.* 1990). The former involves convolving the data with a filter constructed from a physical cluster model, with the angular galaxy distribution fit to a King profile and the magnitudes fit to a redshifted Schechter luminosity function. The wavelet algorithm entails smoothing the binned galaxy data with a mexican–hat function, which can be thought of as a series of local background–subtracting bandpass filters. We are presently testing and comparing results from both of these using large simulations. These methods will of course also be used to find overdensities in the spectroscopic sample.

The use of photometric redshifts (Connolly *et al.* 1995) gives us the tantalizing opportunity to carry out a full pseudo–3D selection of clusters from the entire SDSS photometric survey. To facilitate finding clusters, we will construct from the photometric data, as they are reduced, a compressed, coarsely–binned data 'cube' describing the structure of the photometric galaxy survey, consisting of the position of each galaxy, its r' magnitude and its photometric redshift (see Chapter) which can be easily searched for the density enhancements corresponding to clusters. Methods for searching for clusters in pseudo-3D space are also under development by Kepner *et al.* (1996).

Once more, we can use existing cluster catalogues to gauge the size of this 2D catalogue. The EDSGC finds the surface density of clusters and groups found by looking for galaxy overdensities, to be $\simeq 0.5$ per square degree (Lumsden *et al.* 1992) complete down to $m_{10}(b_j) = 18.5$ (magnitude of the tenth brightest cluster member). Considering that our galaxy catalog needs to be defined to ≈ 2 mags fainter than m_{10} to identify clusters, we will be able to identify clusters to $m_{10}(r') \simeq 20.5$. This yields a surface density of ~ 5 clusters and groups per square degree, producing a catalogue of $\sim 10^4$ rich clusters, an order of magnitude larger than any other catalogue in existence. The smaller groups will be harder to detect at high redshift. A further option will be to co–add the imaging data in the individual SDSS bandpasses and search for angular galaxy overdensities complete to over half a magnitude fainter than this.

There is an alternative approach to cluster finding, pioneered by Dalcanton (1996) and Zaritsky *et al.* (1996). They use the drift-scan data of Schneider, Schmidt, & Gunn (1994) to identify extended low-surface-brightness objects; some of these are associated with low-surface brightness galaxies (Chapter 3.8), while others are associated with clusters too faint for the individual galaxies to be identified. Zaritsky *et al.* estimate that there are two high-redshift ($z \sim 1$!) clusters per square degree in the Schneider *et al.* data. The Northern SDSS survey is somewhat deeper than these data, and of course will be in five colors, and the Southern survey will be appreciably deeper, so we can look forward to several tens of thousands of high-redshift cluster candidates for follow-up on large telescopes.

c. Southern Survey

A major objective of the SDSS is to carry out a deep photometric and spectroscopic survey in the southern hemisphere. This survey is discussed in detail in Section 5.5. These data will be analyzed in the same fashion as the main SDSS data and will push our cluster catalogues to even higher redshift. The state-of-the-art for high-redshift cluster searches is the deep CCD survey of Postman *et al.* (1995), carried out on the Palomar 5-m telescope. This survey, a major observational undertaking with current instruments, goes to about 24th magnitude in two bands and covers five square degrees. The SDSS southern survey will go a full magnitude deeper than this, in *five* bands, and will cover about 300 square degrees. The resulting cluster catalog will be a superb data set for the study of cluster evolution, extending the redshift baseline out to $z \sim 1$.

d. Spectroscopy of a Distant, Volume Limited Sample of High Luminosity Elliptical Galaxies

The SDSS plans to obtain redshifts for a sample of the $\sim 10^5$ most luminous red galaxies (hereafter BRGs) between $r' \sim 18^m - 19.5^m$, a depth $\sim 1.5^m$ fainter than that of the main galaxy sample. These objects are frequently located at the centers of high–density regions (clusters), and can be used to measure the volume -limited cluster distribution to $z \sim 0.5$, greatly extending the scale on which large-scale structure can be measured and providing direct information on evolution for many thousands of clusters of galaxies. BRGs are metal-rich, strong-lined elliptical galaxies, and the SDSS spectra should easily be able to provide redshifts to the limit $r' = 19.5^m$ (see Chapter 3.8).

BRGs have intrinsically high luminosities, and therefore sample a much greater volume than does the main galaxy survey. In the main galaxy sample, a limit of $r' \sim 18^m$ corresponds to a redshift of ~ 0.17 for L^* galaxies; however, BRGs one magnitude fainter than this are at a redshift of ~ 0.45 at their mean luminosity. The BRG sample thus measures a comoving volume about four times larger than that measured by BRGs in the main galaxy sample. The sampling statistics are such that the power spectrum can be well determined to scales which overlap significantly those measured by COBE and expected to be measured by MAP (see Figure 3.1.9).

This sample of luminous red elliptical galaxies includes the brightest galaxies in clusters *but is not limited to them*. It has been known for many years that galaxies like BRGs are found in small groups and poor clusters as well as in the richest, highest density clusters. Recent work on the optical identification of ROSAT sources has turned up objects like these in large numbers (see, for example, Figure 3.5.1). The X-ray luminosity of the extended X-ray source shown in this figure, RX J0018.8+1602, is 1.3×10^{44} erg s^{-1}, similar to typical values for Abell-type optically rich clusters. The bright galaxy near the X-ray centroid is almost certainly associated with the X-ray emission; the nearby optical point source shown in Figure 3.5.1 is unlikely to be the X-ray emitter since a spectrum taken with the ARC 3.5 meter telescope shows it to be an r = $19.5^{\rm m}$ G or K star.

The ARC images show that the giant elliptical galaxy at the position of RX J0018.8+1602 has a cD-type radial profile, an r magnitude of 19.2 and a color g - r = 3.0± 0.4 (giving a rest-frame color of g-r = 0.9, Frei and Gunn 1994). The nearby radio source 54w084 - also at z = 0.54 - has a similar color.

Both RX J0018.8+1602 and 54w084 are likely candidates for the SDSS BRG sample, and illustrate the power of this sample for finding clusters of galaxies which otherwise would have been missed by standard cluster finding algorithms. This example also shows that the SDSS sample is ideal for finding distant X-ray clusters and probing cluster evolution to significant redshifts.

These luminous red galaxies are rare, but not nearly so rare as are the great clusters; estimates of their number indicate that they are plentiful enough to provide powerful leverage for large scale structure but are distinct and rare enough that they will not overwhelm the main galaxy sample. The objects in this extended sample are well represented in the main galaxy sample as well, and there will be more than enough information to determine their bias relative to the bulk of the galaxies in the main sample.

The key to the definition of this extended sample is the introduction of the photometric redshift as a selection criterion, without which it would be difficult or impossible to reject lower luminosity objects which would overwhelm the sample. Connolly *et al.* (1995) have shown that accurate colors obtained from photometry in several bands, plus the apparent magnitude, can measure the redshift to an accuracy of 0.03 or better for most galaxies (see Chapter 3.1). It is a happy coincidence that it is precisely for these very red objects that the photometric redshifts are best determined, and redshift errors of at worst 0.02 should be possible from photometry alone. With a photometric redshift, we can estimate the *absolute* magnitude of the galaxy, and target only the highest luminosity galaxies at a given apparent magnitude. It might be noted that this class of objects is at the red edge of the color distribution, and one achieves much the same result by targeting only the *reddest* galaxies at a given apparent magnitude, since earlier galaxian type and greater redshift both make objects redder. The photometric redshift, however, offers a much more precise criterion.

The exact size of the sample selected this way depends on how tightly it is defined and on some as yet very poorly known astronomical parameters. We can estimate this number as follows. The luminosity functions of both early- and late-type galaxies are similar and are described roughly by a Schechter (1976) function with an alpha of about -1. (Note, however, that *real* brightest cluster galaxies in very rich clusters are known *not* to be part of the general Schechter function, but their numbers are small enough that they are unlikely to affect our estimate very much.) The numbers in the visible bands grow roughly like $10^{0.4m}$, or inversely as the flux density. Thus the distribution of *absolute* magnitudes at a given *apparent* magnitude is the luminosity function weighted by one power of the flux, or, roughly, a pure exponential. Therefore, at a given absolute magnitude, about 1/e = 37% of the sample is brighter than L^*. The mean magnitude of the brightest cluster galaxies is 1.8

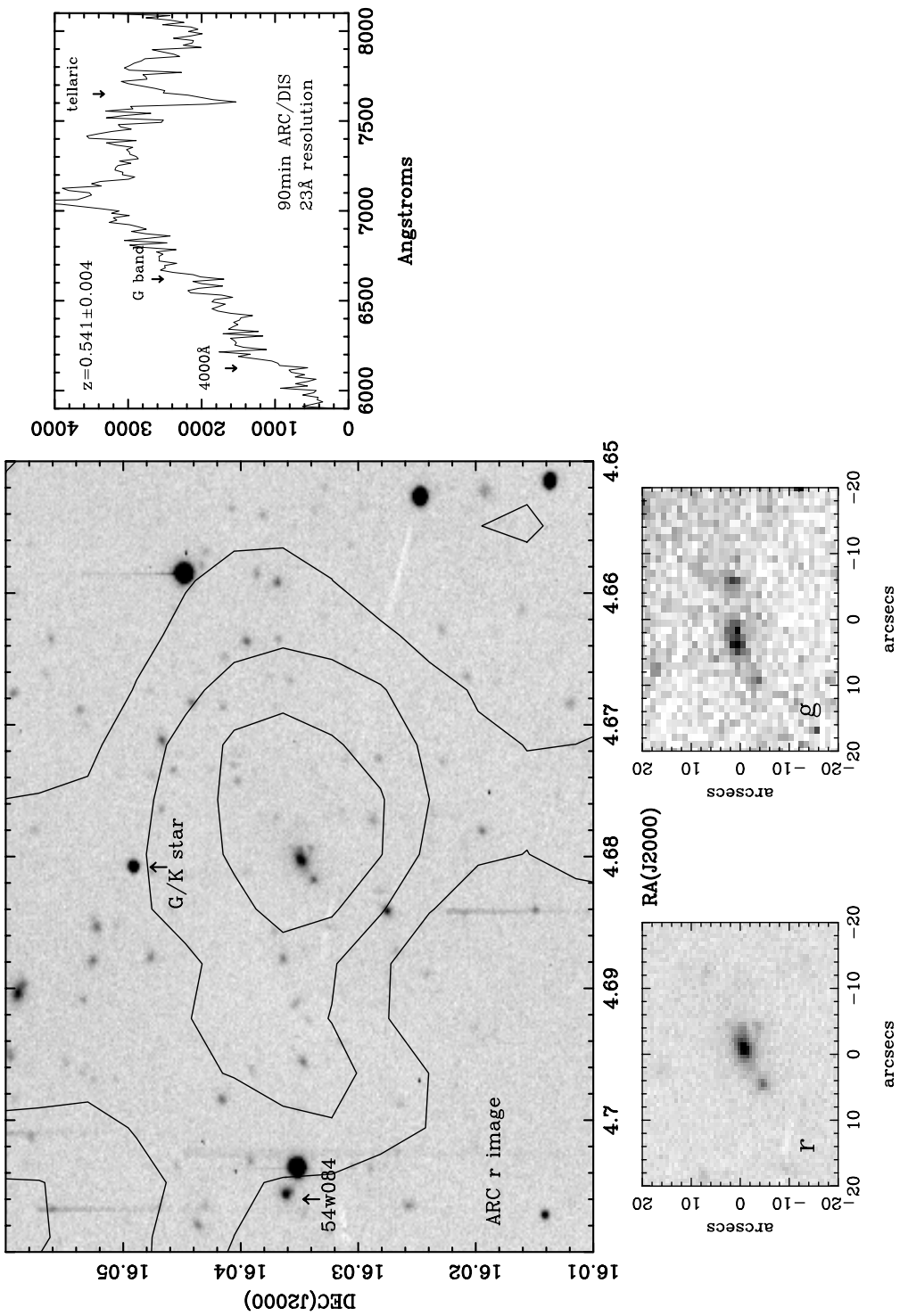

Figure 3.6.1 Data taken at the ARC 3.5 m telescope for the X-ray source RX J0018.8+1602. This source was serendipitously detected in the deep ROSAT PSPC pointing towards the (in)famous cluster of galaxies CL0016+16 at z = 0.546. The X-ray source is coincident with a *single* giant elliptical galaxy which is at the same redshift as CL0016+16 and is surrounded by a host of much fainter galaxies Connolly *et al.* 1996).

magnitude brighter than L^*, and they have a dispersion of about 0.35 magnitudes (and the dispersion is somewhat smaller for metric magnitudes, Lauer and Postman 1994; Postman and Lauer 1995). Since we want to be sure to include nearly all the brightest cluster galaxies, we want to include objects with luminosities brighter than about 1.1 magnitudes above L^*; these represent about 12% of the sample. We also make a cut in color, and restrict the sample to the reddest galaxies. In order not to bias the sample, one needs to be relatively generous, but the dispersion in the intrinsic colors is small, about 0.05 or less in B-V. With generous errors in colors introduced by the uncertainty in the photometric redshift, it would appear that accepting the reddest 30% at a given photometric redshift is adequate to ensure that we have no worse than a 2-sigma cut in the color of the bright ellipticals. Thus at a given apparent magnitude, the sample of bright ellipticals is about 3.5% of the total. We wish, however, to go fainter. To reach a BRG with the mean luminosity of these objects at $z = 0.5$ we need to reach a brightness level of $r' = 20.4$ in the 3 arcsecond fiber, about 1.5 magnitudes fainter than the typical main sample galaxy. There are four times as many objects in this sample than at the cutoff of the main sample, but one-fourth of those are already in the main sample, so we have three times as many *additional* objects, or 10% of the main sample. Simulations indicate that redshift-quality spectra for bright ellipticals at $r'(3'') = 20.4$ are readily obtainable, and that, indeed, one could even go significantly fainter: simulations for a z = 0.5 galaxy 0.7 magnitudes fainter still produce spectra which would yield good redshifts (see Chapter 3.8).

This sample will be accommodated by reducing the main galaxy sample by about 10%, or about $0.1^{\rm m}$ in limiting magnitude. What do we gain in return? The volume of the new sample is four times that of the main sample, and indeed the sample is very nearly volume-limited, so that it does not thin out nearly so severely as the main sample does at large distances. Estimates of the shot-noise contribution to the power spectrum $P(k)$ for this sample indicate that the measurements become shot-noise limited only for scales larger than one Gigaparsec h^{-1}, and at that scale the shot noise and cosmic variance together contribute an error of about 20% to $P(k)$ (see Figure 3.1.9). This scale is well beyond the peak in the power spectrum predicted by CDM models and overlaps well into the range explored by COBE. The window function of the main sample restricts us to scales more than a factor of two smaller. Indeed, these are scales on which the cluster correlation function (to say nothing of the properties of the clusters themselves) is expected to evolve appreciably, and quantitative measurement of this evolution has the potential to distinguish strongly between various cosmological models (e.g., Peacock 1996). The bias of the BRG sample, which is known to exist and be large, can be adequately explored within the main sample. The redshift to which the BRG sample reaches encompasses the range in which there are indications that there are strong evolutionary effects in cluster properties, and some of these may be seen within the sample. It will also allow a comparison of the evolution of optical and X-ray properties of clusters; there are indications that there may be little evolution in X-ray properties of clusters to z = 0.5, even though significant evolution in the optical properties is seen over this redshift range (Nichol *et al.* 1996). The sample will be an important bridge to the cluster survey in the south, which will reach objects with (photometric) redshifts in excess of unity, and to the scales probed by the quasar absorption lines (see Chapter 3.7).

The sample can also be used to study the evolution of stellar populations in these red galaxies. Combining the spectra in bins of redshift will give spectra with very high signal - to - noise ratios up to z = 0.5 or so, and detailed population synthesis models can be fit. There will be some 20,000 spectra of nearly identical systems between redshifts 0.4 and 0.5; the composite of these, even divided into several different categories, should be breathtaking.

Finally, since this sample is a safe *superset* of the brightest cluster galaxies, it greatly simplifies the cluster target selection process; these objects are targeted on the basis of their photometric properties alone (and the photometric redshifts are calculated from a simple

quadratic expression in the colors, so their determination does not place a burden on the target selection software) and do not depend on finding associated density maxima. The clusters working group has identified a way of finding density enhancements using a simple catalog structure constructed by the target selection pipeline, as described above; we still wish to construct such catalogs for use in other investigations, but it would not be necessary to run a separate piece of code between target selection and tiling in order to calculate the density maxima.

3.6.3 *The Combination of SDSS and X–ray Data*

With advances in satellite instrumentation, X–ray studies of clusters of galaxies have recently provided important new information on the general cluster population. For example, ROSAT has proved to be one of the most successful X–ray satellites for cluster studies both in its pointing and survey phases. Deep ROSAT PSPC cluster pointings have shown that a large fraction of clusters have substructure and complicated internal dynamics, indicative of recent evolution. In the optical, such studies have been hampered by the need for large numbers of cluster redshift measurements. Also, the ROSAT pointings have shown that the X–ray emitting gas comprises the largest fraction of the baryonic mass in clusters (greater than or equal to the sum of all the galaxies) and that, in total, there is more baryonic matter than expected from nucleosynthesis if $\Omega = 1$. In parallel, the ROSAT All-Sky Survey (RASS) is now producing objective catalogues of X–ray clusters that are challenging the present optical catalogues in terms of volume sampled and reduced systematics like projection effects. Romer *et al.* (1994) recently presented the first results from such a catalogue showing that their large–scale distribution was quantitatively similar to that of the optical clusters.

It is clear to us that a marriage of the X–ray and optical data on clusters of galaxies will be extremely beneficial to both parties. On a practical front, optical clusters suffer from projection effects where the alignment of galaxies along the line–of–sight can give rise to a phantom cluster. At low redshift, this will not be a problem for the SDSS due to the spectroscopic survey; however, above $z \simeq 0.1$ this problem will start to surface again. Projections are less severe in X–rays and the coincidence of an X–ray source (especially if extended) with an optical overdensity is a powerful indication of a true mass concentration.

For X–ray cluster searches the major problem is mis–classification. Clusters represent only $\sim 10\%$ of all X–ray sources detected, and at faint flux limits clusters are indistinguishable from the dominant AGN/star populations (Chapter). Previous optical follow–up of X–ray sources (Gioia *et al.* 1990) has been laborious and painful, usually requiring both deep imaging and high–quality spectroscopy. *The severity of these problems cannot be overstated, especially at intermediate to high redshifts ($z > 0.15$)* - a small error in the identification of stars/AGNs as clusters/groups will give rise to large contamination in the subsequent X–ray cluster candidate catalogues. Therefore, having a full optical database, with 5 colors and substantial spectrophotometry for each X–ray source, will be an incredible advantage and will, in almost all cases, provide an unambiguous candidate for the X–ray emission.

These points are vividly illustrated by two on–going X–ray cluster surveys. First, Romer (1994) has presented details on her work in constructing an objectively–selected cluster catalogue from the RASS. In her thesis, she defines a sample of 486 possible X–ray clusters in a region of sky centered on the South Galactic Pole. This catalogue is a combination of the available digitized optical photographic plates and the ROSAT survey data. It has already taken her and collaborators over 3 years, with substantial amounts of 4–meter telescope time, to securely identify 129 of these candidates: 15% of them are AGNs, stars and normal galaxies. Furthermore, an additional 10% of the candidates have resolved X–ray emission (a good indication of cluster X–ray emission) but do not coincide with an optical galaxy overdensity on the photographic plates. Deep imaging of a subset of these

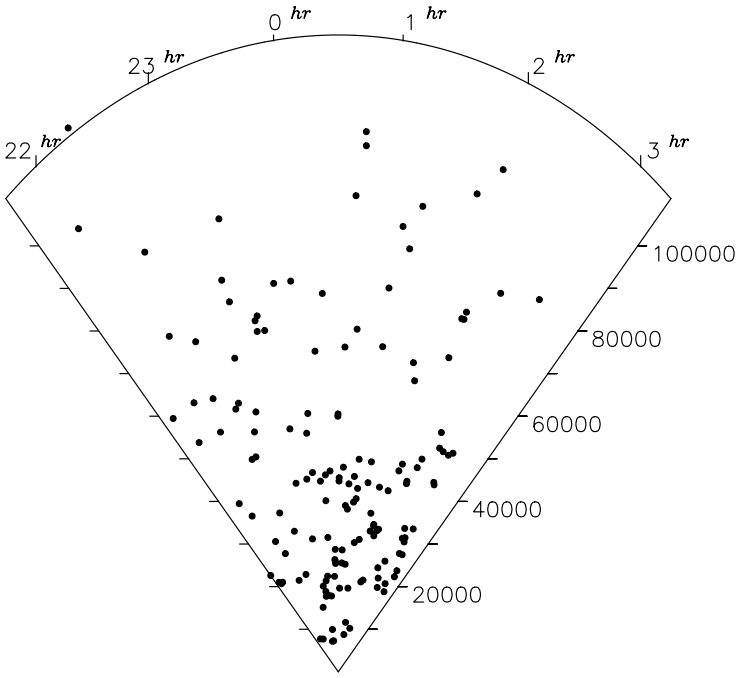

Figure 3.6.2 Right ascension cone diagram of X-ray clusters. The plot shows the redshift distribution of 154 clusters selected from the ROSAT All–Sky Survey (Romer 1994). The declination direction has been compressed and ranges from 0° to −50°. The Sculptor Supercluster (Guzzo *et al.* 1992) is easily visible at $cz \simeq 35000\,km/s$ and dominates the nearby large–scale structure. This supercluster is probably one of the most massive structures presently known. The degree of statistical clustering seen in this sample agrees very well with previous optical results (Nichol *et al.* 1992). The incompleteness beyond $cz \simeq 40000\,km/s$ is due not to the X–ray data, but to the incompleteness of optical identifications and redshift measurements, since the existing photographic catalogues do not go deep enough.

has uncovered several rich clusters at $z \geq 0.3$ (Romer; private communication). Finally, the cluster candidate list was drawn from the superset of $\simeq 5000$ X–ray sources detected in total over her region. Presently, little is known about the number of clusters lost in this larger X–ray source list. The redshift distribution of these clusters is shown in Figure 3.6.2. The distribution is very incomplete beyond $cz \simeq 40,000\ km\ s^{-1}$ because the existing optical cluster catalogues, based almost entirely on photographic plates, are not deep enough. As an example of the advances we can expect to make with the SDSS, Figure 3.5.3 shows an r' band image at the position of an extended ROSAT source observed with the ARC 3.5 m telescope (Nichol *et al.* 1996). A faint cluster, well below the limit of the available photographic plate surveys, is clearly visible.

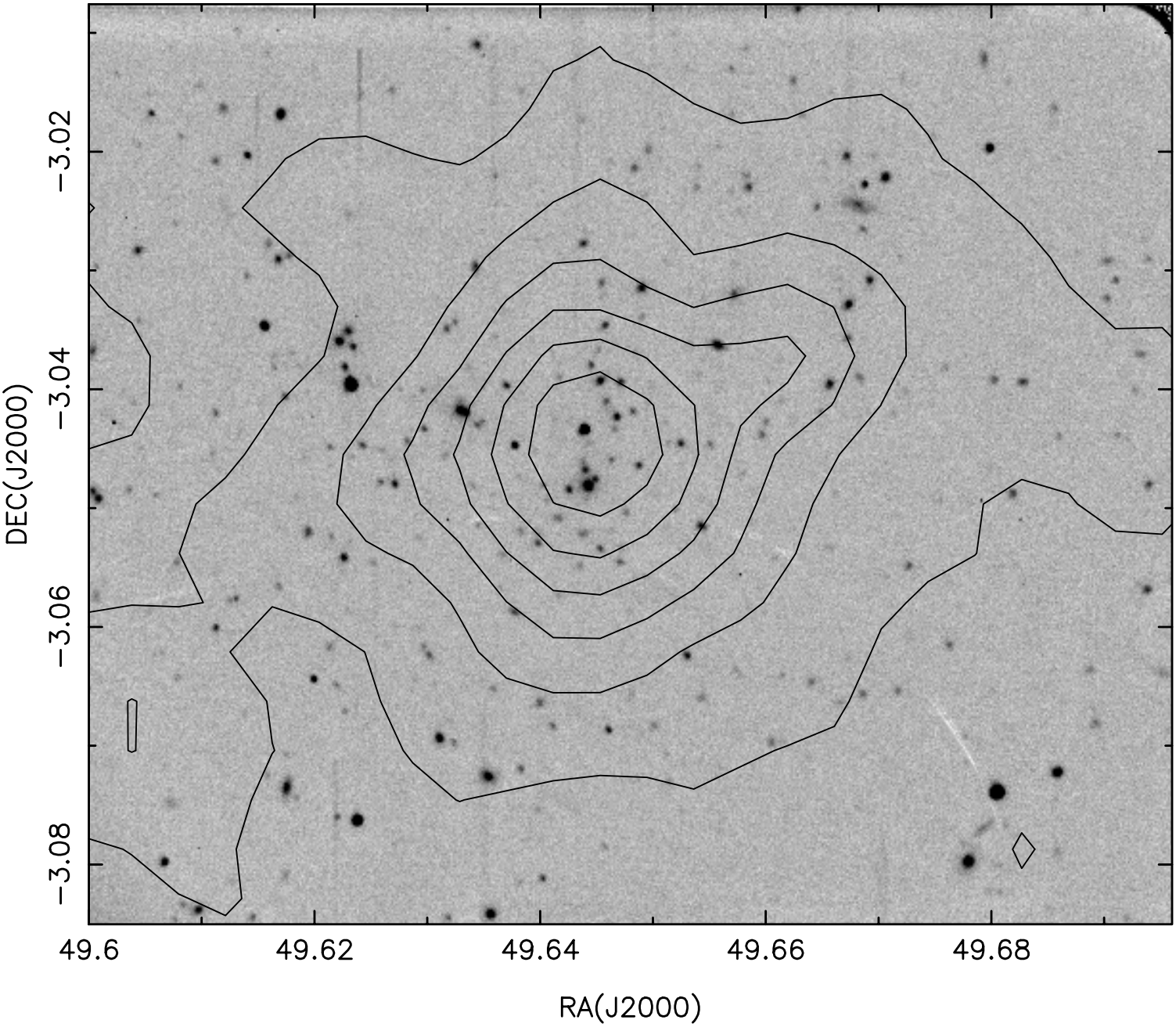

Figure 3.6.3 Discovery of an X-ray cluster. The gray-scale image shows a cluster of galaxies at the position of an extended ROSAT source imaged at r' band using the Digital Imaging Spectrograph on the ARC 3.5 m telescope. The X-ray contours are from ROSAT observations, and the limiting magnitude of the optical image is about $22.5^{\rm m}$ (from the Serendipitous High Redshift Archival ROSAT Cluster Survey; Nichol *et al.* 1996).

Second, several groups are trying to improve upon the work of the Einstein Medium Sensitivity Survey (EMSS; Gioia *et al.* 1990) and search for clusters serendipitously detected in the ROSAT PSPC pointing archive. The ROSAT International X–ray and Optical Survey (RIXOS) has been the first to announce their results, finding a huge discrepancy in the number of clusters detected compared to the EMSS, an order of magnitude fewer (Castander *et al.* 1995). This is remarkable considering that RIXOS samples nearly 3 times more volume than the EMSS. More recent results from the other groups suggest that this discrepancy may be due to object mis–classifications in both of these surveys (Rosati *et al.* 1995, Nichol *et al.* 1996).

Clearly, the optical follow–up of all X–ray sources is ultimately the only way to construct a complete and robust subsample of X–ray clusters. However, using conventional methods, this could take decades. For instance, the WGACAT catalogue produced by N. White at GSFC, which is a database of all point sources detected in over 3000 PSPC ROSAT pointing observations, contains $\simeq 68,000$ sources. To date, *only 10% have been identified*, and the rate is a strong function of X–ray flux. This catalogue, even though it is biased towards X–ray point sources, could contain $\simeq 5000$ X–ray clusters to $z < 1$. The SDSS will provide the identification and classification of tens of thousands of ROSAT sources, redshifts for thousands of X-ray clusters, a search for extended X-ray halos without optical clusters, optical clusters without X-ray halos, the dependence of X-ray emission on structure, redshift, richness, galaxy morphology mix, - - - the list goes on. The structure and evolution of clusters, superclusters and large–scale structure made possible by the combination of SDSS, ROSAT and WGACAT is certain to provide major progress in this field.

3.6.4 *SDSS Cluster Science*

In this final section, we explore the scientific advances in cluster research expected from the whole SDSS dataset. We will highlight areas that we, as a collaboration, will expect to pursue over the course of the SDSS. Obviously, when the data become available to all, it is certain that much more and more ingenious cluster research will be extracted from this huge database. We will also address some of the scientific rewards of combining the SDSS with other databases like the ROSAT catalogues. This will undoubtedly lead to a fuller understanding of the cluster population since the galaxies and the intracluster gas constitute most of the baryonic mass in clusters.

a. Cosmological Implications

Clusters of galaxies present an efficient method of mapping the distribution of matter on the largest possible scales (Bahcall 1988). Present large cluster redshift surveys still sample larger volumes of space than galaxy redshift surveys and provide the only observations at, or near, the expected turnover in the underlying density power spectrum ($\sim 100\,h^{-1}$ Mpc; see Figure 3.5.4). An example of the present status of cluster redshift surveys is shown in Figure 3.6.2. The SDSS will put such studies on a much firmer basis because of its larger volume (possibly two orders of magnitude more cluster redshifts), its objective identification methods, and the use of cluster-finding algorithms that are less subject to projection contamination than Abell–type algorithms. These conclusions will be further strengthened with the combination of the optical and X–ray data (see Figure 3.6.2 for example).

The high amplitude of the cluster correlation function $\xi_{cc}(r)$ was one of the earliest pieces of statistical evidence for strong clustering on very large scales (Bahcall & Soneira 1983 and Figure 3.6.4) and still remains in conflict with most, though not all, popular models of galaxy formation (Bahcall & Cen 1993, Croft & Efstathiou 1994). Kaiser (1984) suggested that the amplification of ξ_{cc} relative to the galaxy correlation function reflects the tendency of clusters to form near high peaks of the primordial density fluctuations; this idea later

Figure 3.6.4 The correlation functions derived from three large cluster redshift surveys. The data are from Romer *et al.* 1994 (X–ray; Sample 6), Postman *et al.* 1992 (Abell clusters) and Dalton *et al.* 1992 (APM clusters). This demonstrates the power of cluster correlations for probing the large–scale structure on scales near the turnover in the power spectrum of the underlying density field at $\sim 100\,h^{-1}$ Mpc.

became the inspiration for theories of biased galaxy formation. With the SDSS we will be able to measure ξ_{cc} on large scales free from the systematic effects which bedevil existing samples (*cf.* Olivier *et al.* 1990). For example, the "shallow" SDSS catalog, identified solely in redshift space, will itself be large enough to yield a robust measure of ξ_{cc} at nearby redshifts ($\sim$ 1000 clusters). We will also have a much larger catalog of clusters, identified in the galaxy photometric data with cluster redshifts obtained in the spectroscopic survey, which will yield very accurate correlation measurements and allow us to examine the dependence of ξ_{cc} on cluster properties such as richness, velocity dispersion, morphology, and X–ray luminosity (*cf.* Bahcall & West 1992). Beyond this, the BRG sample will have the required redshift

baseline to search for evolution in the clustering of clusters out to $z \sim 0.5$. This latter observation may give us our clearest view yet of the true underlying large–scale structure.

On larger scales, catalogs of nearby superclusters, constructed using Abell clusters with known redshifts, have revealed the largest structures presently known in the universe; $\sim 150h^{-1}$ Mpc (Bahcall & Soneira 1984). There are also claims for still larger systems, $\sim 300h^{-1}$ Mpc (Tully 1987), but these are still controversial (Postman *et al.* 1989). Existing galaxy redshift surveys probe a much smaller volume than do the cluster catalogs, but to the extent they can be compared, the superclusters identified from the cluster distribution match up well with structures in the galaxy distribution; the Perseus-Pisces supercluster (Haynes & Giovanelli 1988), and the Great Wall (Geller & Huchra 1989) are both clearly seen in the distribution of clusters and superclusters. The SDSS redshift survey will clarify the relationship between the cluster distribution and the full galaxy distribution. The larger cluster sample should yield definitive answers about structure at very large scales. Cluster peculiar velocities, obtained directly from $D_n - \sigma$ measurements (Lynden-Bell *et al.* 1988) and indirectly from the anisotropy of the cluster correlation function in redshift space (Chapter 3.1), will provide information about the internal dynamics of superclusters.

Finally, it has been known for a long time that brightest cluster galaxies are good standard candles (Sandage 1961; cf., Section 3.6.2). This was more recently confirmed in a very detailed study of the photometry of 100 of these objects by Lauer and Postman (1994), showing that the dispersion is as small as $0^m.35$ and can probably be reduced significantly by the use of velocity dispersion data. The Hubble diagram given by the few thousand BRGs obtained from the SDSS should demonstrate the uniformity of or find significant nonuniformity in the Hubble flow out to redshifts approaching $z = 0.3$. Distances to the nearby sample of clusters can be obtained by the Tully-Fisher (Tully & Fisher 1977) and $D_n - \sigma$ (Lynden-Bell *et al.* 1988) techniques (cf., Section 3.1.4), thus allowing us to establish the zero point of the BRG Hubble diagram and obtain a global value of the Hubble constant. We will, parenthetically, be able to confirm or deny the reality of the very large-scale flow found by Lauer and Postman (1994) in their study, and extend their techniques to a very much larger sample.

b. Cluster Properties

The complete cluster survey will allow a detailed investigation of intrinsic cluster properties. From the 2-D and 3-D information, we will be able to determine accurately such properties as cluster richness, morphology, density and density profile, core radius, velocity dispersion profile, optical luminosity, and galaxy content (morphological fractions and cD galaxies), and to look for correlations between these properties. These studies can be supplemented with the vast array of data from other passbands, for example, the FIRST 1.4 GHz survey, which covers the same area as the SDSS with 5 arcsecond resolution and a completeness limit of $\simeq 1$ mJy. This survey will provide an exquisite catalogue of radio sources in nearby clusters, as well as aiding in the detection of X–ray bright, distant clusters through the presence of a wide–angle radio source (Nichol *et al.* 1994).

With these quantities in hand, we can study individual clusters in great detail. Measurements of galaxy density and morphology, velocity dispersion, and X-ray emission will shed light on the nature of the intracluster medium and its impact on the member galaxies, and on the relative distribution of baryonic and dark matter (*cf.* White *et al.* 1993a). High surface mass density clusters will be candidates for gravitational lenses; we can search the photometric data for systematically distorted background galaxies, especially in the south, and target these clusters for yet deeper imaging surveys with larger telescopes. These lensing studies will allow us to map out the dark matter distribution within a host of clusters (Tyson *et al.* 1990; Kaiser & Squires 1993).

Although the SDSS dataset will not break new ground in the detailed study of any one cluster, with a uniform database covering an enormous statistical sample of clusters, we can characterize the cluster population statistically with much greater accuracy than has been done before (*cf.* Bahcall 1977, 1997). This statistical description will give us new insights into a number of issues associated with structure formation and evolution:

• We will be able to calculate velocity dispersions for ~ 1000 clusters with $z \lesssim 0.2$, compared to ~ 100 clusters with reliable dispersion measurements available now (Mazure *et al.* 1996). For somewhat nearer clusters we will have well-sampled galaxy density profiles and velocity dispersion profiles; these allow much more accurate cluster mass determinations than those provided by global virial mass estimators, which are the standard recourse when only a small number of galaxy redshifts is available. In addition, some clusters in this sample will undoubtedly possess high quality X–ray spectral information from EXOSAT, ASCA and, in the future, AXAF, to provide a further independent handle on the mass estimates. The distribution of nearby cluster masses puts tight constraints on the mass fluctuation amplitude on cluster mass scales, for any specified value of Ω (Bahcall & Cen 1993, White *et al.* 1993b). The mass and dispersion distributions can also be used to distinguish initially Gaussian and non-Gaussian primordial density fluctuations (Weinberg and Cole 1992; Bartlett *et al.* 1993). Beyond the SDSS, the next generation of large aperture telescopes can be used to observe a sample of SDSS distant clusters, thus measuring the distribution of velocity dispersions, and mass, at higher redshift. Due to the homogeneous nature of the SDSS data, we should be able to confidently compare these distant results to the $z = 0$ universe and, for the first time, quantify any observed evolution in the distribution of cluster velocity dispersions.

• With optical luminosities, galaxy profiles, and velocity dispersion profiles, we can measure accurate mass-to-light ratios for a large sample of clusters. These provide constraints on Ω and the bias parameter of galaxies in clusters (cf., Carlberg *et al.* 1996). The infall patterns at larger separations yield another set of constraints on cluster masses and Ω from scales near the cluster turn-around radius (Kaiser 1987; Regös & Geller 1989). It will be especially interesting to compare these constraints from small and intermediate scales to the results obtained from the analyses of large scale clustering anisotropies (*cf.* Section 3.1.3); a complete theory of galaxy formation will need to explain any observed dependence of bias parameters on scale and environment.

• At present, there are over 400 specific PSPC pointings towards known clusters of galaxies (mostly Abell clusters). Such data can also be supplemented with *Einstein* observations. When combined with the SDSS data, this database will be crucial to studies of the relative distributions of luminous (baryonic) and dark matter, and hence of the nature of the dark matter itself. More powerful constraints will be obtained by using dark matter maps derived from lensing of background galaxies. The ratio of baryonic mass to dark mass can lead to a strong constraints on various cosmological parameters and test our understanding of nucleosynthesis (White *et al.* 1993a).

• The above combined sample of nearby X–ray/optical clusters will also shed light on the formation histories of nearby clusters, since mass elements that have collapsed into them, especially at recent epochs, will remain at higher densities. Such measurements are very hard to perform without the fully–sampled nature of the SDSS spectroscopic and photometric data; substructure may be lost outside a limited field–of–view. This is clearly demonstrated in recent work on the Coma cluster, where the combination of extensive X–ray and optical data highlighted the presence of a large subcluster falling in, or out, of the main overdensity. No longer can Coma be thought of as the archetypical relaxed system since it is obviously involved in recent evolution.

The frequency of such subclustering and non-virial structures in clusters at low redshift

(Geller 1990) tells us whether clusters are dynamically old or young. In a gravitational instability model with $\Omega \approx 1$, structure continues to form today, while in a low-Ω, open universe, clustering tends to freeze out at moderate redshift. Cluster profiles and substructure statistics thus provide a diagnostic for Ω that depends on the history of clustering (Richstone *et al.* 1992), an approach quite different from those described earlier, which depend on measuring the mean mass density.

• In addition to the properties of clusters themselves, the properties of galaxies in clusters are of great interest. It is well known that elliptical and S0 galaxies are more common in rich clusters than in the field, and that the fraction of early type galaxies depends on cluster density (Dressler 1984). It has recently been suggested, however, that a morphology-*radius* relation is a better description of existing data than a morphology-density relation (Whitmore *et al.* 1993). The SDSS will provide a powerful database for examining such issues, going beyond statistical characterization to detailed studies of the effects that might produce morphological segregation. Furthermore, for nearby clusters, galaxies will be classified *spectroscopically* as well as morphologically, and will have excellent colors. The interaction of individual galaxies with the X–ray emitting intracluster medium and tidal fields from their neighbors can be studied in detail using the excellent morphological, photometric, and spectroscopic data. To give just one example, measurements of the Hα equivalent width of spirals as a function of velocity relative to the cluster rest frame and distance from its center may reveal the gas stripping and star formation histories of galaxies which are passing through the intracluster gas for the first time.

• One of the most important contributions the SDSS will make in cluster research will be in the study of cluster evolution. This is particularly important since it is potentially the most powerful diagnostic of the physical model causing structure formation. Current investigations are limited to a small numbers of well–known high-redshift clusters, but an even more serious obstacle is that high and low redshift samples of clusters are selected in very different ways. As a result, it is difficult to know whether high redshift concentrations represent fully-formed clusters, or are high-density cores on which more mass will accrete by the present. The first would be unexpected in an $\Omega \approx 1$ model, but the second would be natural. The SDSS cluster sample will clearly remedy both of these shortfalls. The high and low redshift clusters will be selected from the same data using the same algorithms, so they can be compared directly to each other, and to numerical simulations. The number of clusters will be large; we will have redshifts for BCGs to $z \approx 0.5$, and for more distant clusters we can obtain fairly accurate estimated redshifts from the photometry alone (using the colors and apparent luminosity function of member galaxies.) The southern survey will probe the cluster population out to redshifts of order unity (the i' filter will be most useful for this).

Here we discuss a few examples of cluster evolution studies that will benefit from our data;

• The Butcher–Oemler (1984) effect is the strongest optical evidence for significant cluster evolution out to $z \simeq 0.5$. However, as discussed by Kron (1995), these studies have been based on a rogue's gallery of well–studied clusters: "the clusters that formed the foundations of the Butcher–Oemler effect in 1984 are evidently very exceptional objects – otherwise, objects discovered in the mid–1950's would not continue to appear so prominently in current research efforts". Using only the SDSS database, we will transcend such problems and quantify the Butcher–Oemler effect for thousands of clusters as a smooth function of galaxy richness, redshift and cluster morphology. By combining the SDSS with X–ray data, we can further investigate the relationship between the Butcher–Oemler effect and the density/temperature of the intracluster gas, which still remains relatively unknown. This combined dataset will directly test certain Butcher–Oemler hypotheses; for example, is the increased blue galaxy population due to ram–pressure enhanced star formation in infalling cluster galaxies (Gunn

& Dressler 1988)? If so, one would expect a strong correlation between the X–ray properties of clusters and their corresponding galaxy populations as a function of cluster radius and velocity dispersion. At low redshift ($z < 0.1$), the high–quality SDSS data will allow us to move beyond the simplistic Butcher–Oemler definition of optical cluster color/morphology evolution by basing our studies solely on the spectrophotometric data. At higher redshifts, the five color photometry will be invaluable, since the K-corrections can be accurately computed.

• The observed space density of clusters as a function of redshift can place tight constraints on the evolution of the underlying density field. X–ray observations suggest a rapid decrease in the density of bright X–ray clusters (Henry *et al.* 1992), while at optical wavelengths, the space density appears to remain almost constant out to $z = 0.5$ (Postman *et al.* 1995). At present, many authors simply wish to advocate the merits of one cluster selection method over the other, yet the combination of these two results may be telling us something fundamental about cluster evolution and the enrichment/heating of the intracluster gas. Using the SDSS, we will be able to examine such results and quantify any subtle selection bias at play. If such effects are ruled out, or quantified, we should then be able to de–couple the physics of cluster formation from the underlying cosmological density evolution. Again, this will only be possible via the unique marriage of the SDSS and X–ray data.

• Finally, at intermediate redshifts, there already appears to be a strong correlation between the optical morphology of clusters and their X–ray characteristics. This is illustrated by the work of Gioia and Luppino (1994) and Postman *et al.* (1995): over 90% of the $z > 0.3$ EMSS clusters contain a dominant, central galaxy (cD), while in the optical, the fraction is only $\simeq 25\%$. Furthermore, the fact that a rich optical cluster is *not* an X–ray emitter is equally interesting and can only be determined through the combination of the objective X–ray and optical data. Such phenomena may be related to the above discrepancies in the cluster space densities and a direct indication of their contrasting evolutionary histories: relaxed, X–ray bright clusters may have a greater number of cluster crossing times, thereby enriching the intracluster medium and allowing the formation of a dominant galaxy via cannibalism. We will be able to carry out detailed cluster morphological studies (both photometrically and spectroscopically) testing such hypotheses. For example, do X–ray faint, optically rich systems have a significantly different E/S0 population compared to centrally condensed, X–ray bright clusters? If they do, this would be indicative of different formation epochs since these galaxies are dominated by old stellar systems.

References

Abell, G.O. 1958, *Ap. J. Suppl.*, **3**, 211.

Abell, G.O., Corwin, H.G., & Olowin, R.P. 1989, *Ap. J. Suppl.*, **70**, 1.

Bahcall, N. A. 1977, *Ann. Rev. Astron. Astrophys.*, **15**, 505.

Bahcall, N. A. 1988, *Ann. Rev. Astron. Astrophys.*, **26**, 631.

Bahcall, N. A. 1997, in *Astrophysical Quantities*, ed. A. Cox, AIP Press, in press.

Bahcall, N. A., and Cen, R. 1992, *Ap. J. (letters)*, **398**, L1.

Bahcall, N.A., & Cen, R. 1993, *Ap. J. (letters)*, **407**, L49.

Bahcall, N.A., & Soneira, R.M. 1983, *Ap. J.*, **270**, 20.

Bahcall, N.A., & Soneira, R.M. 1984, *Ap. J.*, **227**, 27.

Bahcall, N.A., & West, M.J. 1992, *Ap. J.*, **392**, 419.

Bartlett, J.G., Gooding, A.K., & Spergel, D.N. 1993, *Ap. J.*, **403**, 1.

Butcher, H., & Oemler, A. 1984, *Ap. J.*, **285**, 426.

Carlberg, R.G., Yee, H.K.C., Ellingson, E., Abraham, R., Gravel, P., Morris, S., & Pritchet, C.J. 1996, *Ap. J.*, **462**, 32.

Castander, F.J. *et al.* 1995, *Nature*, **377**, 39.

Colless, M., 1989, *M. N. R. A. S.*, **237**, 799.

Collins, C.A., Nichol, R.C., & Lumsden, S.L. 1992, *M. N. R. A. S.*, **254**, 295.

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

Connolly, A.J., Szalay, A.S., Koo, D.L., Romer, A.K., Holden, B., Nichol, R.C., & Miyaji, T. 1996, *Ap. J. (letters)*, **473**, L67.

Croft, R.A.C., & Efstathiou, G., 1994, *M. N. R. A. S.*, **267**, 390.

Dalcanton, J.J. 1996, *Ap. J.*, **466**, 92.

Dalton, G.B., Efstathiou, G., Maddox, S.J., & Sutherland, W.J. 1992, *Ap. J. (letters)*, **390**, L1.

Dodd, R.J., & MacGillivray, H.T. 1986, *A. J.*, **92**, 706.

Dressler, A. 1984, *Ann. Rev. Astron. Astrophys.*, **22**, 185.

Frei, Z., & Gunn, J.E. 1994, *A. J.*, **108**, 1476.

Geller, M. 1990, in *Clusters of Galaxies: Space Telescope Symposium Series # 4*, W. Oegerle, M. Fitchett, & L. Danly, eds. (Cambridge: Cambridge University Press).

Geller, M., & Huchra, J.P. 1989, *Science*, **246**, 897.

Gioia, L., Henry, J. P., Maccacaro, T., Morris, S. L., Stocke, J., Wolter, A. 1990, *Ap. J.*, **356**, L35.

Gioia, L., & Luppino, G. 1994, *Ap. J. Suppl.*, **94**, 583.

Gunn, J. E., & Dressler, A. 1988, *Large Scale Structure of the Universe*, IAU Symposium 130, eds. Audouze *et al.*

Guzzo, L., Collins, C. A., Nichol, R. C. & Lumsden, S. L., 1992, *Ap. J.*, **393**, L5.

Gunn, J. E., & Dressler, A. 1988, *Towards Understanding Galaxies at Large Redshift*, eds. Kron & Renzini.

Haynes, M.P., & Giovanelli, R. 1988, in Large Scale Motions in the Universe: A Vatican Study Week, edited by V.C. Rubin and G.V. Coyne, S.J. (Princeton: Princeton University Press), p. 31.

Henry, J. P., Gioia, L., Maccacaro, T., Morris, S. L., Stocke, J., Wolter, A. 1992, *Ap. J.*, **386**, 408.

Huchra, J.P., Henry, J.P., Postman, M., Geller, M.J., 1990, *Ap. J.*, **365**, 66.

Kaiser, N. 1984, *Ap. J. (letters)*, **284**, L49.

Kaiser, N. 1987, *M. N. R. A. S.*, **227**, 1.

Kaiser, N., & Squires, G. 1993, *Ap. J.*, **404**, 441.

Kepner, J.V., Fan, X., Bahcall, N.A., Gunn, J.E., & Lupton, R.H. 1996, in preparation.

Kron, R.G., 1995, "The Butcher–Oemler Effect: Lecture 8", Sante Fe Lecture Series.

Lauer, T. R. & Postman, M., 1994, *Ap. J.*, **425**, 418.

Lumsden, S. L., Nichol, R. C., Collins, C. A., and Guzzo, L., 1992, *M. N. R. A. S.*, **258**, 1.

Lynden-Bell, D., Faber, S. M., Burstein, D., Davies, R. L., Dressler, A., Terlevich, R. J., &

Wegner, G. W., 1988, *Ap. J.*, **326**, 19.

Mann, R.G., Heavens, A., & Peacock, J.A., 1993, *M. N. R. A. S.*, **263**, 798.

Mazure, A. *et al.* 1996, Å310, 31.

Nichol, R.C., Collins, C.A., Guzzo L., & Lumsden S.L., 1992, *M. N. R. A. S.*, **255**, 21P.

Nichol, R.C., & Connolly, A.J., 1995, *M. N. R. A. S.*, **279**, 521.

Nichol, R.C., Romer, K., Holden, B., Ulmer, M. Collins, C. and Burke, D. 1996, *ApJ* (in press).

Nichol, R.C., Ulmer M.P., Kron R.G., Wirth G.D., & Koo D.C. 1994, *Ap. J.*, **432**, 464.

Olivier, S., Primack, J.R., Blumenthal, G.R., Dekel, A., & Stanhill, D. 1990, *Ap. J.*, **356**, 1.

Peacock, J.A. 1996, MNRAS, in press (astroph/9608151).

Picard, A. 1991, PhD. Thesis, California Institute of Technology.

Postman, M., Lubin, L.M., Gunn, J.E., Oke, J.B., Hoessel, J.G., Schneider, D.P., & Christensen, J.A. 1995, *A. J.*, **111**, 615.

Postman, M., Huchra, J.P., & Geller, M. 1992, *Ap. J.*, **384**, 404.

Postman, M., & Lauer, T. 1995, *Ap. J.*, **440**, 28.

Regös, E., & Geller, M.J. 1989, *A. J.*, **98**, 755.

Romer, A.K., 1994, PhD Thesis, Univ. of Edinburgh.

Romer, A.K., Collins, C.A., Cruddace, R.G., MacGillivray, H., Ebeling, H., Bohringer H., 1994, *Nature*, **372**, 75.

Rosati, P., Della Ceca, R., Burg, R., Norman, C., & Giacconi, R. 1995, *Ap. J. (letters)*, **445**, L11.

Sandage, A., 1961, *Ap. J.*, **133**, 355.

Schechter, P.L. 1976, *Ap. J.*, **203**, 297.

Schneider, D.P., Schmidt, M., & Gunn, J.E. 1994, *A. J.*, **107**, 1245.

Slezak, E., Bijaoui, A., & Mars, G., 1990, Å, 227, 301.

Struble, M.F., & Rood, H.J. 1991, *Ap. J. Suppl.*, **77**, 363.

Tully, R. B. 1987, *Ap. J.*, **323**, 1.

Tully, R. B., & Fisher, J. R., 1977, *Astr. Ap.*, **54**, 661.

Tyson, J.A., Wenk, R.A., & Valdes, F. 1990, *Ap. J. (letters)*, **349**, L1.

Weinberg, D.H., & Cole, S. 1993, *M. N. R. A. S.*, **259**, 652.

White, S.D.M., Navarro, J.F., Evrard, A.E., & Frenk, C. S. 1993a, *Nature*, **366**, 429.

White, S.D.M., Efstathiou, G., & Frenk, C.S. 1993b, *M. N. R. A. S.*, **262**, 1023.

Whitmore, B.C., Gilmore, D.M., & Jones, C. 1993, *Ap. J.*, **407**, 489.

Zaritsky, D., Nelson, A.E., Dalcanton, J.J., & Gonzalez, A.H. 1996, preprint (astro-ph/9612021)

Zwicky, F. *et al.* 1961–1968, *Catalog of Galaxies and Clusters of Galaxies* (Pasadena: California Institute of Technology).

3.8. Quasars

3.8.1 *Introduction*

In each spectroscopic observation, we will devote between one-tenth and one-sixth of the

fibers to a survey of quasar candidates. These are defined as stellar objects with colors well separated from the stellar locus.

The cumulative quasar counts in π steradians as a function of B magnitude are given in Table 3.7.1. The data are based on the results of Crampton, Cowley, and Hartwick (1987; cf., La Franca and Christiani 1996). Quasars on our system have $g' \approx B - 0.3$, but the transformation is affected at the level of several tenths of a magnitude at different redshifts by the emission lines. We give approximate g' values in the table.

Table 3.8.1. Quasars in π Steradians

B	g'	$N(< B)$
16	15.7	100
17	16.7	1300
18	17.7	11300
19	18.7	52900
20	19.7	172000
21	20.7	414000

The estimated limiting magnitude for obtaining redshifts of quasars with the proposed survey telescope is $g' \approx 19.7$. One can go significantly fainter with quasars than galaxies for two reasons (Section): quasars, being point sources, have a larger fraction of their light captured by the fibers than do galaxies, and redshifts for quasars will be based on measurements of strong emission lines, whereas in most cases one must rely on absorption features to determine redshifts for galaxies.

The proposed instrumentation of the Sloan Digital Sky Survey will therefore allow detailed study of the order of 100,000 quasars brighter than $g' \approx 19.7$. For comparison, the largest samples of quasars from individual surveys compiled in the last decade (cf., Hartwick and Schade 1990; La Franca and Christiani 1996, for reviews) include the works of Boyle, Shanks, and Peterson (1988) (423 quasars with $z < 2.2$), Hewett, Foltz, and Chaffee (1995) (1055 quasars with $z < 3.4$), Warren *et al.* (1991a,b, 1994) (85 quasars with $2.2 < z < 4.5$), Schneider, Schmidt, and Gunn (1994) (140 quasars with $2.1 < z < 4.8$), Christiani *et al.* (1995) and La Franca and Christiani (1996) (327 quasars with $z < 2.2$). The large, well-defined data base produced by the SDSS can be used to address many pressing questions concerning the nature of quasars, their evolution, quasar clustering, galaxy formation at high redshifts, and the properties of the intergalactic medium.

3.8.2 *Selection of Quasar Candidates*

The SDSS will use the colors and morphology of objects to identify quasar candidates in the photometric data. There have been several successful "multicolor" surveys for quasars based on photographic material, for example the programs described by Koo and Kron (1988), Boyle, Shanks, and Peterson (1988), and Warren *et al.* (1991a,b, 1994). The last project, using six broad bands ranging from u to i, discovered over 100 quasars (85 in the complete sample), including the first object with a redshift larger than four (Warren *et al.* 1987). The quasar with the highest known redshift ($z = 4.90$) was found in a multicolor survey using drift-scanning CCDs (Schneider, Schmidt, and Gunn 1991b).

The proposed quasar selection scheme for the SDSS is quite similar to previous studies; objects of stellar appearance with colors that lie well outside the stellar locus will be flagged

for spectroscopic investigation. The SDSS will have several improvements (besides areal coverage) over previous multicolor quasar surveys. The survey data base is produced by CCDs, with their associated advantages (linear response, photometric accuracy, better sky subtraction) over photographic material. The colors for each object are measured virtually simultaneously, so source variability will not be a confusing factor. The wavelength coverage of the SDSS filter set (3540 Å $< \lambda <$ 9240 Å) has been matched by one previous CCD survey (Hall *et al.* 1996a,b), but this was over a very limited area of sky. The reddest bands in particular will permit identification of quasars with redshifts as high as six and possibly seven!

Multicolor surveys, while a very successful method for finding quasars, do present some technical challenges. The fundamental basis for identifying quasars in such programs is the difference between the spectral energy distributions (SEDs) of quasars and stars. To a large degree the SEDs of individual quasars show the same features: a smooth, roughly power-law continuum and a number of broad emission lines of certain ions. There are, however, significant variations in the values of the slopes of the power law, and in the strengths and widths of the emission lines. Quasar continua are strongly affected by absorption from the Lyman α forest, as well as Lyman limit systems. In addition, the spectra of a small fraction (5% – 10%) of quasars possess strong, broad absorption features (the Broad Absorption Line, or BAL quasars; see section 3.7.4*h*). A further complication is that the survey colors of a given quasar SED depend on the redshift of the object. The Warren *et al.* (1994) paper presents an excellent discussion of the technical challenges of quasar detection in multicolor surveys.

We illustrate the nature of the SDSS quasar identification problem with Figure 3.8.1, which shows the locations of the expected quasar tracks in the color-color diagrams as a function of redshift for the SDSS filter system. The quasar tracks used are the median tracks of a Monte-Carlo simulation of quasar colors in the SDSS filter system. The emission spectra of quasars are simulated according to observed distributions of the power law continuum ($\alpha = 0.5 \pm 0.3$, Francis 1996) and a total of 36 emission lines and FeII features. The absorption features include Lyman α forest and Lyman Limit System (LLS) absorption simulated according to the procedures of Moller and Jakobsen (1990). The upper two and lower left panels display the results for the three sets of adjacent color pairs. The filled circles represent the stars contained in the spectrophotometric atlas of Gunn and Stryker (1983); the open circles are the white dwarfs studied by Greenstein and Liebert (1990). The relative density of star points in these figures does not reflect the expected relative density of objects in the survey, nor the effect of photometric errors. In particular, the number density of hot main sequence stars is strongly over-represented in these plots, especially at high galactic latitudes; we have simply plotted all of the objects in the two catalogs to show the positions of known stellar types.

Magnitudes in the SDSS system of the quasar tracks are marked with plusses at intervals of 0.1 in redshift. The graph in the lower right is the mean i' magnitude of a quasar with the absolute magnitude of 3C 273 ($V = -27$). Such a quasar is brighter than 20th magnitude in the i' band out to a redshift of about five.

Figure 3.8.1 shows the general characteristics of the quasar/star separation problem for SDSS, but one should keep in mind that: 1. the quasar positions in color space, especially for those of high redshift, have a big dispersion, which is mainly due to the distribution of absorbers in the line of sight; 2. the statistics of the Lyman α and Lyman Limit absorptions used in the simulation are based on observations of quasars with redshifts smaller than 4.5 (Figure 3.7.7), while the colors of quasars of redshifts larger than this are based on the extrapolation of the absorption to higher redshift; 3. emission line compact galaxies, which are one of the principal contaminants in multicolor QSO surveys (Hall *et al.* 1996a) are not

Figure 3.8.1 The colors in the SDSS filter system of quasar models. The plot shows the median track from a Monte Carlo simulation of quasar spectra as a function of redshift. The filled circles are the colors of stars from the atlas of Gunn and Stryker (1983), the open circles are white dwarfs from Greenstein and Liebert (1990). The panel in the lower right is the i' magnitude as a function of redshift of a quasar with the luminosity of 3C273.

included in these plots. Some of these have very strong H-α and [OIII] emission, and have similar colors to quasars.

A detailed study of the distribution of point-like sources in the color space, including a detailed Galactic model to simulate the distributions of stars and white dwarfs, and a quasar distribution based on the known quasar luminosity function and estimates of the color distribution of compact emission line galaxies is currently underway. Figure 3.7.2 shows work in progress along these lines by Xiaohui Fan, a graduate student at Princeton University. This is a model color-color-color diagram ($u' - g'$ vertically, $g' - r'$ horizontally, $r' - i'$ into the page) of stellar objects brighter than $r' = 20$ magnitude that would be seen at the North Galactic pole. The distribution of objects in three dimensions is shown in the center, and the projection onto each of the principal two-dimensional planes is shown on the edges. The red dots are stars, using the stellar population models and colors of Ng (1994). Notice that unlike the color-color diagrams of the previous figure, the hottest stars are simply not present: there are very few stars earlier than F towards the poles! The blue points represent quasars; the vast majority of them are beyond the blue end of the stellar locus, representing the fact that blue quasars with $z < 2.5$ are much more common than higher-redshift, and therefore redder, quasars. The green points are white dwarfs, with colors from the simulations of Bergeron (private communication), and with mass distribution and luminosity function from Liebert *et al.* (1988). Photometric errors have been included in this diagram. This work is far from complete: among other things, we do not model the presence of compact emission-line galaxies, binary stars, the reddest stars, and other important contaminants. However, these diagrams will be invaluable as we develop our quasar selection criteria.

The quasar tracks in Figure 3.8.1 reveal the overall features of their color distributions : 1. little change in the colors at low redshifts as a succession of relatively weak emission lines marches through the filters; 2) a move to the left (blue) in the diagrams when the Lyman α emission line enters the bluest of the three filters represented in each diagram; 3) a large shift to the right (red) as the Lyman α emission enters the middle filter and the absorption is seen by the blue filter; 4) eventually, when the optically-thick LLS absorption begins to show up in the bluest filter, the majority of the radiation will be absorbed by the LLS, and quasars become undetectable in those filters. As a result, their colors will be very different from those of any stars or low-redshift compact galaxies.

At redshifts smaller than 2.1, the u' filter measurements are crucial for separating the colors of quasars and stars. The ultraviolet measurement places quasars well away from main-sequence stars (these stars have a strong Balmer jump not present in quasars; the main problem is confusion of the quasars with white dwarfs; cf., Green, Schmidt and Liebert 1986). Employing 5 colors will help in separating them, since the Rayleigh-Jeans tail of a black body ($f_\nu \sim \nu^2$) is very different from the quasar power law ($f_\nu \sim \nu^{-0.5}$). At redshifts larger than 2.5, the u' filter is no longer a good quasar/star discriminant; once Lyman α enters the g' filter, the quasar enters the realm of the main-sequence stars, and have very similar colors to the F-type stars (this is analogous to the z = 2.2 limit for UVX surveys).

The main place where quasars are difficult to separate from stars lies in the range z = 2.5 – 3.0. This is a common feature for all broad-band multicolor surveys, since in this redshift range, the three spectral features of quasars, i.e. absorption due to the Lyman α forest, Lyman α emission and the power law continuum redward of Lyman α, are redshifted to approximately the positions of three features prominent in F-type stars: the Balmer jump, the maximum in the stellar continuum and the gradual decline after that. The large width of the broad-band filters smooths these features and gives quasars and F type stars very similar colors. The selection effects in this range are strong and quite dependent on the number density of F-G stars in the observed direction. An accurate estimate of the quasar density

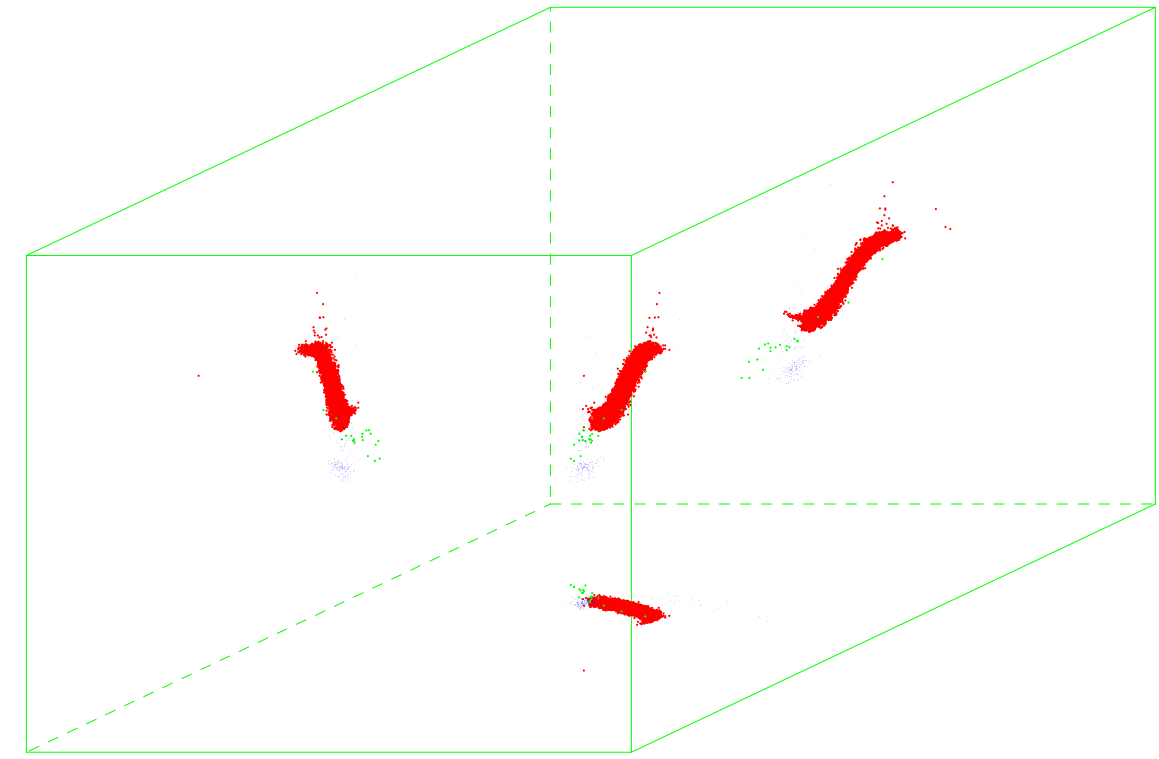

Figure 3.8.2 A color-color-color diagram ($u' - g'$ vertically, $g' - r'$ horizontally, $r' - i'$ into the page) of the expected distribution of stellar objects in the North Galactic Pole direction. The three-dimensional distribution is shown in the center, with projections on the three planes on the sides. Red points are stars, green points are white dwarfs, and blue points are quasars. Photometric errors are included in this figure.

in this redshift range needs detailed modelling of the distribution of quasars and stars in color space. The only way to tie this down observationally is a *complete* survey of all stellar objects over some limited area of sky. As discussion in Section 5.5, we plan to do exactly this in the Southern stripe.

When the redshift is greater than 3.5, the existence of LLS in quasar spectra makes them rather easy to distinguish from stars. The APM survey (Storrie-Lombardi *et al.* 1996) and the DPOSS survey (Kennefick *et al.* 1996) use broad-band colors to successfully find high redshift quasars. The SDSS color system also has no problem in separating quasars from stars in this regime, as we saw above.

An ideal quasar selection algorithm would be (a) uniform over the survey area, such that the quasar population in different regions of the sky can be compared directly (b) complete, in that quasars of all types are included, and (c) efficient, in that the number of non-quasars

Figure 3.8.3 Fit of the stellar locus (open squares) in color space. The data are from Trevese *et al.* (1994). Known stars are dots and quasars are filled squares. Sources that were chosen as quasar candidates are circled.

targeted is minimized. These three requirements pull us in different directions, given the changing stellar densities and populations as a function of direction, and the presence of quasars with broad-band colors essentially indistinguishable from stars. We thus cannot simultaneously optimize all three. Our approach is to enforce the uniformity requirement over only half of the survey area, and not require 100% completeness (which would be impossible anyway) over the entire survey area. As mentioned above, the completeness of the quasar sample will be quantified with the parallel survey of *every* stellar object over an appreciable region in the Southern stripe (Section 5.5).

Based on these simulations and data, and with these requirements in mind, the SDSS quasar working group has developed an algorithm to select quasars from the point-source objects found in the photometric observations. The scatter of stars around the mean stellar locus in color-color-color-color space is modelled as a Gaussian distribution with width varying as a function of position along the locus; quasars are defined as outliers from this distribution. In order that the quasar selection be uniform, we require that the stellar locus and its scatter be defined *a priori*; this is a major task of the test year. The stellar population, and the stellar density itself, of course varies strongly as a function of direction. For this reason, the quasar survey will be done in two parts. A uniform and deep survey of quasars will be done in the "cap" region, the half of the survey area at lowest stellar density according to the star-count models of Bahcall and Soneira (1981). This would account for roughly 75% of the quasar candidates surveyed over the sky. In the "skirt", the increased stellar density would mean that the number of false positives would become prohibitively high (especially considering that the changing stellar population will cause the stellar locus to move), and thus we back off and select only the brightest candidates with the highest statistical confidence of being quasars. In both cap and skirt, stellar objects associated with radio sources from the FIRST survey (Becker *et al.* 1995) will be targeted independent of their colors; estimates are that this will represent between 5 and 10 extra fibers per plate. The details of the quasar selection are still very much under development, and indeed refining the quasar selection criteria will be the most important and time-consuming job of the test year (Section 5.3). Our aim is for roughly 70% efficiency, and thus we plan to target of the order of 140,000 quasar candidates to get 100,000 bona fide quasars, again the bulk of them in the low stellar density $\pi/2$ ster of the survey footprint. The quasar counts in Table 3.8.1 imply that a g' limit of roughly 19.7 is appropriate, but this will need to be carefully calibrated during the test year, and of course this will need to be augmented with additional selection in the redder bands for those quasars whose redshifts put Lyman α beyond the g' band.

As an illustration, the prototype algorithm has been tested using a set of UJFN magnitudes and object identifications from the SA 57 survey of Trevese *et al.* (1994). Selection of objects which lie more than 3σ from the stellar locus (note that the "thickness" of the stellar locus is a function of color) yielded 79% completeness with 47% contamination, as shown in Figure 3.8.3. Much of the contamination is due to white dwarfs (lower-left-hand corner of the plot), and as mentioned above, we can reduce this contamination considerably by use of all the color information at our disposal. The contamination should be further reduced with the smaller photometric errors the SDSS data will have, and reduced further still by not including objects in regions of color-color space in which quasars are very rarely found (although this will unfortunately reject some fraction of the rare quasars with very unusual SEDs; cf., Section 3.7.4i).

The use of the FIRST data will help investigate the frequency of quasars highly reddened by internal dust (cf. Sanders *et al.* 1989, Webster *et al.* 1995, Shaver *et al.* 1996; Malhotra, Rhoads, and Turner 1996). This should prove very interesting when the broad-band spectra of quasars are constructed from SDSS and 2MASS data. The FIRST quasars will also prove invaluable for understanding our selection effects; without it, we would not know of the existence of those quasars which lie directly in the stellar locus (cf., Figure 3.8.3). Faint

stellar objects associated with ROSAT sources (Chapter and Section 3.7.4g) are also very likely to be quasars, and will further increase the completeness of the quasar sample. Again, the completeness of the sample will be quantified with the Southern stripe survey of all stellar objects over an appreciable area of sky.

3.8.3 *Spectroscopy of Quasars*

During the course of the Sloan Digital Sky Survey, spectra of approximately 100,000 quasars will be obtained with the SDSS spectrograph. The spectra, covering wavelengths from 3900 Å to 9200 Å at a resolution of 2000, will be of very high quality by current standards; at present only a few hundred quasar spectra exist with such resolution, signal-to-noise ratio, and wavelength coverage.

Figure 3.7.4 displays a simulated SDSS spectrum of a quasar with a redshift of 3.4 and $g' = 19.7$. These data should be typical of the lowest-quality quasar spectrum that will enter the SDSS data base. The upper two panels are the data from the blue side of the spectrograph; the lower two graphs are the red channel spectra. The data have been compressed for display purposes (each spectrograph channel in the figure has 1000 points, whereas the actual spectra will contain 2000 points per channel). It is clear that the spectra will contain a wealth of information about the properties of the quasar itself, as well as about the absorption lines due to material in the intergalactic medium.

3.8.4 *Quasar Science Programs*

a. Clustering

The history of the study of the large scale spatial distribution of quasars has been almost entirely written in the past decade. The first large systematic investigation was performed by Osmer (1981); he found no significant clustering at $z \approx 2$ from a sample of 174 quasars found in a CTIO objective prism survey. Many subsequent programs, using relatively small sample sizes, also failed to detect any signs of quasar clustering (see Shaver 1988 for a review).

This picture of randomly distributed quasars has changed in the past few years because of substantial increases in sample sizes, in particular the availability of large low-redshift samples (cf., Crampton 1992). The analysis by Shanks, Boyle, and Peterson (1988) of their sample of 423 quasars with $z \leq 2.2$ produced strong evidence for the existence of quasar clustering on small (less than 10 h^{-1} Mpc) scales. Iovino and Shaver (1988) found a similar result for $z < 1.5$ quasars, but they could detect no significant clustering for quasars at redshifts above 1.5. A recent study by Andreani and Christiani (1992) combined the results from a number of surveys to produce a sample of ≈ 1600 quasars; unlike Iovino & Shaver, their data was consistent with no change in the two-point correlation function as a function of redshifts for $z \leq 2.15$.

Crampton, Cowley, and Hartwick (1989) discovered what is to date the most convincing grouping of quasars; a total of 23 quasars with $1.0 < z < 1.2$ located within a 2° field was identified from their CFHT survey. Another interesting set of probable physical pairs was discovered in a CCD/grism survey (Schneider, Schmidt, and Gunn 1994); among the 90 quasars with redshifts larger than 2.7 found in this program, there are three pairs of quasars that have an angular separation of less than $7'$ and have $\Delta z < 0.05$ (cf., Kundić 1996). The strength of the quasar correlation function suggests that quasars are located in small groups of galaxies (Bahcall and Chokshi 1991).

At the present time the "conventional wisdom" (if this term can apply to such a rapidly changing subject) is that quasars are clustered on scales less than 10 h^{-1} Mpc for $z < 2$;

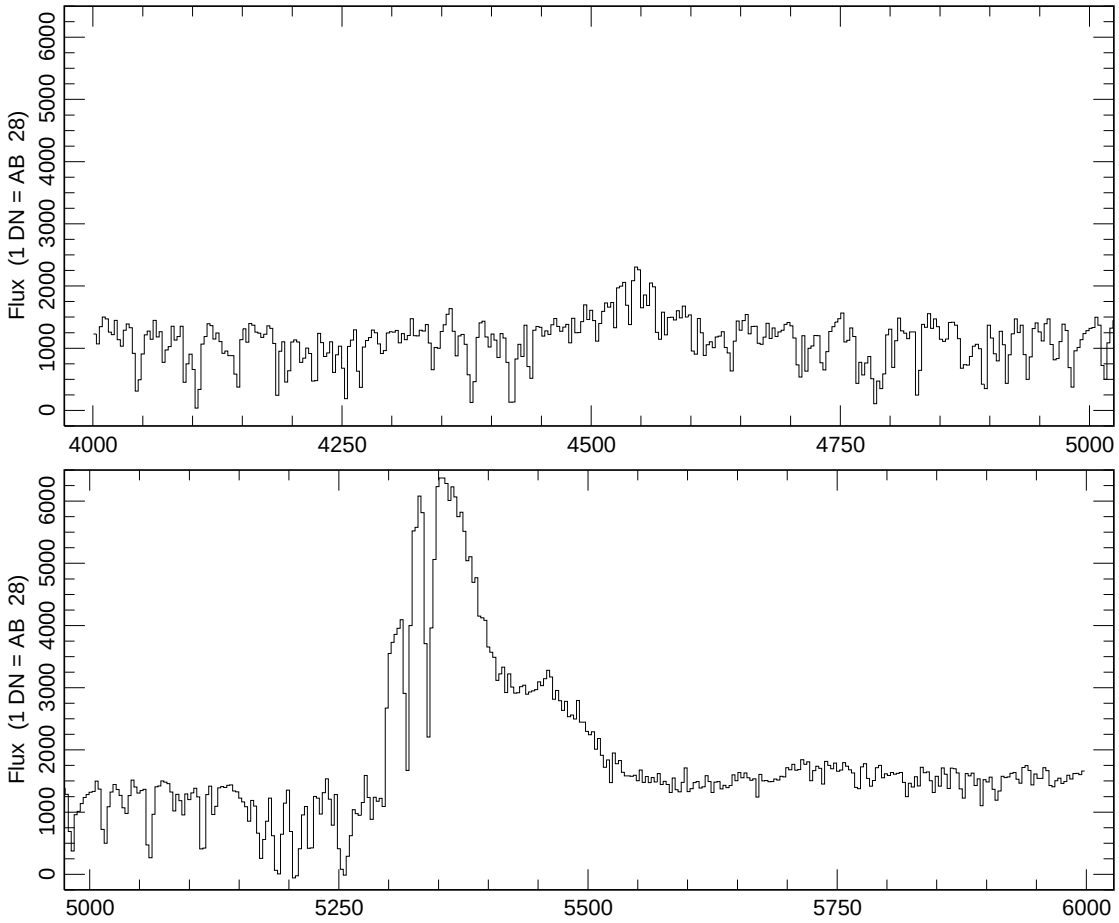

Figure 3.8.4a Simulated SDSS spectrum. The figure shows the blue channel of a $z = 3.4$, $g' = 19.7$ quasar. The spectrum has been rebinned so that there are only 1000 points per spectrum (the actual data will have 2000 points per spectrum).

at distances or redshifts larger than this the clustering disappears. It is clear that this subject would benefit enormously from the existence of a very large sample having well-defined selection properties. The SDSS will in this sense be ideal. The errors on the derived correlation function vary inversely with the sample size at small scales; therefore, we can expect an improvement in the sensitivity to clustering over existing samples by a factor of nearly 200! In the southern survey, we will probably perform a redshift survey of quasar candidates at least a magnitude fainter than that in the North (Section 5.5). This will not increase the volume covered in the survey but will explore less luminous objects with much larger space densities, and so will provide much better definition of the structure on intermediate scales.

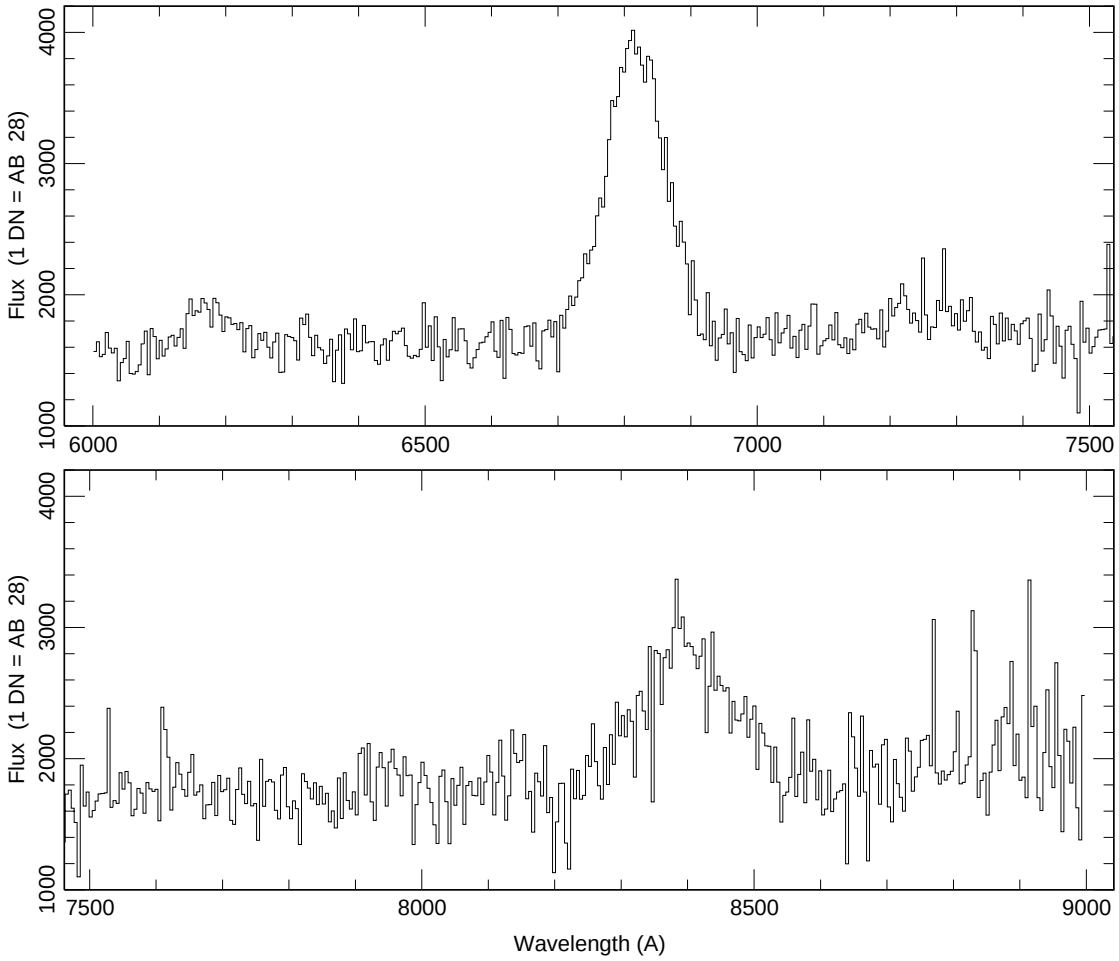

Figure 3.8.4b Simulated SDSS spectrum. The figure shows the red channel of a $z = 3.4$, $g' = 19.7$ quasar. The spectrum has been rebinned so that there are only 1000 points per spectrum (the actual data will have 2000 points per spectrum). The effects of the dichroic filter, and the overlap region of the spectroscopic channels, have been suppressed in these figures.

On the theoretical front, Matsubara and Suto (1996) and Ballinger *et al.* (1996) have pointed out that a comparison of the radial and transverse dependence of the quasar correlation function in principle can yield a constraint on the cosmological constant. This measurement will not be easy; the effect is not large, and is mimicked to some extent by peculiar velocity effects. Nevertheless, the quasar sample will be big enough to attempt this measurement.

b. Evolution

Quasars are in many respects our best probes of cosmic evolution because their enormous intrinsic luminosities (up to more than 10^{13} $L_\odot$) allow them to be detected out to very large

distances, and hence long look-back times. The first ten years of quasar research saw the largest known quasar redshift rise from 0.158 (Schmidt 1963) to 3.53 (Wampler, Robinson, and Baldwin 1973). This increase meant that one could study the most luminous objects and their environments at an era when the Universe was only about 10% of its current age; a startling gain in our grasp of the cosmos. The very existence of quasars at this epoch revealed valuable insights about the time scales for the formation of structure, and the discovery of the multitude of absorption features in the spectra of quasars, in particular the "Lyman α forest" (Lynds 1971), permitted investigation of non-luminous matter distributed throughout the Universe.

The search for high-redshift quasars is motivated by the fact that their existence places interesting constraints on quasar evolution and the early history of the Universe. If quasars are 10^7–10^9 $M_\odot$ black holes accreting matter at the Eddington limit, then the e-folding time scale for the growth of the mass of the central object is $\approx 4.4 \times 10^8$ years (Rees 1984 and references therein; Yi 1996). This is comparable to the age of the Universe at redshifts of 4–5 in currently fashionable cosmologies.

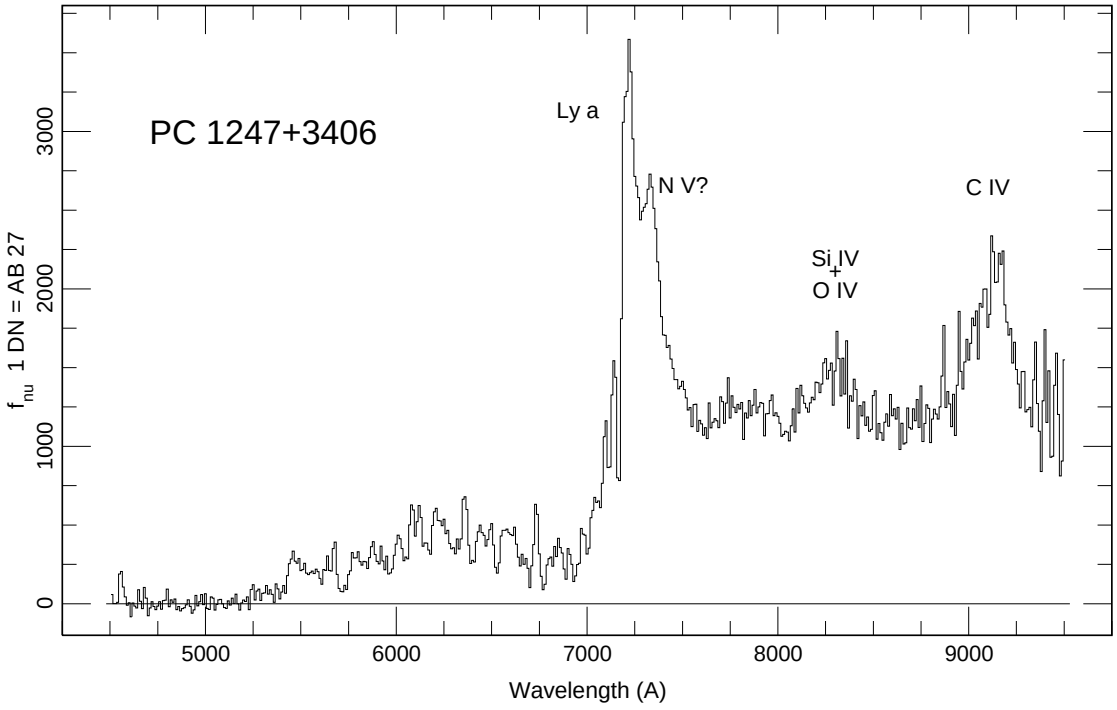

Figure 3.8.5 The spectrum of the most distant known quasar ($z = 4.90$). This object was discovered using a four-filter/CCD survey. Note the strong absorption blueward of the Lyman α emission line (Schneider *et al.* 1991b).

Despite extensive efforts to locate ever more distant objects, the subsequent 13 years saw

the largest known redshift increase only to 3.80 (Hazard, McMahon, and Sargent 1986). The last decade, however, has seen explosive growth in this field; these breakthroughs were made possible by improvements in detector technology and the rapid increase in the capabilities of computer hardware and software (Warren and Hewett 1990). The first redshift greater than 4 was found by Warren *et al.* (1987) in a multicolor survey that employed a filter scheme very similar to that being used for the Sloan Digital Sky Survey. The redshift record was broken at the rate of roughly once per year between 1987 and 1991; the most distant known object as of the time of this writing, also found in a multicolor survey, has a redshift of 4.90 (Schneider *et al.* 1991b).

Several surveys, using both multicolor and slitless spectra techniques, have been undertaken since 1991, but these programs have not discovered any quasars with redshifts larger than 4.9. It is becoming clear that quasars with redshifts of five or larger are exceedingly rare; in order to assemble a reasonable sized sample of such objects one must cover large areas of sky.

This requirement is of course met by the SDSS; this survey should dramatically increase our understanding of the quasar luminosity function at high redshifts. Based on current results, we expect the SDSS to contain several thousand quasars with $z > 3$ and as many as 1000 with $z > 4$. The SDSS data should permit identification of quasars at redshifts well above five, where Lyman α moves into the i' and z' bands. At present, it is a struggle to assemble a uniformly selected sample of 100 quasars with $z > 3$. As of this writing (December 1996), published studies have found about 50 quasars with $z > 4$ using a large range of search techniques.

With the new data, we can expect a resolution of the longstanding debate on the nature of the evolution of the quasar luminosity function and the question of whether the space density truly declines at redshifts larger than 3.5 (Osmer 1982). Figure 3.7.6 shows an estimate of the evolution of the spatial density of quasars with $M_B < -26$ as a function of redshift produced by combining the results of the Hewett, Foltz, and Chaffee (1993) study (solid line for $z < 2.5$) with a recent high-redshift emission line survey (four points for $z > 2.5$; Schmidt, Schneider, and Gunn 1995). This picture of the evolution is qualitatively consistent with the results of Hewett *et al.* (1993), Warren *et al.* (1994), and Kennefick *et al.* (1995), although there are indications that the space density of the most luminous objects does not decline nearly as steeply with redshift as do the quasars used in the analysis that produced Figure 3.7.6 (Irwin, McMahon, and Hazard 1991). The luminosity function at redshifts larger than 4.5 remains quite uncertain, as only six such quasars are known.

As discussed by Efstathiou and Rees (1988), the number of $z \approx 5$ quasars will place strong constraints on the cold dark matter model, where massive condensed structures would be expected to form at a later epoch. Turner (1991) describes the theoretical difficulties posed by the existence of the known $z > 4$ objects; Cavaliere and Padovani (1989) and Haehnelt and Rees (1993) discuss the constraints that the quasar luminosity function places on the intrinsic properties of quasars. The spectra of high-redshift quasars provide information about the production of metals in the early Universe (*e.g.*, Hamann and Ferland 1992). The existence of even a few quasars with redshifts greater than 4.5 has stimulated theoretical investigations of the processes that could produce such rapid formation of massive objects (Loeb 1993; Loeb and Rasio 1994).

Since the discovery of individual quasars at redshifts larger than five would have important scientific implications, it is planned that a small number of fibers (3–5 per spectroscopic field) will be devoted to examination of faint objects (down to $r' \approx 22$) that are promising candidates for $z > 5$ quasars. Although the quality of these spectra will be rather low,

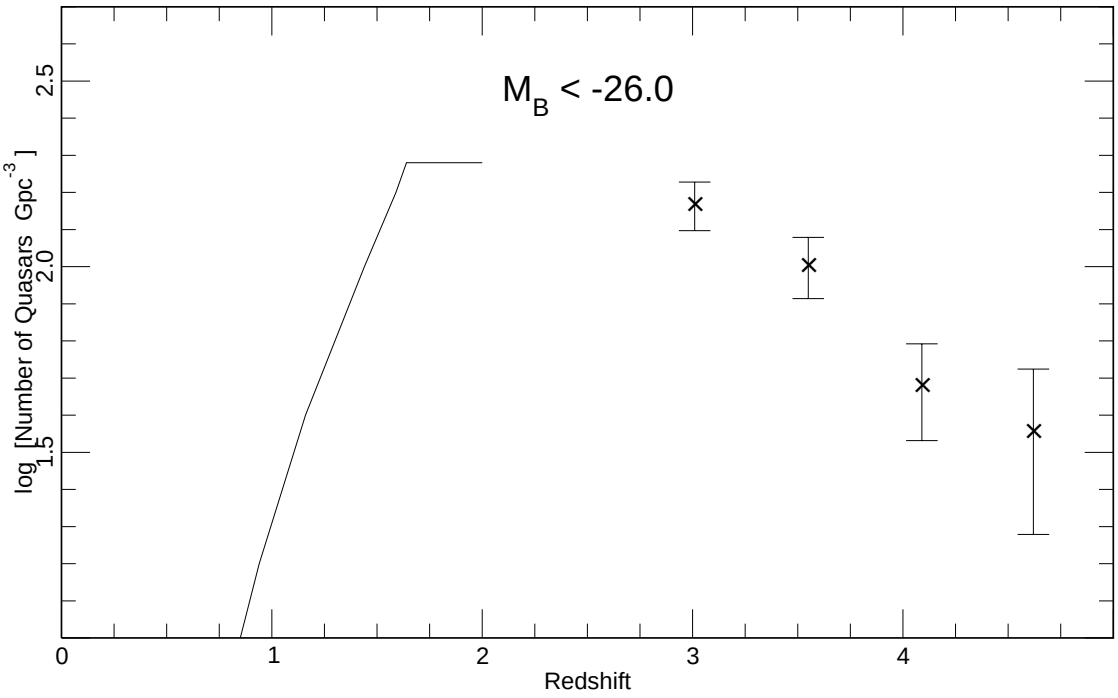

Figure 3.8.6 Quasar evolution. The solid line is the current best estimate of the evolution of the spatial density of luminous quasars with $M_B < -26$.

the signature of a high-redshift quasar (strong Lyman α line combined with significant intergalactic absorption) is sufficiently unusual that a redshift determination will be possible.

The quasars found in the SDSS will also improve our understanding of the quasar luminosity function at lower redshifts. The best low-redshift ($z < 2$) samples typically have a few hundred quasars, while the high-redshift studies typically contain but 100 objects. In practice it is difficult to tie these results together, because the techniques used to find the quasars often change as a function of redshift. The SDSS will produce a sample roughly two orders of magnitude greater in size than current programs, and will do so with a homogeneous selection system. The SDSS will certainly discover a large number of bright, low-redshift quasars; the CCD/grism survey of Schneider, Schmidt, and Gunn (1994) covered only 61 sq degrees and was designed to find $z > 2.7$ quasars, but still managed to discover three previously unknown quasars with $z < 2$ brighter than $r = 16.7$.

The number of bright, high-redshift quasars will be greatly enlarged as a result of this survey. Over 11,000 quasars with $B < 18$ should be discovered based on the numbers in Table 3.8.1; of these, approximately 10% will have $z > 2.5$. Thus, there should be a sample of more than a thousand bright, high redshift objects available for study with 8-m class telescopes, which will have enough light gathering power to permit the use of spectral resolutions as high as 60,000. At these resolutions, the intergalactic medium at high redshift can be studied in unprecedented detail via the absorption lines. (A good example of state-of-the art measurements is the R=36,000 spectrum at a signal-to-noise ratio of 160 of QSO1422+231

by Songaila and Cowie 1996). We will be able to obtain valuable new information on the nature and evolutionary history of the intergalactic medium at large redshifts and on the properties of galaxies that intersect the line of sight but are otherwise too distant and too dim to be seen.

c. Quasars in Clusters of Galaxies

The survey will be ideal for investigating the associations between low redshift quasars and galaxies. Ellingson, Yee, and Green (1991), using a sample of 63 nearby ($0.3 < z < 0.6$) quasars, show that while radio-quiet quasars are rarely found in rich clusters, radio-loud quasars frequently reside in Abell richness class 1 or richer clusters. The distinction between the clustering properties of radio-quiet and radio-loud quasars has recently been questioned (Fisher *et al.* 1996) in work that shows that the quasar-galaxy correlations are very strong. Since the SDSS will provide deep imaging of all of the quasar fields, it will be possible to study the relationship between quasars and their local environments for a very large number of quasars, especially in the Southern stripe. This line of investigation will undoubtedly yield important information about the relation between galaxy environments, interactions, and the phenomenon of active nuclei for many more than the ~ 50 cases studied so far.

d. Absorption Systems

Perhaps the greatest scientific contribution to date arising from quasar research comes not from what their properties reveal about galaxy formation or physical processes under extreme conditions, but their use as "cosmic flashlights" to expose the presence of non-luminous matter. When quasars were discovered whose redshifts were sufficiently large to bring the Lyman α transition into the optical band ($z > 2$; Schmidt 1965), it was quickly realized that studies of the absorption lines in the spectra of these distant beacons might yield valuable insights into the properties of "dark matter" (Gunn and Peterson 1965; Bahcall and Salpeter 1965). The subsequent 30 years have more than fulfilled the early expectations; studies of quasar absorption lines systems have provided vast amounts of information about the physical conditions of the intergalactic medium and the distribution of galaxies. Until recently, the investigation of quasar absorption lines was, to a large extent, confined to intermediate or high redshifts (the strong resonance lines of the cosmically abundant light elements all occur in the ultraviolet), but the past few years observations with the ultraviolet spectrographs of HST have permitted investigation of the local intergalactic medium (Bahcall *et al.* 1993, 1996). Recent reviews of quasar absorption lines are presented by York (1994) and Meylan (1995).

The impact of general intergalactic absorption on the spectra of quasars is vividly demonstrated in Figure 3.7.7. This graph displays the depression (fraction of light absorbed) in quasar spectra at wavelengths blueward of the Lyman α emission line. We ignore the complete absorption which occurs when a strong Lyman limit system is detected. The amount of absorption rises from zero for low redshift quasars (observed with the International Ultraviolet Explorer) to over 50% at redshifts larger than four. Figure 3.8.5, the spectrum of a $z = 4.90$ quasar, shows how pronounced the absorption becomes at high redshift.

Quasar absorption lines are seen from three types of systems: 1) systems with only the Lyman series lines of H I detected (often only Lyman α is seen), with a wide and not strongly clustered distribution in redshift space; 2) systems with lines of high and/or low ionization states of heavy elements; and 3) broad absorption line (BAL) systems.

Systems of type 1 make up the "Lyman α forest" (Lynds 1971). Heavy element lines are not generally seen for H I column densities below 10^{15} cm^{-2}, either because the systems are primordial (Sargent *et al.* 1980) or because the abundances are too low to yield detectable

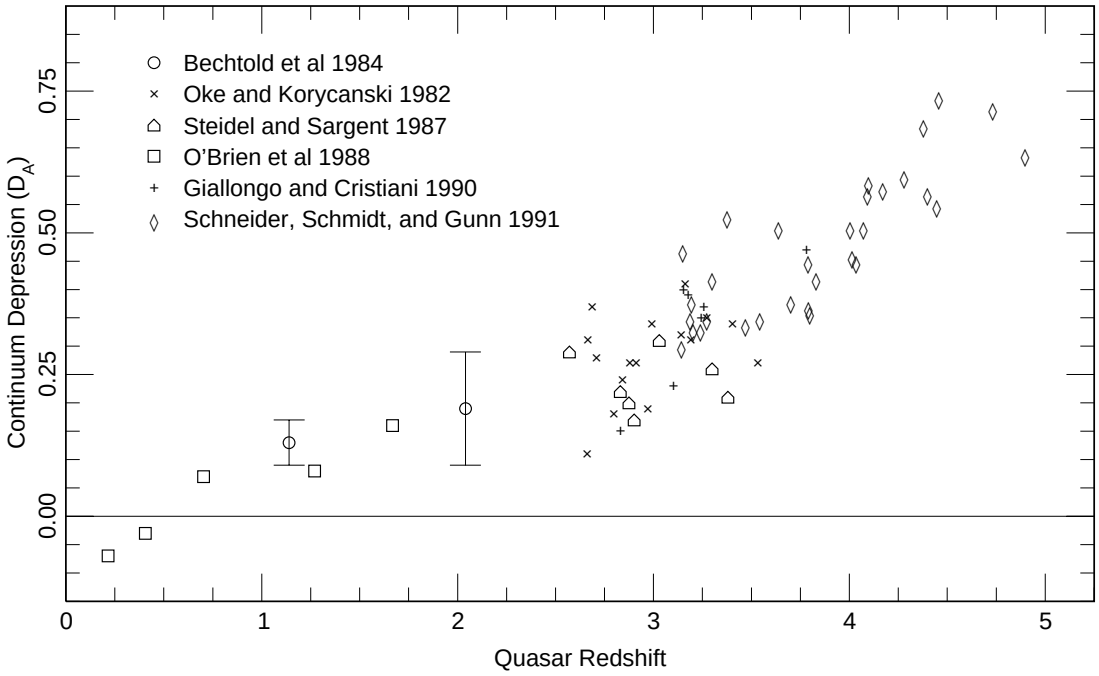

Figure 3.8.7 The amount of absorption due to intergalactic gas (presumably mostly the "Lyman α clouds") as a function of redshift. At redshifts larger than four, more than half of the light is absorbed (Schneider *et al.* 1991a).

absorption features with the instrumentation used. Recent evidence is conflicting on the relationship of these systems to galaxies. They may be directly associated with galaxies (Lanzetta *et al.* 1995), loosely associated with galaxies (York *et al.* 1986, Bowen *et al.* 1996), or not associated with galaxies (Shull *et al.* 1996, see also arguments in Bowen *et al.* 1996).

Systems of type 2 generally form a continuum with type 1 systems, in terms of strength of heavy element lines; the latter are weak but detectable as the column densities increase from 10^{15} to 10^{17} cm^{-2} (Meyer and York 1987, Tytler and Fan 1994, Songaila and Cowie 1996). At higher column densities, the Lyman limit of H I is detectable at rest frame 912 Å (Tytler 1981; see the region 5180 to 5280 Å in Figure 3.8.5). Above 10^{20} cm^{-2}, the Lyman α and a few other lines of the Lyman series develop damping wings that are discernible even in the heavily blended forest of weaker lines. Abundances of the heavy elements, such as O, Si, Fe, Zn and Cr are uncertain for many observational reasons, but all indications are that at high z, the abundances are very low (0.1 to 1% of Solar), rising to just about 10% of Solar at $z < 1$ (Timmes *et al.* 1995). When systems of type 2 are within a few thousand kilometers per second of the redshift of the background QSO, they are referred to as "associated "and may provide a sample of gas processed by and ejected from the QSO.

Systems of type 3 contain deep absorption troughs on the short wavelength side of the main QSO emission lines, with velocity widths of up to 10,000 km/sec. There is some weak evidence that they may be related to the narrower associated systems (Jannuzi *et al.* 1996).

As many as 10% of QSOs show the BAL phenomenon (Weymann *et al.* 1991); see Section h below.

The Lyman α forest (type 1 lines) makes up the general depression that increases in redshift as noted above (Figure 3.8.7). The number of these lines increases with redshift as $(1+z)^{2.3}$ past z=2 (Murdoch *et al.* 1986, Lu *et al.* 1991), whereas the number of heavy element line systems per unit z decreases above z=2 (Steidel 1990). The latter effect may be largely related to the decrease of the abundance of heavy elements with increasing redshift. The Lyman α forest has more lines at low z than expected from the above law (Bahcall *et al.* 1993), whereas too little is yet known about the heavy element line systems at $z < 1$ to make a general statement.

The origin of the type 2 systems is still controversial. Some may be associated with galaxian disks (Wolfe *et al.* 1986), some may be associated with star forming building blocks of modern galaxies (Yanny and York 1992, Pascarelle *et al.* 1996) and some may be loosely associated with the outer parts of modern galaxies (Churchill *et al.* 1996). Whatever the origin, the Lyman limit absorption from these systems accounts for the removal of large amounts of radiation shortward of the wavelength that corresponds to $(1+z) \times 912$ Å . Since the system redshifts are generally uncorrelated from one quasar to the next on large scales, the absorption from the limits introduces scatter in the QSO tracks of Figure 3.8.1. These same Lyman limit systems offer a direct probe of neutral hydrogen in the Universe at early times (Meiksin and Madau 1993) and provide the mean free path of ionizing radiation at these epochs.

The SDSS quasar spectra will be well suited, in wavelength coverage, spectral resolution, and signal-to-noise ratio, for many absorption line studies (although it certainly will not compete with the very high-resolution studies of absorption lines carried out, e.g., with the HIRES instrument on the Keck telescope; cf., Songaila and Cowie 1996 for an example of the spectra that are attainable). The Lyman α, C IV, and Mg II lines will be visible at redshifts 2.3–6.4, 1.6–4.8, and 0.5–2.2, respectively, and the Lyman-limit can be seen at redshifts greater than 3.4. A one hour exposure on a $g' = 19.7$ object with the SDSS spectrograph will produce a spectrum with a signal-to-noise ratio (per 2.5 Å resolution element) of 14 at 5500 Å. The spectra will have a 5σ limit on the observed equivalent widths of unresolved absorption lines of ≈ 0.9 Å, corresponding to rest frame equivalent widths a factor of $(1+z)$ smaller, and comparable to the limits of various surveys of absorption lines extant in the literature. At high redshifts, the SDSS spectra will not have sufficient resolving power to separate the individual lines in the Lyman α forest, but the properties of the Lyman α lines can still be investigated in some detail using statistical techniques developed in the past few years for quasar absorption line studies (Jenkins and Ostriker 1991; Webb *et al.* 1992; Press, Rybicki, and Schneider 1993). The largest published absorption line surveys (e.g., Steidel and Sargent 1992) contain at most ≈ 100 quasars each having data for less than an octave of wavelength coverage, in most cases much less. This survey will contain uniform data for 100,000 objects with more than an octave of coverage. If one adds up the redshift range that will be covered for all of our quasars, we find $\Delta z = 80{,}000$ for $0.5 < z < 1.5$, $\Delta z = 60{,}000$ for $1.5 < z < 2.5$, and $\Delta z \approx 4{,}000$ for $2.5 < z < 3.5$.

Current absorption line studies show that one can expect to see about one galaxian absorption system per unit redshift per quasar. The lines in which these can be detected include C IV, $\lambda_{\rm rest} = 1550$ Å; Si IV, 1393, 1402 Å; Si II, 1526 Å; Al II, 1670 Å; Fe II, five lines in the range 2300-2600 Å; Mg II, 2795, 2802 Å; and Mg I, 2852 Å. At the equivalent width limits of our faintest quasars, two-thirds of absorption-line systems in the literature have only C IV lines at $\lambda > 4000$ Å; this doublet is easily recognizable with the 2.5 Å SDSS spectral resolution. The SDSS spectra should result in the detection of 0.5 absorption systems per unit redshift that contain either multiple species or only C IV lines.

The number of galaxian absorbers per unit redshift increases from $z = 0$ to $z = 2$ to 2.5, then decreases. If we detect 0.5 absorbers per unit redshift, the number of absorbers found in the survey will be 40,000 ($0.5 < z < 1.5$), 30,000 ($1.5 < z < 2.5$), and approximately 2,500 for $2.5 < z < 3.5$. The detected objects have typical separations of tens of Mpc, similar to the separation of the survey quasars.

Key information on galaxies and large-scale structure at large redshifts can be obtained simultaneously with the galaxy and quasar surveys described above. Even though the luminosity of stars in distant galaxies may not be discernible, those galaxies will produce gas absorption lines in spectra of objects behind them. Aggregates containing but 10^{-6} of a galaxy mass can be detected in this way. Steidel *et al.* (1994) have shown that follow-up of such systems on larger telescopes is an ideal way to identify galaxy populations at high redshifts for evolution studies.

The sample of galaxian absorbers is important for a number of studies. First, the sample provides about 1000 objects in each of several volumes the size of the extremely well sampled nearby galaxy surveys, allowing clustering to be measured on very large scales (Quashnock *et al.* 1996; cf., the discussion in Section 3.1.4). The sample size is so large that the correlation function on scales of 50 to 100 Mpc can be determined in each volume to higher precision than we now know the correlation function within the local volume to 100 Mpc, and can be determined separately for fifty volumes of 1 Gpc^3. The evolution of the correlation function on large scales is of great interest in distinguishing low density and high density universes.

A second study follows from the fact that the large samples will include an appreciable subset of strong line systems ($\sim 10\%$) with detectable lines of many elements (those given above plus chromium, nickel, cobalt, copper and zinc). This sample will be adequate, when studied with larger telescopes, to distinguish directly the gradual build-up of grains and of the heavy element pool in galaxies from $z = 4$ to the present (Pettini *et al.* 1994, Lu *et al.* 1996, Lauroesch *et al.* 1996).

Since the sample is so large, one can break the data into subsamples of significant size, selected only by direction. The statistics of the subsamples can be analyzed in a variety of ways (number of absorbers per unit redshift; coarse abundances; cosmic radiation pumping of atomic levels; galactic radiation fields from ionization states of observed species; and changes of these quantities with time) to address the question of the homogeneity and isotropy of the universe on the largest scales yet achieved.

Finally, the survey will greatly enhance the ability to study coherent structure of the intergalactic gas by the inevitable discoveries of pairs (or sets) of quasars with separations of only a few arcminutes (*e.g.*, Jakobsen *et al.* 1986; Fang *et al.* 1996); this subject is discussed in detail in Section 3.1.4.

e. Intrinsic Properties

A large data base on the emission line and continuum properties of quasars will be produced by the survey. Figure 3.8.4 displays a simulated spectrum of a quasar with a redshift of 3.4 and an apparent brightness of $g' \approx 19.7$, our survey limit. Since the quality of all of the quasar spectra obtained in the survey will be at least this high, it is obvious that the set of 100,000 spectra will allow a number of detailed investigations of the intrinsic properties of quasars.

We know that quasar spectra differ significantly in their line widths and strengths, but no obvious pattern has emerged except for the "Baldwin Effect": lower luminosity quasars tend to have stronger emission lines (Baldwin 1977). However, emission line data and photometry

of quasars are very important for testing physical models of quasars, and the new data will be by far the best available for such work. The best continua/emission line studies to date contain of the order of 100 quasars; the analyses presented in Steidel and Sargent (1991) and Tytler and Fan (1992) are examples of the scientific questions that can be addressed using a set of high-quality quasar spectra.

f. Gravitational Lenses

Few topics in astronomy have captured the popular (and professional, for that matter) imagination as vividly as gravitational lenses. The fact that standard theories reproduce the basic observational parameters of several lenses (*e.g.* Q 0957+561, the original "double quasar"; cf., Grogin and Narayan (1996), 2016+112, and 2237+0305) indicates that the general theory of relativity applies on scales of the sizes of galaxies. It also confirms our basic understanding of cosmological distances and redshifts as well as galaxy mass distributions. Measurements of time delays between images (Kundić *et al.* 1996, and references therein) and possible microlensing events (Irwin *et al.* 1989) are among the most exciting observational results of the past few years. In addition, the observed frequency of lensing has been used to place constraints on the masses of galaxies and of the value of the cosmological constant (Turner, Ostriker, and Gott 1984; Fukugita and Turner 1991; Maoz and Rix 1993).

The known cases of lensing have frequently been serendipitous discoveries, though lens surveys at optical wavelengths (Surdej *et al.* 1987; Yee, Filippenko, and Tang 1993) and radio (Lawrence *et al.* 1984) wavelengths have experienced some success. In particular, the Hubble Space Telescope Survey for gravitational lenses (Bahcall *et al.* 1992; Maoz *et al.* 1993) examined Planetary Camera images of 498 quasars, and discovered the most distant known lens (Q 1208+1011 at a redshift of 3.80 (Maoz *et al.* 1992). Most of the known lenses have component separations of several arcseconds; in some of these cases a fairly rich cluster of galaxies contributes appreciably to the lensing mass (although there are a number of controversial "dark" lenses). It has long been predicted (Turner, Ostriker, and Gott 1984) that there should be many cases of image separations of about one arcsecond (arising from lensing by a single galaxy). The calculation of the expected frequency involves rather few assumptions, and one expects of order 10^{-3} of all quasars in the magnitude range 18–21 to have double images with separations of this order (Ostriker and Vietri 1986); in 60% of these cases the images differ in brightness by less than one magnitude. The Palomar Sky Survey lacks the angular resolution required to distinguish these objects.

The SDSS should have an enormous impact on this field because of its large areal coverage, its high angular resolution, and the large dynamic range of its detectors. We expect there to be ≈ 100 lenses in the spectroscopic survey, and about 1000 in the fainter quasar candidate list (objects with colors of quasars but too faint to be included in the spectroscopic survey). In the latter case, the fraction of candidates that are lenses should be extremely high; obviously two closely-spaced faint stellar objects with similar quasar-like colors are even more likely to be a lensed quasar than any single object with quasar-like colors is to be a single quasar. These predicted numbers are likely to be overestimates because of dust extinction in the lensing galaxies, but the larger lensing cross-sections of early-type, big-bulge systems with little or no dust makes this less important than one might think. In any case, many of the most promising potential uses of gravitational lenses as cosmological probes depend on the analysis of a substantial and statistically well controlled sample of lens systems which the SDSS will produce. We will *not* be able to do spectroscopic follow-up on most of our lens candidates, of course, because of our fiber size and finite minimum fiber separation, but the sample will not be so large as to preclude follow-up on other telescopes, and we do not expect any dearth of outside interest in this follow-up.

g. X-ray Properties of Quasars

More than a quarter century has elapsed since the discovery of X-ray emission from the first quasi-stellar object, 3C273 (Bowyer *et al.* 1970). Since that time, observational progress has been steady. We now have integrated X-ray flux measurements for several hundred QSOs, both from targeted observations of objects selected from pre-existing catalogs, and from *post facto* optical identifications of serendipitously detected X-ray sources. What have we learned?

At the highest level, it is clear that studies of QSOs at X-ray wavelengths are essential. Although there are considerable variations from object to object, in general as much flux is dissipated in soft X-rays as in the visible part of the spectrum (*i.e.*, $f_{(0.5-5\ \mathrm{keV})}/f_{opt} \sim 1$), and *most* of the total, bolometric energy dissipation of a typical QSO emerges as ionizing radiation ($f_{(E>1\ \mathrm{keV})}/f_{(E<1\ \mathrm{keV})} > 1$). Furthermore, the X-ray flux is often found to be strongly variable on a timescale of hours, appreciably faster than variations seen in other regions of the spectrum, presumably implying an origin within tens of A.U. of the center of the system. As this is comparable to the Schwarzschild radius of a $10^9\ M_\odot$ black hole, it seems apparent that X-ray observations probably take us as deep into the central emission region of a QSO as is possible to probe.

It is interesting to consider the impact of X-ray observations on our understanding of the physics of QSOs. As the optical and radio properties of QSOs vary widely from object to object, and in X-rays only accurate broadband flux data, rather than dispersive spectroscopy, are available, workers thus far have studied ensemble X-ray characteristics of samples of objects as a function of their optical or radio characteristics. (We note in passing that this approach may change radically within a few years with the launch of AXAF, XMM, and ASTRO-E, when high-resolution X-ray spectroscopy at X-ray flux levels appropriate for dozens of QSOs will finally become routinely available.) Such X-ray studies have provided a variety of tantalizing facts, but limited global advances in our understanding of QSOs. For example, we understand such macroscopic issues as the fact that L_x/L_{opt} is dependent on L_{opt} (Tananbaum *et al.* 1979), and that radio-loud QSOs have systematically different X-ray spectra and luminosity than radio-quiet objects (Avni and Tananbaum 1986; Schartel *et al.* 1996a, 1996b). Furthermore, most workers now agree that the soft component of the diffuse X-ray background radiation is due in the majority to emission from a superposition of distant point sources, chiefly QSOs and related active galactic nuclei, although there are still significant uncertainties remaining (Fabian and Barcons 1992). In addition, it is clear that Broad Absorption Line QSOs are not detectable X-ray sources (Green and Mathur 1996), and that (broadband) QSO X-ray spectra become complex and diverse at E <1 keV, almost surely indicating multi-component emission processes and/or reflection/reprocessing of emission from multiple sites (Laor *et al.* 1994, Fiore *et al.* 1995). While all of these are wonderful achievements indeed, and have brought the field of X-ray observations of QSOs to a vastly more sophisticated state than the first few 3σ flux detections of 25 years ago, it is probably a fair statement that X-ray observations have not yet yielded fundamental new insights into such areas as QSO energy sources or evolution.

Where are the prospects for future progress in this field? As noted above, the era of routine X-ray spectroscopy at $\lambda/\Delta\lambda \sim 10^3$, finally comparable to optical resolution, can be expected shortly, and will surely provide a huge jump forward. We point out here that the availability on the same timescale of the very large (10^5 object) SDSS QSO sample, with its perfectly homogeneous and high signal to noise optical spectroscopy, will also be a new and vital tool in the study of the X-ray properties of QSOs. Indeed, because at the conclusion of the SDSS, $\sim 90\%$ of *all* known QSOs will be SDSS QSOs, it will be virtually impossible to avoid the issue of interaction between extragalactic X-ray astronomy and the SDSS. The "typical" SDSS QSO, with $18 < r' < 19$, will have $-13 < \log f_x < -12$ erg cm^{-2} s^{-1} in the

(0.5–5.0) keV band (Stocke *et al.* 1991), and is thus very well matched to the capabilities of current X-ray imaging surveys such as ROSAT, and future spectroscopic instruments on AXAF, XMM, and ASTRO-E.

The consequences of this match are multiple, but here we consider as examples two specific issues, ROSAT All Sky Survey (RASS) QSOs, and AXAF/XMM/ASTRO-E X-ray spectroscopic targets. As we noted in Chapter , the amalgam of the RASS data with SDSS will yield broadband X-ray fluxes for $\sim 10^4$ SDSS QSOs, each with a high signal-to-noise ratio optical spectrum obtained with the identical equipment and observing protocol. A smaller but still significant number (*i.e.*, orders of magnitude larger than ever before) will also have X-ray spectral data or at least hardness ratios. For the first time, we will have *hundreds* of QSOs in each redshift or luminosity bin, each with accurate, homogeneous X-ray and optical fluxes, to understand issues of luminosity dependences and evolution. For the first time, we will have a huge, homogeneous set of optical emission line fluxes to compare with QSO X-ray fluxes, a potent arsenal to finally unravel such issues as the elusive "Baldwin effect," a striking but poorly understood correlation of ionizing radiation luminosity with emission line strength (*e.g.*, Green 1996). As AXAF, XMM, and ASTRO-E open the era of QSO X-ray line spectra, the SDSS QSO spectroscopic bank will be a mandatory interpretive tool. We will wish to compare X-ray lines from multiply ionized heavy elements with optical lines of lower excitation; SDSS will provide an "on demand" and beautifully homogeneous data archive of the necessary optical spectroscopic data. X-ray edges and spectral slopes will demand knowledge not just of local galactic N_H for interpretation, but also of the presence, redshift, and opacity of intervening intergalactic absorbers, *e.g.*, the Lyα forest, damped Lyα, and Ly-limit systems; again, SDSS provides these data immediately and uniformly for huge numbers of potential targets.

h. BL Lacertae Objects

BL Lacertae objects are a subclass of active galaxy characterized by nonthermal continua with strong radio and X-ray emission, nearly featureless optical spectra, high polarization, and marked intensity variability at all wavelengths (Kollgaard 1994, Urry and Padovani 1995). Most models for the BL Lac phenomena invoke a relativistic beam viewed nearly along the line of sight (Blandford and Rees 1978), but others (for example) invoke gravitational microlensing of background quasars viewed through foreground galaxies (Ostriker and Vietri 1990).

Because of their rarity (only ∼2% of the objects in current QSO/AGN catalogs are BL Lacs) and unusual properties, large and well-defined samples of BL Lacs are eagerly sought, but until very recently have proven extraordinarily difficult to assemble. Their rarity demands large areal sky coverage, and yet their unusual spectral energy distributions and lack of strong spectral features render past QSO/AGN optical-selection (e.g., UV-excess or objective prism surveys) woefully incomplete for BL Lacs. Despite numerous attempts, there are currently only a very small number of homogeneous BL Lac samples of even modest size, and these few have been assembled via exploitation of strong radio or X-ray emission. These few samples have nonetheless produced some fascinating results. For example, the X-ray-selected sample of the *Einstein* Extended Medium Sensitivity Survey (Morris *et al.* 1991) has provided an initial estimate of the luminosity function for BL Lacs that suggests "negative" evolution; that is, in marked contrast to quasars, BL Lac objects appear to be *less* common or *less* luminous in the past. However, even this tantalizing result relies on only a few dozen objects.

Many recent efforts to obtain well-defined BL Lac samples utilize a combination of multi-wavelength information from the optical, X-ray, and radio bands to obtain high selection efficiency. Stocke *et al.* (1991), Perlman *et al.* (1996), and other groups exploit the fact that BL Lacs occupy distinctive region(s) in flux-ratio plots constructed from such multi-

wavelength information; e.g., BL Lacs fall in distinctive regions of f_r/f_{opt} vs. f_x/f_{opt} plots compared with other object classes. (But note there may be some added complexity in that X-ray- and radio-selected BL Lacs may favor somewhat different regions of such plots, e.g. Brinkmann *et al.* 1996). The combination of SDSS optical flux information (in 5 colors) over π ster of sky, coupled with X-ray information from the RASS, and perhaps radio information from FIRST and other radio surveys, will surely provide an excellent opportunity for selection of a large homogeneous BL Lac sample. In fact, the efficacy of utilizing RASS data for very large area combined X-ray/optical/radio BL Lac selection—or even just optical/X-ray selection—has already been demonstrated (e.g., Kock *et al.* 1996, Nass *et al.* 1996). For example, Nass *et al.* (1996) selected 34 BL Lacs in 6000 deg^2 of sky jointly covered by the RASS and the Hamburg QSO (optical) survey (Wisotzki *et al.* 1996); candidate objects were selected initially from their high ratios of f_x/f_{opt}, characteristic of BL Lacs, and then followed up with optical spectra. These authors estimate that an additional ~ 200 BL Lacs await discovery even in this same 6000 deg^2 of sky to a modest X-ray flux limit of $\sim 10^{-12}$ erg cm^{-2} s^{-1}. Thus in the SDSS survey region of π ster, it is expected that available combinations of optical, X-ray, and perhaps radio flux information will aid in the homogeneous selection of a BL Lac sample comprised of several *hundreds* of objects (compared to the handful of extant homogeneous samples, which contain ∼20–60 objects each).

It is worth noting that flux sensitivities reached in SDSS, RASS, and even FIRST are mutually well-matched for a large-area multiwavelength sky survey for BL Lacs. For example, a BL Lac with an X-ray flux of a few times 10^{-12} erg cm^{-2} s^{-1} might be expected to have an optical magnitude ~ 20, and a radio flux of a few mJy. For many of these BL Lacs, SDSS will not only provide optical flux measures in 5 colors, but also at least initial confirming optical spectra as well. For example, RASS X-ray sources also detected in radio catalogs will be amongst those X-ray sources targeted with highest priority for SDSS optical spectra, and many other FIRST radio sources are also intended for SDSS spectroscopy, independent of whether or not they are X-ray sources.

i. Exotic Objects

With the large size of the quasar sample will come the opportunity to study relatively rare objects. For example, quasars with Broad Absorption Lines (BALs) are ejecting matter at speeds of 10% the speed of light or more. The spectrum of a typical BAL quasar is shown in Figure 3.7.8. Establishing what fraction of quasars exhibit the BAL phenomenon and studying their spectra will reveal information on processes occurring very close to the central engine, in particular the geometry of the quasar emission mechanism, the production of metals in the early Universe, and the physics of outflows.

The sample will also reveal other rare objects which at present are mostly curiosities, but which could contain clues to the evolution of abundances in quasars and their relation to more normal galaxies at high redshift. For example, will some primeval galaxies with emission lines be found in the quasar survey? How will they compare with quasars? An example of an exotic object found in a CCD quasar survey (Schneider, Schmidt, and Gunn 1994) is displayed in Figure 3.7.9. The spectrum of this $z = 3.68$ quasar is dominated by strong, relatively narrow emission lines; the observed equivalent width of the Lyman α line is more than 1500 Å. The SDSS quasar data base should contain hundreds of quasars with characteristics that have rarely, if ever, been detected with current programs.

3.8.5 *Summary*

The SDSS will bring quasar samples up to the size of current galaxy surveys from which the first indications of large scale structure and voids were detected. We may expect similarly important discoveries to arise, now at much higher redshift. In addition, the quasar sample

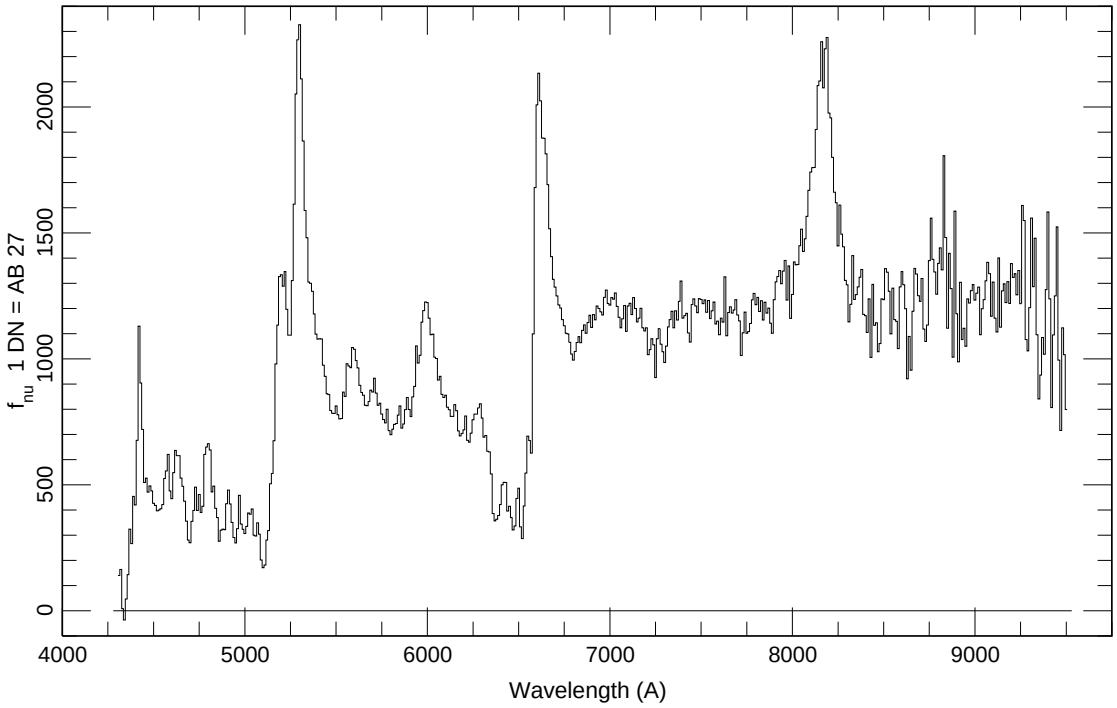

Figure 3.8.8 The spectrum of a Broad Absorption Line (BAL) quasar PC 2330+0125 ($z = 3.28$). The three strongest emission lines, in order of increasing wavelength, are Lyman α, C IV, and C III. Note the deep, broad, blueshifted absorption troughs due to Lyman α and C IV.

will yield a wealth of information on the intrinsic properties of quasars, the evolution of the luminosity function of quasars with time, and the dark intergalactic medium. It is clear that the quasar "spinoff" produced by SDSS represents a great advance over present efforts, and will almost certainly influence our picture of the large scale structure of quasars as significantly as the Shane-Wirtanen catalog did for galaxies.

References

Andreani, P., and Cristiani, S. 1992, *Ap. J. (letters)*, **398**, L13.

Avni, Y., and Tananbaum, H. 1986, *Ap. J.*, **305**, 83.

Bahcall, J.N., Bergeron, J., Boksenberg, A., Hartig, G.F., Jannuzi, B.T., Kirhakos, S., Sargent, W.L.W., Savage, B.D., Schneider, D.P., Turnshek, D.A., Weymann, R.J., and Wolfe, A.M. 1993, *Ap. J. Suppl.*, **87**, 1.

Bahcall, J.N., Bergeron, J., Boksenberg, A., Hartig, G.F., Jannuzi, B.T., Kirhakos, S., Sargent, W.L.W., Savage, B.D., Schneider, D.P., Turnshek, D.A., Weymann, R.J., and Wolfe, A.M. 1996, *Ap. J.*, **457**, 19.

Bahcall, J. N., Maoz, D., Doxsey, R., Schneider, D., Bahcall, N., Lahav, O., and Yanny, B. 1992, *Ap. J.*, **387**, 56.

Bahcall, J.N., and Salpeter, E.E. 1965, *Ap. J.*, **142**, 1677.

Bahcall, J.N., and Soneira, R.M. 1981, *Ap. J. Suppl.*, **47**, 357.

Figure 3.8.9 The spectrum of the strong-lined, $z = 3.68$ quasar PC 1033+4750. The two prominent emission lines are Lyman α and C IV.

Bahcall, N.A., and Chokshi, A. 1991, *Ap. J. (letters)*, **380**, L9.

Baldwin, J.A. 1977, *Ap. J.*, **214**, 679.

Ballinger, W.E., Peacock, J.A., and Heavens, A.F. 1996, preprint (astro-ph/9605017).

Becker, R.H., White, R.L., and Helfand, D.J. 1995, *Ap. J.*, **450**, 559.

Blandford, R. D., and Rees, M. J. 1978, in *Pittsburgh Conference on BL Lac Objects*, ed. A.

N. Wolfe (Pittsburgh U.: Pittsburgh), 328.

Bowen, D. B., Blades, J. C., and Pettini, M. 1996, *Ap. J.*, **464**, 141.

Bowyer, C. S., Lampton, M., Mack, J., and deMendonca, F. 1970, *Ap. J. (letters)*, **161**, L1.

Boyle, B.J., Shanks, T., and Peterson, B.A. 1988, *M. N. R. A. S.*, **235**, 935.

Brinkmann, W., Siebert, J., Kollgaard, R. I., and Thomas, H-C. 1996, *Astr. Ap.*, **313**, 356.

Cavaliere, A., and Padovani, P. *Ap. J. (letters)*, **340**, L5.

Christiani, S. *et al.* 1995, *A&AS*, **112**, 347.

Churchill, C.W., Steidel, C. C. and Vogt, S. S. 1996, *Ap.J.*, in press.

Crampton, D.C. 1992, *The Space Distribution of Quasars*, ASP Conf. Series 21.

Crampton, D.C., Cowley, A.P., and Hartwick, F.D.A. 1987, *Ap. J.*, **314**, 129.

Crampton, D.C., Cowley, A.P., and Hartwick, F.D.A. 1989, *Ap. J.*, **345**, 59.

Efstathiou, G., and Rees, M.J. 1988, *M. N. R. A. S.*, **230**, 5p.

Ellingson, E., Yee, H.K.C., and Green, R.F. 1991, *Ap. J.*, **371**, 49.

Fabian, A. C., and Barcons, X. 1992, *Ann. Rev. Astron. Astrophys.*, **30**, 429.

Fang, Y., Duncan, R. C., Crotts, A., and Bechtold, J. 1996, *Ap. J.*, , 462, 77.

Fiore, F., Elvis, M., Siemiginowska, A., Wilkes, B. J., McDowell, J. C., and Mathur, S. 1995, *Ap. J.*, **449**, 74.

Fisher, K.B., Bahcall, J.N., Kirhakos, S., and Schneider, D.P. 1996, *Ap. J.*, **468**, 469.

Francis, P.J. 1996, *P.A.S.A.* **13**, 212.

Fukugita, M., and Turner, E.L. 1991, *M. N. R. A. S.*, **253**, 99.

Green, P. J. 1996, *Ap. J.*, **467**, 61.

Green, P. J., and Mathur, S. 1996, *Ap. J.*, **462**, 637.

Green, R.F., Schmidt, M., and Liebert, J. 1986, *Ap. J. Suppl.*, **61**, 305.

Greenstein, J.L., and Liebert, J. 1990, *Ap. J.*, **330**, 662.

Grogin, N.A., and Narayan, R. 1996, *Ap. J.*, **464**, 92.

Gunn, J.E., and Peterson, B.E. 1965, *Ap. J.*, **142**, 1633.

Gunn, J.E., and Stryker, L.L. 1983, *Ap. J. Suppl.*, **52**, 121.

Haehnelt, M.G., and Rees, M.J., 1993, *M. N. R. A. S.*, **263**, 168.

Hall, P.B., Osmer, P.S., Green, R.F., Porter, A.C., and Warren, S.J. 1996a, *Ap. J.*, **471**, 1073.

Hall, P.B., Osmer, P.S., Green, R.F., Porter, A.C., and Warren, S.J. 1996b, *Ap. J. Suppl.*, **104**, 185.

Hamann, F., and Ferland, G.J. 1992, *Ap. J. (letters)*, **391**, L53.

Hartwick, F.D.A, and Schade, D. 1990, *Ann. Rev. Astron. Astrophys.*, **28**, 437.

Hazard, C., McMahon, R.G., and Sargent, W.L.W. 1986, *Nature*, **322**, 38.

Hewett, P.C., Foltz, C.B., and Chaffee, F.H. 1993, *Ap. J. (letters)*, **406**, L43.

Hewett, P.C., Foltz, C.B., and Chaffee, F.H. 1995, *A. J.*, **109**, 1498.

Iovino, A., and Shaver, P.A., 1988, in *Large Scale Structure of the Universe*, IAU Symposium 130, J. Audouze *et al.* eds., p. 572.

Irwin, M.J., McMahon, R.G., and Hazard, C. 1991, in *The Space Distribution of Quasars*, ASP Conf. Series 21, D.C. Crampton, ed. (Brigham Young University: Provo), p. 117.

Irwin, M.J., Webster, R.L., Hewett, P.C., Corrigan, R.T., and Jedrzejewski, R.I. 1989, *A. J.*, **98**, 1989.

Jakobsen, P., Perryman, M.A.C., Ulrich, M.H., Machetto, F., and Di Serego Alighieri, S. 1986, *Ap. J. (letters)*, **303**, L27.

Jannuzi, T. T., Hartig, G. F., Kirhakos, S., Sargent, W. L. W., Turnshek, D. A., Weymann, R. J., Bahcall, J. N., Bergeron, J., Boksenberg, A., Savage, B. D., Schneider, D. P. and Wolfe, A. M. 1996, *Ap. J.*, in press.

Jenkins, E.B., and Ostriker, J.P. 1991, *Ap. J.*, **376**, 33.

Kennefick, J.D., Djorgovski, S.G., and de Carvalho, R.R. 1995, *A. J.*, **110**, 2553.

Kennefick, J.D., Djorgovski, S.G., and de Carvalho, R.R. 1996, *A. J.*, **111**, 1816.

Kock, A., Meisenheimer, K., Brinkmann, W., Neumann, M., and Siebert, J. 1996, *Astr. Ap.*, **307**, 745.

Kollgaard, R. I. 1994, *Vistas in Astronomy*, **38**, 29.

Koo, D.C, and Kron, R.G. 1988, *Ap. J.*, **325**, 92.

Kundić, T. 1996, PhD. Thesis, Princeton University.

Kundić, T. *et al.* 1996, preprint (astro-ph/9610162).

La Franca, F., and Christiani, S. 1996, preprint (astro-ph/9612181).

Lanzetta, K. M., Bowen, D. V., Tytler, D. and Webb, J. K. 1995, *Ap. J.*, **442**, 538.

Laor, A., Fiore, F., Elvis, M., Wilkes, B. J., and McDowell, J. C. 1994, *Ap. J.*, **435**, 611.

Lauroesch, J. T., Truran, J. W., Welty, D. E., York, D. G. 1996, *Pub. A. S. P.*, **108**, 641.

Lawrence, C.R., Schneider, D.P., Schmidt, M., Bennett, C.L., Hewitt, J.N., Burke, B.F., Turner, E.L., and Gunn, J.E. 1984, *Science*, **223**, 46.

Liebert, J., Dahn, C.C., and Monet, D.G. 1988, *Ap. J.*, **332**, 891.

Loeb, A. 1993, *Ap. J.*, **403**, 542.

Loeb, A., and Rasio, F.A. 1994, *Ap. J.*, **432**, 52.

Lu, L., Sargent, W. L. W., Barlow, T. A., Churchill, C.W., and Vogt, S.S. 1996, *Ap. J. Suppl.*, **107**, 475.

Lu, L., Wolfe, A.M., and Turnshek, D.A. 1991, *Ap. J.*, **367**, 19.

Lynds, C.R. 1971, *Ap. J. (letters)*, **164**, L73.

Malhotra, S., Rhoads, J., and Turner, E.L. 1996, *M.N.R.A.S.*, submitted (astro-ph/9610233).

Maoz, D., Bahcall, J.N., Schneider, D.P., Bahcall, N.A., Djorgovski, S., Doxsey, R., Gould, A., Kirhakos, S., Meylan, G., and Yanny, B. 1993, *Ap. J.*, **409**, 28.

Maoz, D., Bahcall, J.N., Schneider, D.P., Doxsey, R., Bahcall, N.A., Filippenko, A.V., Goss, W.M., Lahav, O., and Yanny, B. 1992, *Ap. J. (letters)*, **386**, L1.

Maoz, D., and Rix, H.-W. 1993, *Ap. J.*, **412**, 34.

Matsubara, T., and Suto, Y. 1996, *Ap. J. (letters)*, **470**, L1.

Meiksin, A. and Madau, P. 1993, *Ap. J.*, **412**, 34.

Meyer, D. M. and York, D. G. 1987, *Ap. J. (letters)*, **319**, L45.

Meylan, G. 1995, *QSO Absorption Lines*, (Springer-Verlag: Berlin).

Moller, P., and Jakobsen, P. 1990, *Astr. Ap.*, **228**, 299.

Morris, S. L., Stocke, J. T., Gioia, I. M., Schild, R., Wolter, A., Maccacaro, T., and Della Ceca, R. 1991, *Ap. J.*, **380**, 49.

Murdoch, H.S., Hunstead, R.W., Pettini, M., and Blades, J.C. 1986, *Ap. J.*, **309**, 19.

Nass, P., Bade, N., Kollgaard, R. I., Laurent-Muehleisen, S. A., Reimers, D., and Voges, W. 1996, *Astr. Ap.*, , 309, 419.

Ng, Y.K. 1994, PhD. Thesis, Leiden University, The Netherlands

Osmer, P.S. 1981, *Ap. J.*, **247**, 762.

Osmer, P.S. 1982, *Ap. J.*, **253**, 28.

Ostriker, J. P., and Vietri, M. 1986, *Ap. J.*, **300**, 68.

Ostriker, J. P., and Vietri, M. 1990, *Nature*, **344**, 45.

Pascarelle, S. M., Windhorst, R.A., Keel, W. C., and Odewahn, S. C. 1996, *Nature* **383**, 45.

Perlman, E. S., Stocke, J. T., Schachter, J. F., Elvis, M., Ellingson, E., Urry, C. M., Potter, M., Impey, C. D., and Kolchinsky, P. 1996, *Ap. J. Suppl.*, **104**, 251.

Pettini, M., Smith, L. J., Hunstead, R. W., and King, D. L. 1994, *Ap. J.*, **426**, 79.

Press, W.H., Rybicki, G.B., and Schneider, D.P. 1993, *Ap. J.*, **414**, 64.

Quashnock, J., Vanden Berk, D. E., and York, D. G. 1996, *Ap.J.(letters)*, 472, in press.

Rees, M. 1984, *A.R.A.A.* **22**, 471.

Sanders, D.B., Phinney, E.S., Neugebauer, G., Soifer, B.T., and Mathews, K. 1989, *Ap. J.*, **347**, 29.

Sargent, W. L. W., Young, P. J., Boksenberg, A., and Tytler, D. 1980, *Ap. J. Suppl.*, **42**, 41.

Schartel, N. *et al.*, 1996a, *M. N. R. A. S.*, **283**, 1015.

Schartel, N., Walter, R., Fink, H. H., and Trümper, J. 1996b, *Astr. Ap.*, **307**, 33.

Schmidt, M. 1963, *Nature*, **197**, 1040.

Schmidt, M. 1965, *Ap. J.*, **141**, 1295.

Schmidt, M., Schneider, D.P., and Gunn, J.E. 1995, *A. J.*, **110**, 68.

Schneider, D.P., Schmidt, M., and Gunn, J.E. 1991a, *A. J.*, **101**, 2001.

Schneider, D.P., Schmidt, M., and Gunn, J.E. 1991b, *A. J.*, **102**, 837.

Schneider, D.P., Schmidt, M., and Gunn, J.E. 1994, *A. J.*, **107**, 1245.

Shanks, T., Boyle, B.J., and Peterson, B.A. 1988, in *Optical Surveys for Quasars*, ASP Conf. Series 2, P.S. Osmer *et al.* eds. (Brigham Young University: Provo), p. 244.

Shaver, P.A., 1988, in *Large Scale Structure of the Universe*, IAU Symposium 130, J. Audouze *et al.* eds., p. 359.

Shaver, P.A., Wall, J.V., Kellerman, K.I., Jackson, C.A., and Hawkins, M.R.S. 1996, *Nature*, **384**, 439.

Shull, J. M., Stocke, J. T., and Penton, S. 1996, *A. J.*, **111**, 72.

Songaila, A. and Cowie, L. 1996, *A. J.*, **112**, 335.

Steidel, C. C. 1990, *Ap. J. Suppl.*, **72**, 1.

Steidel, C.C., Dickinson, M., and Persson, S.E. 1994, *Ap. J. (letters)*, **437**, L75.

Steidel, C.C., and Sargent, W.L.W. 1991, *Ap. J.*, **382**, 433.

Steidel, C.C., and Sargent, W.L.W. 1992, *Ap. J. Suppl.*, **80**, 1.

Stocke, J. T., Morris, S. L., Gioia, I. M., Maccacaro, T., Schild, R., Wolter, A., Fleming, T. A., and Henry, J. P. 1991, *Ap. J. Suppl.*, **76**, 813.

Storrie-Lombardi, L.J., McMahon, R.G., Irwin, M.J., and Hazard, C. 1996, *Ap. J.*, **468**, 121.

Surdej, J., Magain, P., Swings, J.-P., Borgeest, U., Courvoisier, T. J.-L., Kayser, R., Kellerman, K.I., Kuhr, H., and Refsdal, S. 1987, *Nature*, **329**, 695.

Tananbaum, H. *et al.* 1979, *Ap. J. (letters)*, **234**, L9.

Timmes, J. X., Lauroesch, J.T., and Truran, J. 1995, *Ap. J.*, **451**, 468.

Trevese, D., Kron, R.G., Majewski, S.R., Bershady, M.A. and Koo, D.C. 1994, *Ap. J.*, **433**, 494.

Turner, E.L. 1991, *A. J.*, **101**, 5.

Turner, E.L., Ostriker, J.P., and Gott, J.R. 1984, *Ap. J.*, **284**, 1.

Tytler, D. 1981, *Nature*, **291**, 289.

Tytler, D., and Fan, X.M. 1992, *Ap. J. Suppl.*, **79**, 1.

Tytler, D., and Fan, X.M. 1994, *Ap. J.*, **424**, 87.

Urry, C. M., and Padovani, P. 1995, *Pub. A. S. P.*, **107**, 803.

Wampler, E.J., Robinson, L.B., and Baldwin, J.A. 1973, *Nature*, **243**, 336.

Warren, S.J., and Hewett, P.C. 1990, Rep.Progress.Phys., **53**, 1095.

Warren, S.J., Hewett, P.C., Irwin, M.J., McMahon, R.G., Bridgeland, M.T., Bunclark, P.S., and Kibblewhite, E.J. 1987, *Nature*, **325**, 131.

Warren, S.J., Hewett, P.C., Osmer, P.S., and Irwin, M.J. 1991a, *Ap. J. Suppl.*, **76**, 1.

Warren, S.J., Hewett, P.C., and Osmer, P.S. 1991b, *Ap. J. Suppl.*, **76**, 23.

Warren, S.J., Hewett, P.C., and Osmer, P.S. 1994, *Ap. J.*, **421**, 412.

Webb, J.K., Barcons, X., Carswell, R.F., and Parnell, H.C. 1992, *M. N. R. A. S.*, **255**, 319.

Webster, R.L., Francis, P.J., Peterson, B.A., Drinkwater, M.J., & Masci, F.J. 1995, *Nature*, **375**, 469.

Weymann, R., Morris, S. I., Foltz, C. B., and Hewett, P. C. 1991, *Ap. J.*, **373**, 23.

Wisotzki, L., Koehler, T., Groote, D., and Reimers, D. 1996, *Astr. Ap. Suppl.*, **115**, 227.

Wolfe, A.M., Turnshek, D.A., Smith, H.E., and Cohen, R.D. 1986, *Ap. J. Suppl.*, **61**, 249.

Yanny, B. and York, D.G. 1992, *Ap. J.*, **391**, 569.

Yee, H.K.C., Filippenko, A.V., and Tang, D. 1993, *A. J.*, **105**, 7.

Yi, I. 1996, *Ap. J.*, **473**, 645.

York, D. G. 1994, in Proceedings of the International School of Physics "Enrico Fermi": Course CXXII, ed: J. Silk and N. Vittorio (Amsterdam: North Holland).

York, D. G., Dopita, M., Green, R. and Bechtold, J. 1986, *Ap. J.*, **311**, 610.

3.9. Galaxies

3.9.1 *Introduction*

The optical properties of galaxies are at the core of this proposal. The survey will result in a catalog of 5×10^7 galaxies with good surface photometry in five bands and high-quality

moderate-resolution spectroscopy from $3900 - 9200$Åfor the brightest one million of these. The survey will thus result in an enormous improvement in our understanding of the properties of the ensemble of galaxies and their environment. Examples of problems which can be addressed (Section 3.8.5), are the correlations between galaxy morphologies, colors, and spectral properties, the density-morphology relation, which (in all its nasty detail) lies at the heart of relating the distribution of galaxies to the large-scale distribution of matter, the calibration of distance indicators and the investigation of the dependence of their zero-points on detailed morphology and environment, the luminosity function of galaxies with a sample large enough to investigate *its* detailed dependences on morphology and environment, the statistical incidence of almost any known class of peculiarity, either morphological or spectroscopic, the incidence of interaction and its dependence on environment, such environmental questions as the multiplicity function for groups and its dependence on group morphology, and the dynamics, morphology, and population of binary systems and small groups. In addition, there are several new areas of galaxy studies which will be made possible by the unique features of the data base produced by this survey. These include searches for very compact and for very extended low surface brightness galaxies, the use of the u'-band and z'-band data and the analysis of the detailed spectra as an adjunct to the image morphology. The discussion here is very far from exhaustive; we have selected a few of what we consider to be the most exciting possibilities.

The homogeneity of the spectroscopic survey is of very great concern, as it to a very large extent determines the ultimate usefulness of the spectroscopic sample. The photometric survey will go several magnitudes fainter than the spectroscopic survey so that, in the end, how well the spectroscopic sample is defined depends not on any limitations imposed by the photometric data but on how well we can define our selection criteria, as discussed in Sections 5 and 3.1.2.

Many of us feel that the photometric sky survey alone more than justifies this project. Almost all research on the statistical properties of nearby galaxies has until very recently been based on the Palomar Sky Survey and the ESO-B Survey and on the catalogs obtained from these and other photographic plate archives (de Vaucouleurs, de Vaucouleurs and Corwin 1976; Zwicky *et al.* 1960-68; Lauberts 1982; Sandage and Tammann 1981; Nilson 1973; Lauberts and Valentijn 1989; de Vaucouleurs *et al.* 1991). All of these catalogs have been selected by eye examination of plate material and suffer from incompletenesses and inhomogeneities at many levels. Some recent work has used digital catalogs produced at other wavelengths, most notably the IRAS Point Source Catalog (1988), and very recently, catalogs of galaxies have been produced over large areas from digital scans of photographic plates, with calibration from CCD sequences (Maddox *et al.* 1990a; Collins *et al.* 1989; Picard 1991), but the photometric errors and systematic offsets are much larger than for CCD work. The advantage to extragalactic astronomy of the proposed sensitive, uniform, photometrically and astrometrically accurate, digital sky survey cannot be overstated. In this chapter, we describe the nature of the data, both photometric and spectroscopic, we will obtain for galaxies, and the science we hope to do with it.

3.9.2 *Galaxies and Image Processing*

The photometric pipeline with which we will process the photometric images is described in some detail in Section . Here we describe a few of the aspects specific to galaxy science.

The first issue that comes up is identifying the galaxies themselves, in particular distinguishing them from the much more numerous stars. The two-dimensional image of every object that is identified by the object finder will be fit to three models: a star (i.e., the point-spread function, or PSF), a pure exponential profile (Freeman 1975) of arbitrary inclination, convolved with the PSF, and a pure de Vaucouleurs (1948) profile of arbitrary inclination,

convolved with the PSF. It is feasible, in fact, to fit composite models (i.e., bulge + disk); it is not clear yet, however, how stable the results are, or whether the computer time this would entail is prohibitive. In any case, star-galaxy separation can be done in essentially an optimum way: a galaxy is any object whose goodness of fit to the galaxy models is superior to the PSF fit. We are currently carrying out extensive tests of this approach with model and real images. These fits will be done on the images in all five colors individually and will work with excellent data, since we plan to use only the finest seeing conditions (we expect images of order 1 arcsecond FWHM for the photometric survey; see Chapters 5 and 13). We thus expect the survey to be much more complete than previous catalogs for objects such as very compact galaxies and galaxies with bright starburst and nonthermal nuclei, since the nuclei will have quite different colors from the main body of the galaxy, and the object will be more extended in some bands than in others.

The resulting galaxy counts in different colors will be invaluable for studies of galaxy evolution, large scale structure (section 3.1) and the distribution of Galactic extinction (Chapters and). Table 3.8.1 shows the number of galaxies per steradian per magnitude interval in the J Band, which is close to the photographic band; $J = B - 0.1^m = g' + 0.4^m$. These counts are averaged from several sources, as summarized in Figure 15 of Kron (1980) and Figure 8 of Yoshii and Takahara (1988). Thus in π steradians we expect a total of 10^6 galaxies to J $= 18.7^m$ (corresponding roughly to our spectroscopic limit of $r' \sim 18.2$) and about 5×10^7 galaxies to $J = 22.5^m$, which represents roughly the 5-sigma detection limit for extended objects (galaxies) in the survey.

Table 3.9.1. Galaxy Counts per Steradian per Magnitude

J	g'	A	dlogA/dm
16	15.6	1.49×10^4	0.55
17	16.6	5.16×10^4	0.53
18	17.6	1.70×10^5	0.51
19	18.6	5.29×10^5	0.45
20	19.6	1.34×10^6	0.38
21	20.6	3.06×10^6	0.35
22	21.6	6.73×10^6	0.37
23	22.6	1.65×10^7	0.41

The model-fitting algorithm will classify as galaxies extended objects within our own Galaxy, such as planetary nebulae and highly evolved stars with circumstellar envelopes. There are likely to be few enough of these in the survey region ($\lesssim 100$) that they are an infinitesimal perturbation on the galaxy counts (although they are very interesting on their own account; see Section) and in any case can be distinguished on the basis of their colors.

Of course, model fits will not deal properly with the roughly 2 million objects in our sample which will be saturated (i.e., brighter than $r' \sim 14$). Because we expect at most a few tens of galaxies to saturate in this survey, identifying these as stars is not a terribly difficult problem. However, very bright stars have quite extended wings and diffraction spikes, as well as bleed trails, and the object finder must not split these up into thousands of faint and highly clustered galaxies. We have developed code which characterizes the extended wings of the PSF, and subtracts it off the brightest objects; cf., Figure 4.14 shows the effect. Those

parts of the object for which this subtraction is not effective will be masked; our tests show that the number of spurious objects that remain is negligible.

One of the more problematic aspects of galaxy photometry is deblending: how to deal with multiples of objects whose isophotes overlap. This is of course of concern for star-galaxy separation, because of the possibility that overlapping stellar images can be taken as a pair of galaxies, and the difficulties of doing model-fitting and photometry for a galaxy with superposed star, or a pair of galaxies. This is currently the focus of most of our work on the photometric pipeline, but we have demonstrated that our sector array approach to 2-D photometry (Section) does an impressive job of rejecting stars and cosmic rays in the outer parts of galaxies. In addition, we have developed software to fit local maxima in extended images to a PSF and subtract them if the fit is good, which should be able to remove stars very close to the center of a galaxy. The much more difficult problem of deblending a close pair of galaxies brings up astrophysical questions as well. Most people would agree that NGC 5194 and NGC 5195, the two galaxies which make up M51, the Whirlpool Nebula, should indeed be considered two galaxies. At the opposite extreme, the nucleus of Arp 220 is double on the scale of 1" in the near-IR (Graham *et al.* 1990), but few would argue that this should therefore be considered two galaxies. Where between these two extremes should one draw the line? The answer to this question, which in practice will be a function of redshift, will have repercussions on any number of galaxy statistics, from large-scale structure, to luminosity functions, to number counts.

A related area of concern is the analysis of crowded fields. While these are less of a problem at high Galactic latitudes than at low we will scan through several globular clusters and must be able to recognize when we have done so. Crowded field algorithms for the analysis of star images such as those described by Lupton (1986) and Stetson (1987) would be prohibitively expensive in terms of processing time, and so our approach is to make the code robust enough to recognize when to throw up its hands and mark a region as too confused for further analysis. Of course, most such regions (e.g., globular clusters) will indeed be known about a priori.

Finally, galaxies are extended objects, and thus a galaxy whose center is close to the edge of a CCD strip will not necessarily have its entire image on the frame. In particular, the overlap between adjacent strips is $1'$ (Chapter 5), and thus the photometry of objects that are bigger than this will not be complete. The fraction of galaxies of size d that are affected by this is clearly of the order of d/D, where $D = 13.5'$ is the CCD field width. Integrating this over the diameter distribution of galaxies shows that there will be roughly 3000 galaxies over the survey area which will be affected by this. As we describe in Section 3.1.2, we will target for spectroscopy *all* galaxies in the sample with diameters greater than $1'$, so that we need not worry about performing accurate photometry for those objects at the edge of a scan.

3.9.3 *Galaxy Morphology*

Morphological typing of galaxies is important in a number of applications. For example, the relative mix of galaxy types depends quite strongly on the local number density (Dressler 1980, 1984). One needs to have a quantitative calibration of the morphology-density relation that includes the fact that the luminosity functions of ellipticals and spirals are different and hence that the relative mix varies with the limiting absolute magnitude. Whether or not the observed trends continue into very low density regions is not known, though there are suggestions to the affirmative, and this survey will finally bring into existence a sample big enough to answer questions like this one (*cf.* the discussion in section 3.1.3). It is clear that the question "do the galaxies trace the mass?" is an *a priori* exceedingly ill-posed one, since different kinds of galaxies live in different kinds of places, and the only way to clarify the

situation is to address the detailed morphological mix and its relationship to the density. Studies of the color-morphology relation (Humason, Mayall and Sandage 1956; Bothun, Schommer and Sullivan 1982; Giovanelli and Haynes 1983; Roberts and Haynes 1994) show that galaxies of a given morphological type have a very wide range in $B-V$, and may well have systematically different colors (and therefore have different kinds of stellar populations and mass-to-light ratios) in different kinds of environments. We know as well that the luminosity functions of different kinds of galaxies are different (Sandage *et al.* 1985; Binggeli *et al.* 1988), and all these effects must be considered carefully to answer the much-better posed but still very difficult question "Does the baryonic matter associated with galaxies trace the mass?" We thus consider both the morphological and the color data to be an absolutely necessary part of the survey for large-scale structure studies (cf., the discussion in Section 3.1.3), without which the much more laboriously obtained redshift data are much less useful.

Hubble's (1926) morphological classification scheme for galaxies has proved to be an exceedingly powerful tool for studies of galaxies. His "tuning fork" diagram has been shown to be, to first order, a sequence of bulge to disk ratios, from the diskless elliptical galaxies to the bulgeless late type spirals. This classification scheme has also been shown to have considerable physical significance. Disks are bluer than bulges, since they have continuous star formation while bulges do not, so that the global color of a galaxy (bulge + disk) is roughly correlated with its morphological type. Further, disks are dynamically 'cold', are very thin in the z direction and are supported by rotation, while bulges are dynamically hot, have comparable dimensions in all three coordinates and are supported by the random motions of their stars. These shapes contain fundamental information about the formation history of a galaxy and about the nature of the density fluctuations in the early universe from which the galaxies formed.

The shapes and sizes of galaxies are, however, affected by other things than formation. Galaxies evolve: they merge with neighbors; they accrete gas from smaller neighbors and from continued infall; they lose gas to star formation, to galactic winds and to ram pressure sweeping by intergalactic gas. The colors, size, and shape of a galaxy are related to its dynamical and star-formation history. Ideally we would like to do a study of color, luminosity, and morphology as a function of cosmic time from the formation epoch to the present. We clearly cannot do this with the present survey, with which we will be able to extract some morphological information to redshifts of about 0.2, but we will be able to characterize in detail the *current* morphological mix and its dependence on environment for comparison with the necessarily small samples of very distant objects obtained with large telescopes (*cf.* Appendix). It is not currently possible to do this in any reliable way with any existing database.

As described above, the photometric pipeline will not do a detailed morphological analysis of the objects detected in the images, but will confine itself to two-dimensional model fits to stellar, exponential, and de Vaucouleurs profiles. Our long-term plan is to develop what we call an "atlas-image pipeline", in which we would carry out detailed morphological and photometric analyses of the brighter several million objects in the sample. This can be done after the fact, because the results of these analyses are not needed for the running of the survey itself (e.g., we do not need this information for galaxy target selection), and because the photometric pipeline cuts out atlas images of every detected image, precisely for the purpose of such a posteriori analyses.

As we develop and test such software, and think about the sort of galaxy science that can be done with the images, it is useful to have examples of galaxies with known morphologies as they will appear at different redshifts. These are very useful to see, for example, how a morphological classifier becomes biased with redshift, and thus to see how well we can

distinguish spirals and ellipticals as we approach the photometric limit of the images. We have used a series of images in g and r of Hercules cluster ($z = 0.036$) galaxies taken by Dressler, Gunn, and Schneider. This cluster contains galaxies of all morphological types at a common distance. The redshift of the Hercules cluster is 0.036. The appearance which these galaxies would have at larger redshifts in our survey (from 0.072 to 0.288 in g', and from 0.216 to 0.814 in r') have been simulated by reducing the scale of each image by the appropriate amount, applying the effect of redshift on the magnitude (*i.e.* the opposite of the K-correction), convolving with a point spread function obtained from the stellar imaging, repixelling, and adding noise appropriate for our effective exposure time and quantum efficiency. These redshifted images are also shown in Figure 3.8.1. The luminosities of these objects are such to put them all beyond the spectroscopic limit at $z > 0.2$; as the images show, for $z < 0.2$, spirals and ellipticals can still easily be distinguished, and finer gradations in morphological type can be made in most cases.

We have started to investigate several approaches to the quantitative analysis of morphology, in addition to the exponential and de Vaucouleurs model fits described above.

One such is to look at galaxies in $\theta, \log r$ space; in such a space, logarithmic spiral arms will be straight lines. Frei (1994) has developed code to do such an analysis, and was able to quantify the multiplicity of the arm pattern, its brightness amplitude, and the width of the arms. We hope to modify this code to run in an efficient and stand-alone mode to analyze the literally millions of galaxy images which are of high enough signal-to-noise ratio to allow such an analysis.

Another approach developed by M. Edwards and R, Lupton is to investigate the possibility of fractal structure in spiral galaxy disks using our digital images of nearby spiral galaxies. The results to date are very promising; evidence for fractal structure has indeed been found (a result of great interest on its own). Edwards has devised a method for computing the Hausdorff (1919) measure of a galactic isophote and has shown that it depends on the morphological type and on the bulge to disk ratio. Very roughly, the Hausdorff measure is the ratio of the length of a closed isophote at some scale ϵ to the area enclosed by that isophote. A smooth curve can be successively approximated by polygonal segments of scale ϵ which converge to a finite total length, that of the curve, as $\epsilon \rightarrow 0$. The same procedure carried out on a fractal curve produces a length which increases without bound. The Hausdorff measure thus characterizes the behavior of a curve at scale ϵ. This measure shows a transition between higher and lower values as the isophote intensity is increased and can therefore be used as a measure of the bulge to disk ratio. The method can also be generalized to deal with low resolution, noisy data and can thus be used to compare galaxy properties over a wide range of distance using the SDSS imaging data.

We continue our discussion of galaxy morphology below in Section 3.8.5, but let us first quantify the nature of the spectroscopic data we will obtain for galaxies.

3.9.4 *The Spectroscopic Survey*

The goal of this project is to obtain redshifts for 10^6 galaxies, which involves, of course, obtaining spectra for all of the galaxies. The details of the expected spectral coverage, sensitivity and resolution are discussed in Chapter 11.

What might the spectra from the survey look like? It turns out to be quite difficult to answer this question in detail, because data of the quality which will be obtained in this survey (spectral coverage, signal to noise ratio and resolution) simply do not exist at present for large representative samples of either stars or galaxies. Figure 3.2 shows our partial answer. Spectra of two model galaxies are shown, crudely synthesized following approximately the

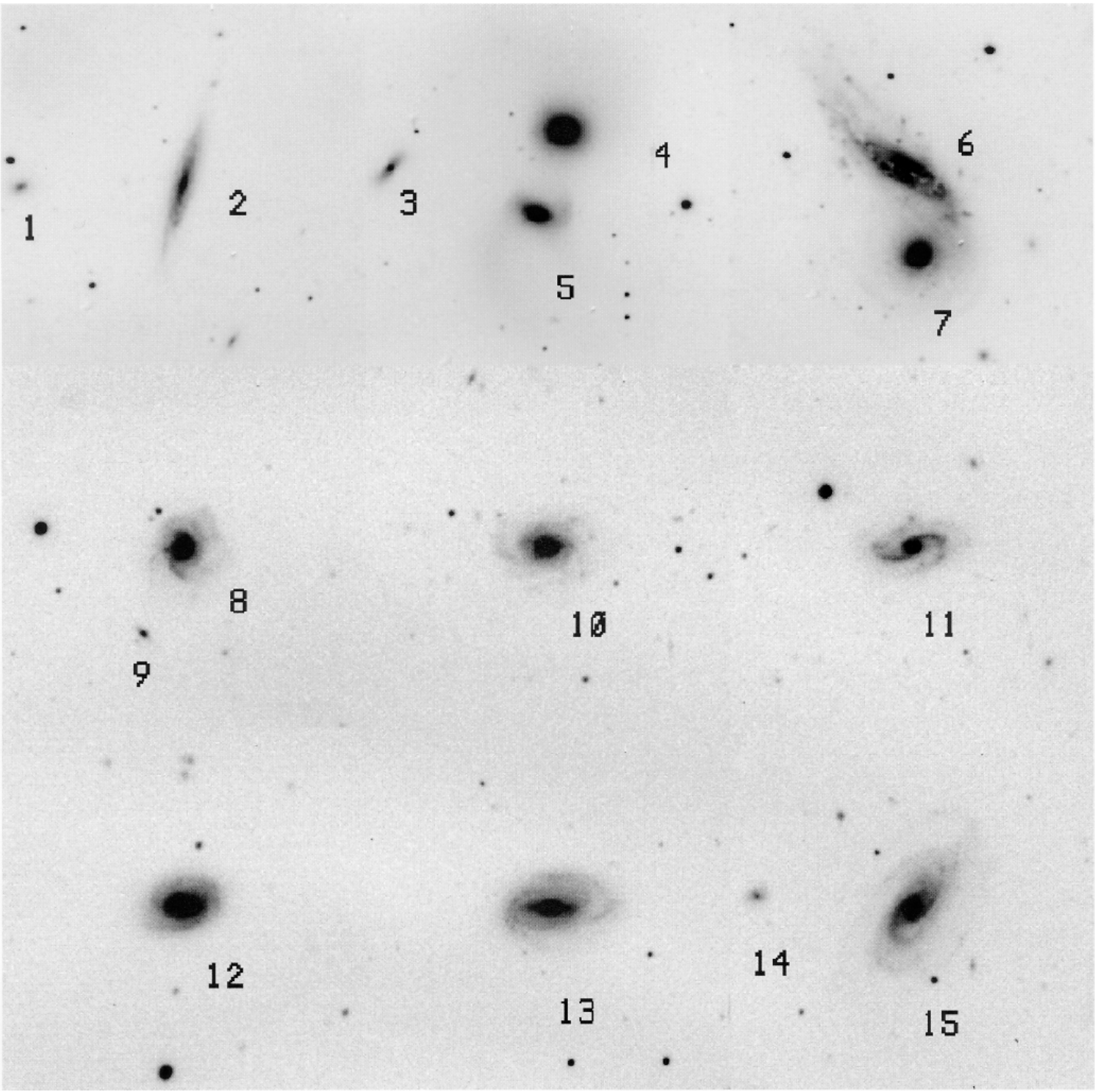

Figure 3.9.1a A mosaic of nine Hercules cluster galaxy fields as they will appear in the survey in g' at the Hercules redshift of 0.036. The panels which have visibly higher noise are noisier than the survey data will in fact be; they were taken with the PFUEI imager on the Hale Telescope with a very bright sky at large zenith angle. The others are four-shooter images taken under very much better conditions; all are in the g band.

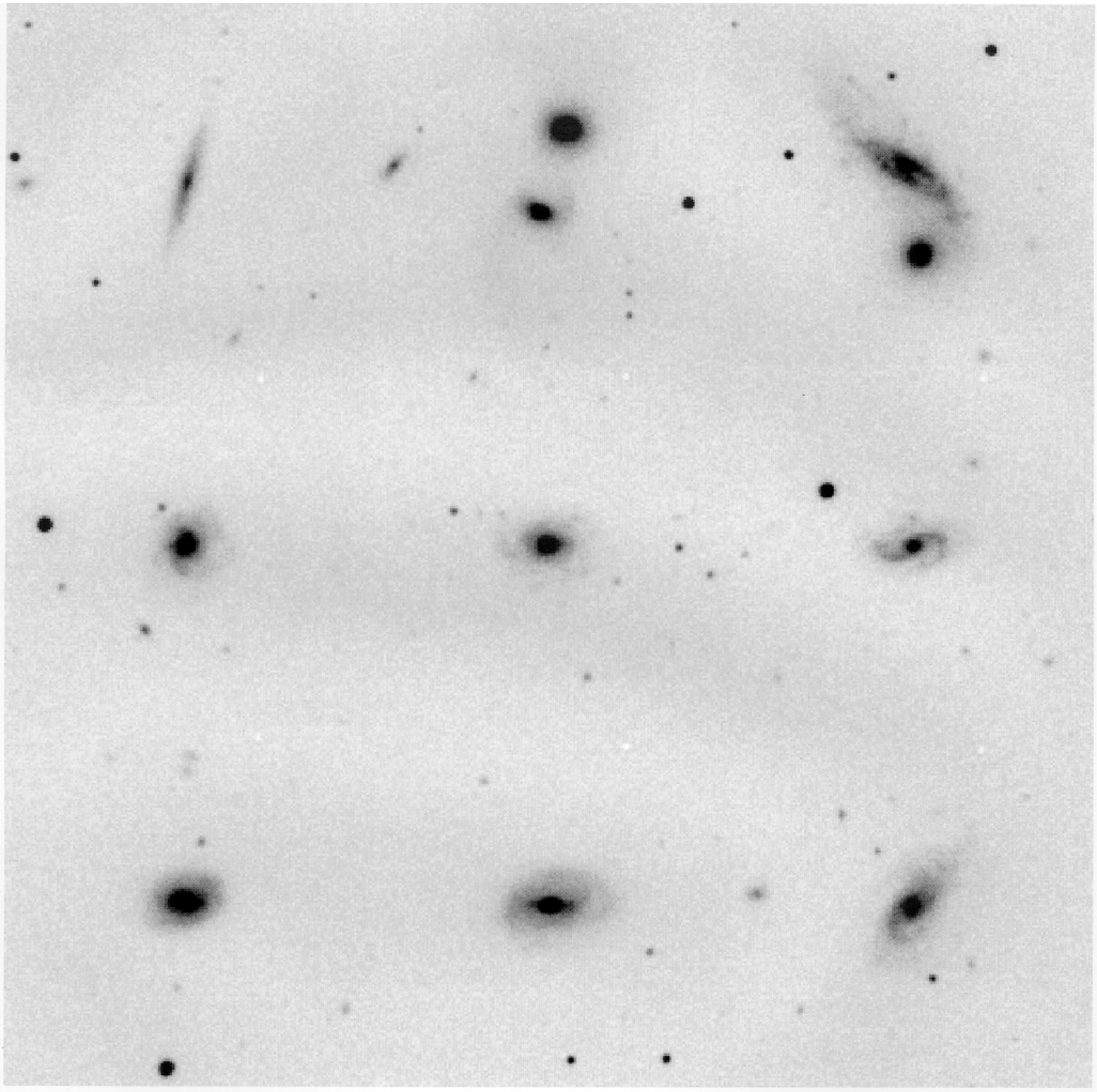

Figure 3.9.1b The galaxies of panel a as they would appear at a redshift of 0.072. Most of the structure in the bright galaxies remains.

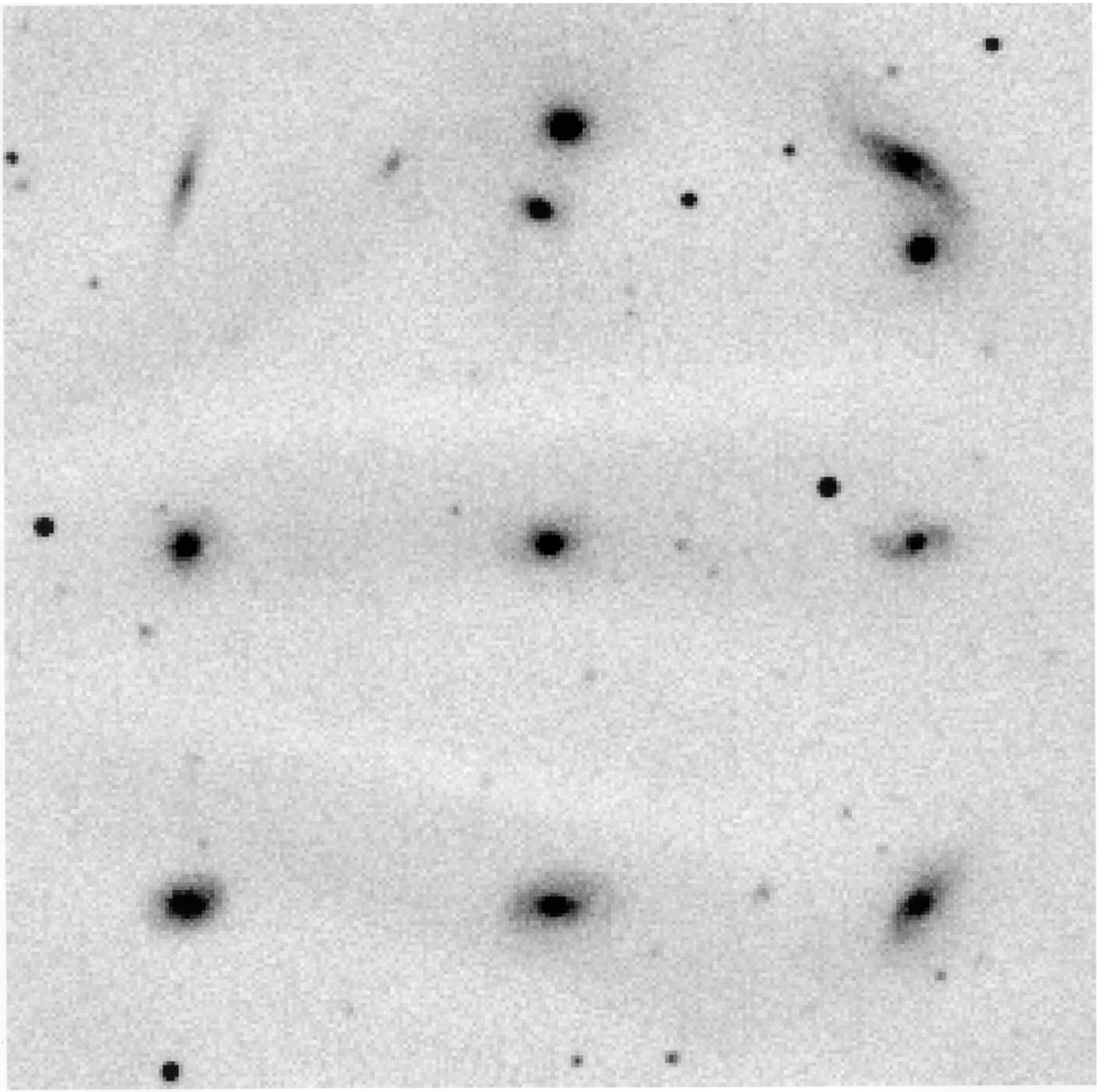

Figure 3.9.1c The galaxies of panel a as they would appear at a redshift of 0.144. The detail in the spiral structure of the bright galaxies is disappearing, but the bulges and disks are still quite distinct, as are the faint extensions in the interacting pair (6,7).

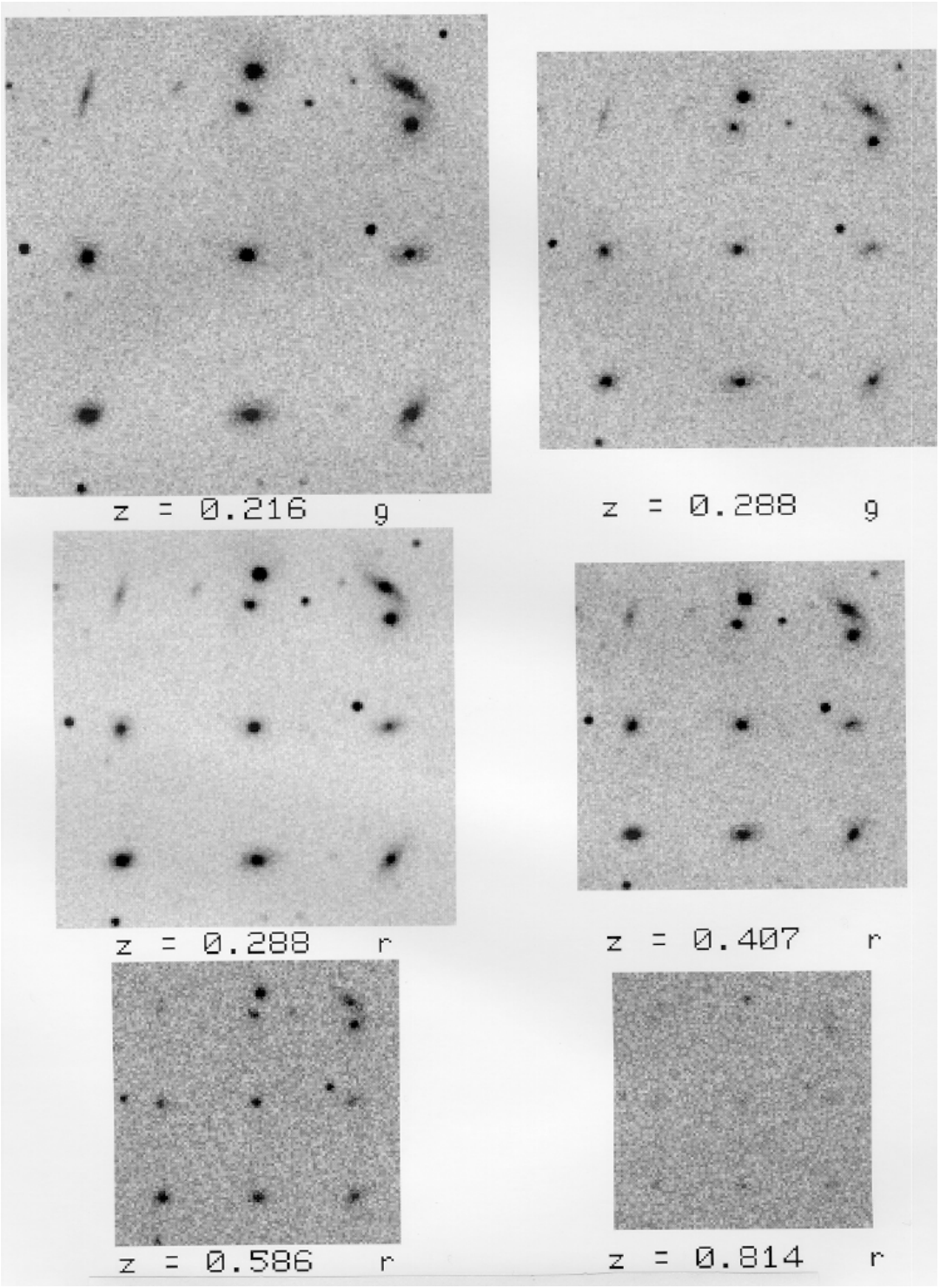

Figure 3.9.1d The galaxies of panel a as they would appear in g' at redshift of 0.216, at which distance most of the bright galaxies are near the survey limit, and 0.288, at which almost all of them have dropped out of the sample. Also shown are simulated r'-band images at redshifts of 0.288, 0.407, 0.586, and 0.814. All these images are to the same scale.

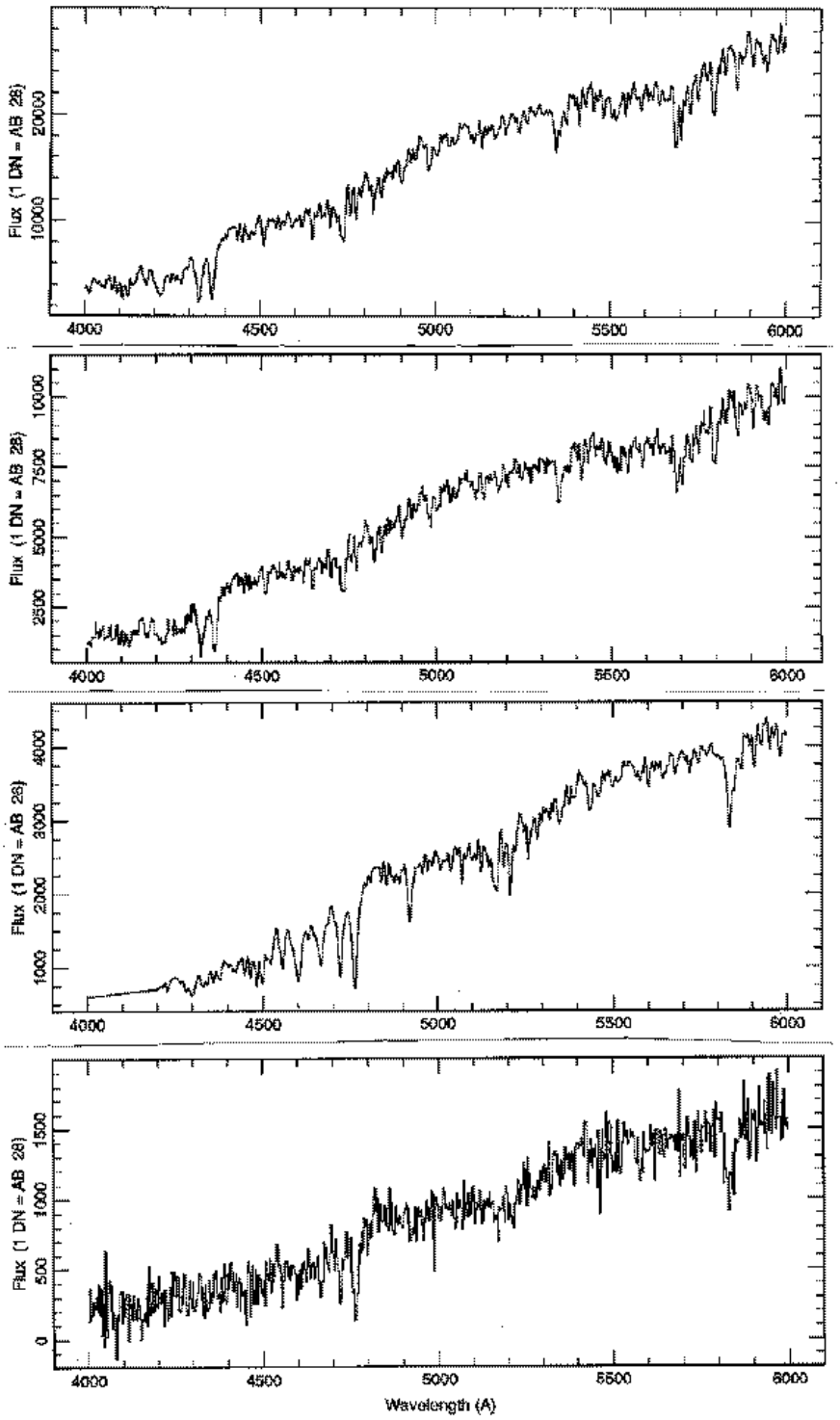

Figure 3.2 Simulated spectra for a $g' = 17.6$ elliptical galaxy and a $g' = 19.7$ spiral, showing the signal input and the output spectra. See text for details.

prescription of Bruzual (1983) from the stellar spectra in the atlas of Jacoby, Hunter and Christian (1984). These spectra cover the wavelength range of 4000 Å to 6000 Å (the blue channel) at a resolution of 1000, half that of the SDSS spectrographs (although the velocity dispersion broadening in a typical galaxy spectrum more than makes up for the difference). The bottom two panels show a model spiral at $z = 0.2$ with $g' = 19.7$, *i.e.* well beyond the survey magnitude limit. This model galaxy has Bruzual's star formation parameter $\mu = 3$, and corresponds to the star formation history expected for a middle spiral galaxy. The top two panels are for a galaxy with $\mu = 5$, (*i.e.* early star formation, as would occur in an E or S0 galaxy), with $g' = 17.6$ and z = 0.1. The top panel of each pair gives the spectral distribution before detection by the telescope, *i.e.* the 'input', noiseless spectrum. The lower panel shows the spectrum as observed by the SDSS, *i.e.* with noise added and as it would appear after sky subtraction (accounting for the spikes in the simulated spectra). The intensity difference between the two is due to the loss incurred by the finite fiber diameter. The bottom-most spectrum in Figure 3.2 is one of the most difficult types of spectra from which to obtain a redshift. The galaxy is faint, the hydrogen and other lines are weak and the 4000 Å break is weak. However, it is more than good enough to obtain a redshift and velocity dispersion accurate to about 20 km s^{-1}, and it represents the expected results from a galaxy more than a magnitude fainter than our average limit. The spectrum of the brighter object illustrates another factor, that one does not do as well as one might expect for brighter galaxies because a smaller fraction of light is captured by the fiber.

An absolutely integral part of this survey will be the acquisition of a large library of stellar spectra taken with the SDSS spectrographs, either during the test year, during the time when the Moon is up, or both. These spectra will act as templates for calculating the redshifts and velocity dispersions of the galaxies and also as raw material for population synthesis modeling. It is important to realize that the excellent five color imaging data which will be obtained from the photometric survey will allow the spectra to be photometrically calibrated. Fiber spectroscopy always carries with it the question of how well centered the fibers really are on the object during the exposure; we will attempt to address this problem by taking four very short and heavily binned exposures with the telescope offset slightly in the four cardinal directions (in elevation and azimuth). These data used in conjunction with the image data should tell us unambiguously for each object where the center of the fiber is, and furthermore allow us to make corrections for chromatic differential refraction to do quite accurate spectrophotometry (see Chapters 5 and 11 for a fuller discussion.)

The spectrophotometry and the high signal to noise ratios for the spectra of even the faintest galaxies in the spectroscopic sample may allow the redshifts and velocity dispersions to be calculated by direct fitting of the stellar spectra to the galaxy spectra in wavelength space (e.g., Rix and White 1992; Connolly *et al.* 1995), instead of using the Fourier quotient or Fourier cross-correlation techniques which have become standard in this field (Sargent *et al.* 1977; Tonry and Davis 1979; Statler 1995). Direct fitting, if it can be done, has several advantages: it makes use of the information in the continuum slope and in the spectral breaks, and it allows the variation of the signal to noise ratio across the spectra to be properly taken into account. The algorithms we are developing are described in Chapter .

The spectroscopic data produced by the SDSS will give, if appropriately calibrated, an invaluable source to study stellar content of galaxies and their evolution. There exist only a few dozen galaxies for which good spectroscopic data with this wide spectral coverage and moderately high dispersion exist; the SDSS data will increase this by several orders of magnitude. The K-correction, which is a basic ingredient in studying distant galaxies, has been calculated from only a few galaxies in each morphological class (Coleman, Wu and Weedman 1980; Kennicutt 1992).

The spectroscopic data from SDSS will allow stellar synthesis models to a level of sophisti-

cation not previously possible: for example, the coverage to 9000Å will allow us the determination of the fraction of early and late M-stars, and fine resolution will tell us about A and F stars. The understanding of stellar populations is crucial for the study of galaxy evolution. Particularly interesting will be the spectral synthesis and star-forming history of irregular galaxies, which are now believed to be the most rapidly evolving population, responsible for the excess of blue galaxies at faint magnitudes (Glazebrook *et al.* 1995).

Our redshift survey will reach beyond $z = 0.15$, corresponding to a look-back time of 2 Gyrs; this may be enough to see the effects of evolution (indeed, Maddox *et al.* 1990 find evidence for galaxy evolution in the number counts at the magnitude limit of the redshift survey). The effects we will be looking for are subtle, but we will have excellent data for large numbers of objects.

3.9.5 *Science*

(a) The Properties of the Ensemble of Galaxies

A number of elementary statistical characteristics of galaxies, such as the luminosity function, the color distribution, the frequency of morphological and profile types, the frequency of characteristic surface brightnesses and the various correlations among galaxy global properties remain poorly determined. This is because there is no large catalog of galaxy image profiles in more than one band for normal field galaxies chosen in some unbiased way. The SDSS will completely change this situation by providing excellent multi-color photometry for an enormous sample of galaxies. The use of the u' and z' filters in the photometric survey will be an especially important contribution, since little galaxy photometry has been done at these wavelengths. The $U - B$ colors of galaxies span a wide range and are critically diagnostic of star formation, and, in the images, of the presence of even small amounts of obscuring dust.

Studies of the galaxy luminosity function and other properties and of evolution effects can be made both on the larger photometric sample of 5×10^7 galaxies and on the spectroscopic sample. The large sample of galaxies to our imaging limit could be subdivided into groups of similar apparent magnitude and apparent size to test for redshift effects and to test how different sample selection changes the apparent galaxy relationships.

The measurement of redshifts allows the physical properties of galaxies, such as their diameters and luminosities, to be determined. To date, the determination of distribution functions such as the galaxy luminosity function have of necessity been quite crude because the statistically reliable samples have been small. Thus, the assumption is often made that the form of the luminosity function is the same in rich clusters as in the field and that different types of giant galaxies have the same shape of luminosity function. It is unlikely that such uniformity actually prevails (*e.g.* Sandage, Binggeli and Tammann 1985) and it is important to determine how the luminosity function does vary. Figure 3.3 shows the best determination to date of the galaxy luminosity function in optical bands (Loveday *et al.* 1992). These data indeed show that the spiral and elliptical luminosity functions are quite distinct. Although the luminosity function is well-determined near the knee, it is uncertain at the bright end, and even more so at the faint end; indeed, the slope of the luminosity function at the faint end has been a point of a great deal of controversy (*cf.* Efstathiou *et al.* 1988 and references therein). This has been a great source of uncertainty in the modeling of the faint galaxy counts and tests for evolution (cf., Koo, Gronwall, and Bruzual 1993; Koo and Kron 1992; Gronwall and Koo 1995). With a total sample of 10^6 galaxies one can subdivide the sample into many groups according to the spectral type, surface brightness, local environment etc. and still have excellent statistics for determining the shape of the luminosity function and

its normalization in luminosity and in amplitude. Similarly, one can construct bivariate luminosity functions, with the second independent variable being color, diameter (*e.g.* Sodré and Lahav 1993), or even flux in a waveband other than the optical. With a sufficiently large number of galaxies, it will be possible to search for features in the luminosity function (aside from the characteristic roll-over at bright luminosities). Such features might be especially evident in the distribution of diameters for disk galaxies grouped according to spectral type, which might be suggestive of physical processes that operate only above some threshold, for example in the surface density of gas in the disk.

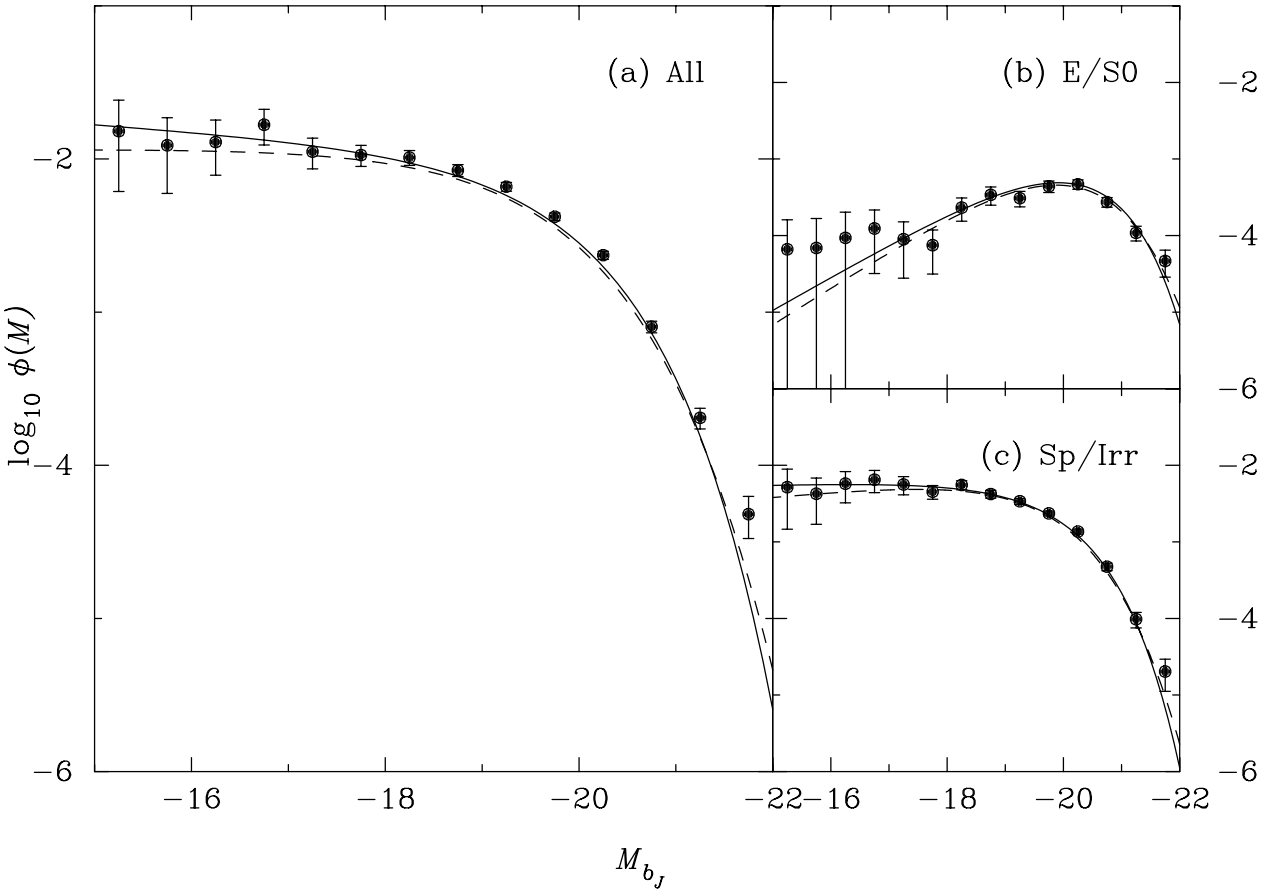

Figure 3.3 The distribution of luminosities derived from a sparse magnitude-limited sample of 1769 galaxies. Note the large statistical errors at faint and at bright luminosities. Note also the difference between the luminosity functions of elliptical and spiral galaxies. The curves drawn are fits to a Schechter (1976) curve. From Loveday *et al.* (1992).

When we start to contemplate the full multivariate distribution function of galaxies as a function of all relevant observable properties, what we are really doing is looking for the underlying physical correlations between these properties. A full understanding of these correlations would go a long way towards putting the morphological classification of galaxies

on a firm physical basis, and would supply invaluable clues for galaxy formation models and the origin of the range of galaxy morphologies. One well-developed technique for quantifying these correlations is Principal Component Analysis, or PCA (cf., Murtagh and Heck 1987). Previous PCA studies performed on samples of bright galaxies (e.g., Whitmore 1984; Efstathiou and Fall 1984; Han 1995) have shown the existence of at least two significant dimensions both for disk and elliptical galaxies, although the parameters included in the analysis are different for the two classes of galaxies (Okamura *et al.* 1989, Kormendy and Djorgovski 1989). The dimensionality and the shape of the manifold must ultimately reflect the physics of galaxy formation and evolution.

However, the small size of the samples used (20-150 galaxies) and the limited range of data for each galaxy leaves many important questions unanswered:

1. How secure is the existence of two significant dimensions and how significant is the third dimension in the statistical sense?

2. What is the relation between the manifolds of disk and elliptical galaxies in the same parameter space?

3. What is the dependence of the manifold on environment?

4. How do star formation rates correlate with the significant dimensions of the manifold?

The SDSS will produce excellent data to investigate these questions. The spectroscopic sample, in particular, will include 10^6 galaxies with superb morphological, spectroscopic, and photometric data, residing in a wide range of environments. The number of galaxies with uniform data available is unprecedented, of course. The complete redshift survey will allow local (i.e., several Mpc) density to be included in the PCA analysis. The spectra will allow a host of spectral diagnostics to be developed, including absorption and emission line strengths and ratios, which can also be correlated with the morphological and photometric properties of the galaxies. It will also become possible to analyze subsamples under different environments in order to see the universality of the manifold. Finding either universality or change among the subsamples would give important clues to the understanding of galaxy formation. Finally, new correlations between distance-dependent and distance-independent quantities will give rise to new distance indicators, which will of course be very useful in peculiar velocity studies (Section 3.1.4).

We can of course apply PCA to the much larger number of galaxies in the deeper photometric sample, although the number of available parameters for each object will be smaller, both because of absence of redshifts and spectra, and because the images will not be as detailed. Distance independent parameters such as colors, surface brightness, and concentration index can be included in the analysis.

This work offers many opportunities for further investigations. For example, photometry from large-scale surveys in other bands, such as the near-infrared (the Two Micron All Sky Survey; cf., Kleinmann 1992), the radio continuum (the FIRST survey; Becker, White, and Helfand 1995; cf., Condon 1992), and the ROSAT X-ray data (cf., Fabbiano 1989) can be included in the PCA analysis, yielding insights into the physical processes that give rise to these emission. On the longer term, this work could evolve to more sophisticated tools than PCA, using eigenvalue techniques such as the Karhunen-Loève (K-L) Transform (for recent astronomical applications, cf., Connolly *et al.* 1995; Vogeley and Szalay 1996; Tegmark, Taylor, and Heavens 1996).

(b) Low Surface Brightness Galaxies

It has long been argued (Disney 1976; McGaugh, Bothun, and Schombert 1995; Sprayberry, Impey, and Irwin 1996) that existing galaxy surveys are strongly biased against finding low surface brightness galaxies, an argument made stronger by the recent discovery of appreciable numbers of large galaxies with central surface brightnesses of $\mu_V > 25$ mag per square arc-sec (Schombert and Bothun 1988; Schombert *et al.* 1992; Dalcanton 1995; Impey *et al.* 1996) and galaxies with dramatically large HI mass to optical luminosity ratios (e.g., Carignan and Beaulieu 1989; Impey *et al.* 1990). The low surface brightness population is very important to quantify, because it may represent an appreciable fraction of the baryonic content of the universe, and have much to tell us about thresholds for star formation and the fragility of disks (cf., Impey and Bothun 1989; Schombert *et al.* 1990). Indeed, given our standard picture of hierarchical clustering and merging in the universe, it is unclear how such low-surface brightness disks can possibly survive. The low-luminosity end of the galaxy luminosity function is quite uncertain, as mentioned above, and tying it down will require a full inventory of low surface brightness systems. Indeed, the discovery of objects like Malin 1 (Impey and Bothun 1989) shows us that there exist populations of low surface brightness galaxies with appreciable integrated luminosities; it is unknown if there are enough of these to appreciably contribute to the galaxy luminosity function at the bright end (McGaugh 1996). It has been suggested that the faint blue galaxies seen in deep photometric surveys are going through a starburst and have faded to low surface brightnesses today (e.g., McGaugh 1994; Ferguson and McGaugh 1995); such a scenario can be constrained only by characterizing the low surface brightness population at the present.

Very little is known about the large-scale distribution of low-surface brightness objects. Although it does not appear that moderate surface brightness galaxies (i.e., those with mean surface brightnesses high enough to appear in the UGC, although still faint with respect to "ordinary" galaxies) fill the voids traced by brighter galaxies (Bothun *et al.* 1986; Thuan, Gott, and Schneider 1987; Eder *et al.* 1989), they do show lower-amplitude clustering on small scales than do "ordinary" galaxies (Bothun *et al.* 1993; Mo, McGaugh, and Bothun 1994; Sprayberry *et al.* 1995). The fragility of their disks would argue that they should not be found in dense environments (Bothun *et al.* 1993), although surveys in clusters have found many galaxies of low surface brightness (e.g., Bothun, Impey, and Malin 1991; Ulmer *et al.* 1996). At the brighter surface brightness limits of the SDSS spectroscopic survey (cf., Section 3.1.2), one can search for differences in the density fields traced by galaxies of different surface brightnesses.

Objects like Malin I present further mysteries. It is one of several low surface brightness galaxies known with a dwarf Seyfert nucleus. It is completely unclear how such a seemingly fragile system can form and feed a massive black hole in its center. These objects are rich in HI but deficient in CO, and have blue colors with very weak star formation (Schombert *et al.* 1990). With only a handful of such galaxies known, it is difficult to come up with models to explain this series of seemingly paradoxical results.

The SDSS should be superb for finding low surface brightness galaxies. The drift-scan mode means that we should be able to flatfield with high precision: if we can flat-field to a (pessimistic) 1% in normal staring mode, then we should be able to do a factor of $\sqrt{N}$ better in drift-scan mode, where $N = 2048$ is the number of pixels across the CCD. Thus we should have systematic problems 9 magnitudes below sky, or at about $\mu_V = 30.7$ mag per square arcsec; in any case, we should be limited by photon statistics for any reasonable object. Thus we find that if we assume a Gaussian profile for these objects, the 5σ detection limit of galaxies with a scale length of $16''$ is $r' = 19.5$ (a factor of 10 fainter than Malin 1!), with a central surface brightness of 27.5 mag arcsec^{-2}. If they have exponential profiles (e.g., McGaugh, Schombert, and Bothun 1995), our efficacy will be somewhat lower. Dalcanton

(1995) has used the drift scan data of Schneider, Schmidt, and Gunn (1989) to demonstrate the advantages of this approach to find low surface brightness objects; her survey is perhaps the deepest yet over substantial areas of the sky. We should be able to go to a surface brightness limit two magnitudes fainter still when the multiple scans of the Southern Stripe (Chapter 5.5) are stacked.

Low surface brightness galaxies are not the only very diffuse extended objects which can be detected in this work; we will also detect a large number of the diffuse, extended high latitude reflection nebulae (e.g., Guhathakurta and Tyson 1989), and high-redshift clusters of galaxies (Dalcanton 1996; Zaritsky *et al.* 1996; cf., Chapter 3.6). Algorithms for finding and distinguishing among these objects are going to be tricky, and this remains work for the future. Photometric measurements of the larger low surface brightness galaxies will be biased low by the determination of the local background sky in the automated reduction of the SDSS images, so that much work will be involved in carrying out simulations and going back to the raw data to calibrate and correct for this effect.

(c) The u′ *Emission from Elliptical Galaxies*

Most elliptical galaxies have red colors ($B-V=1.0$) reflecting their old stellar populations (Faber *et al.* 1989). However, a small fraction (about 5% of elliptical galaxies in a magnitude limited sample) have colors considerably bluer than this, with $B-V \leq 0.8$. These galaxies also tend to contain more cold interstellar matter (detected via far infrared, CO and HI emission; Lees *et al.* 1991) and to be of lower luminosity than the redder galaxies. Is the color-luminosity relation for ellipticals due to metallicity effects or is it indicative of slow continuous star formation? What are the interrelations among metallicity, color and luminosity? Are the blue ellipticals really ellipticals? The combination of excellent colors, the u' band flux densities, the high quality surface brightness/morphology information and the high dispersion spectra (which include information on absorption line strengths, on velocity dispersions and on emission line strengths) for a large number of early type galaxies should allow definitive answers to these questions. They are important for several reasons: what really is the star formation history of the typical elliptical galaxy? How is this affected by environment? How reliable are elliptical galaxies as standard candles and how well can their distances be calibrated? How does star formation work in an elliptical galaxy, which provides a very different environment than does the disk of a spiral galaxy? How did these galaxies form, and what does this tell us about the early structure of the Universe? Can galaxies have formed as recently as recently than $z \sim 1-2$, which some currently fashionable cosmological models predict? This is only one of the many possible studies of galaxies which will be enabled by a catalog of well determined morphologies.

(d) Systematics of Spiral Structure

The SDSS will obtain high resolution photometric images in five colors of several thousand galaxies larger than 2 arcminutes; most of these will be spirals. This will provide a large sample which will allow quantitative studies of spiral structure and allow such quantities as the strength of the spiral pattern to be measured as a function of color. The density wave theory of spiral structure predicts that the spiral pattern should be strong in the dynamically cold material and much weaker in the dynamically hot material (the old disk stars). However, near infrared (2 μm) images of M51 show that the arm-interarm brightness contrast is greater at K than in the visible (Rix and Rieke 1993). The great contrast in the visible can be attributed to the effects of dust and young stars, but it is difficult to explain the infrared result. The images discussed herein should allow this question to be investigated for a very large number of spiral galaxies of many different types and allow us to infer whether the arms are adequately described by more-or-less classical density wave theory or not, and what are the conditions (mass of the galaxy, presence of a satellite, etc).

which give rise to spiral structure in the young and old material.

(e) Active Galactic Nuclei

It is well known that low-level quasar-like activity is found in the nuclei of large numbers of galaxies, and it is estimated that as many as 50% of nearby galaxies show some signs of non-thermal activity in their centers (e.g., Filippenko and Sargent 1985; Ho, Filippenko, and Sargent 1995; Ho 1996), although only $\sim$ 2% of optically selected galaxies are of classical Seyfert type 1 or 2 (Blandford *et al.* 1990). Thus the SDSS spectroscopic survey can be expected to find of the order of 20,000 classical Seyfert galaxies, several times the number currently known (although many might be classified as quasars depending on their luminosity), with superb photometry and spectroscopy for each, and more than an order of magnitude more weaker AGNs.

Active galaxies manifest themselves in a large number of ways. The optical spectra of AGNs often show broad emission lines. Seyfert 2 galaxies have lines with FWHM of order 500 km s^{-1} (appreciably broader than the rotation speeds of all but the most massive galaxies), while Seyfert 1 galaxies show widths of up to many thousands of km s^{-1}.

The SDSS spectra will cover a large range of wavelengths, allowing one to search for many emission lines diagnostic of nuclear activity; the spectral range from [O II]3726, 3729 Å to [S III]9069 Å is potentially observable at low redshifts. Line-ratio diagnostics such as have been developed by Baldwin, Phillips, and Terlevich (1981), Veilleux and Osterbrock (1987) and Osterbrock *et al.* (1992) will select complete samples of AGNs; indeed, the quality and quantity of the SDSS spectra will be such that these diagnostics will be refined substantially. Similarly, the survey will be sensitive to rare types of AGNs: those with uncommon emission lines or emission line ratios. At the $\approx$ 2 Å resolution of the SDSS spectrographs, one can distinguish AGNs by their line widths, and for the higher signal-to-noise ratio spectra, one can search for quite low amplitude AGN components (cf., Filippenko and Sargent 1985). To do this properly, especially for AGNs with luminosities small compared to that of their host galaxies, will require careful subtraction of the underlying stellar continuum spectrum (cf., Filippenko and Sargent 1988), which will require specialized software to be developed as the survey progresses.

The scientific returns from such an analysis are obvious. Models for the formation of emission lines in the narrow and broad line regions of AGNs (cf., Davidson and Netzer 1979; Netzer and Ferland 1984) are rarely faced with high quality spectra covering essentially the entire optical window. Having line strengths and widths for such a range of lines for many thousands of spectra will test these models to their limit, and inspire the next generation of models. We plan to stack the spectra of many individual AGNs of a given class to obtain generic spectra at very high signal-to-noise ratio, which can be used to search for very faint emission lines (cf., Francis *et al.* 1991; Lawrence and McCarthy 1996). We can also use the more sophisticated tools of PCA and the K-L Transform (see above) to find the set of eigenspectra that span the space of observed spectra (cf., Francis *et al.* 1992; Connolly *et al.* 1995). In addition, these studies will be the first complete inventory of active galaxies in our neighborhood. The luminosity function of AGN (or even more interesting, the mass function of the putative black holes at the centers of these galaxies) at the present is wildly uncertain; existing samples are small and are based on inaccurate and ill-defined criteria. This survey will measure the luminosity function directly, and thus yield insights into the evolution of QSO's and their connection to the AGN population. In particular, the increase in the comoving density of quasars as a function of redshift z suggests that inactive or weakly fueled black holes abound in nearby galaxies; the luminosity function of local AGN to faint levels will yield important insights into this question.

Comparison of the spectroscopic and morphological properties of these galaxies may yield insights into the starburst-AGN connection (cf., Filippenko 1992) (especially if emission lines like $H\alpha$ can be decomposed into broad and narrow components), and unified models for AGN. In particular, it remains a mystery why Seyfert galaxies are associated with spirals, while radio galaxies are exclusively found in elliptical galaxies (cf., Balick and Heckman 1982 and the recent discussion in Bahcall *et al.* 1996).

The optical continuum of AGNs is another distinguishing property. It is usually power-law in form, $f_\nu \propto \nu^{-0.7}$. However, in galaxies in which the AGN is comparable or weaker than the stellar light of the host galaxy, it can be difficult to separate the non-stellar and stellar continuum, especially in the presence of reddening. If the non-thermal continuum of the AGN population can be measured, models of the ionizing background at low redshifts will be strongly constrained (cf., the discussion in Williams and Schommer 1993).

AGNs manifest themselves photometrically as well. In many galaxies, they appear as an unresolved nucleus in the stellar profile, becoming more prominent in the blue (cf., Maoz *et al.* 1995). Unresolved nuclei are not a unique indicator of activity, however, as many elliptical galaxies are now being found with unresolved cusps, even at HST resolution (cf., Lauer *et al.* 1995). However, a galaxy with a central region dramatically bluer than the galaxy as a whole is a good AGN candidate.

The Southern photometric survey, with its multiple scans of a single strip, will yield large numbers of candidate AGNs through their variability. This will yield an AGN sample selected completely independently of its spectral or color characteristics, and may be an ideal way to discover previously unrecognized classes of AGN (cf., Trevese *et al.* 1989). This survey will almost certainly discover new types of spectral variability of AGNs with careful measurements on long time baselines.

Finally, AGNs manifest themselves by their emissions in other wavebands, in particular the radio. The VLA 1.4 GHz FIRST survey (Becker, White, and Helfand 1995; cf., Appendix) is covering the same area as the Northern SDSS area. At the 1 mJy flux limit of this survey, the radio source population is nearly an equal mixture of high-redshift radio galaxies, and nearby star-forming galaxies (it is the latter population that exhibits a tight correlation between the radio and far-infrared luminosity; cf., Condon 1992; Crawford *et al.* 1996). Thus those radio galaxies that are optically bright enough to be spectroscopic targets will be mostly star-forming galaxies, while the high-redshift population will be identified photometrically, to be followed up on larger telescopes. Estimates are that 60% of the radio sources will have optical counterparts brighter than the Northern photometric survey limit.

From the ROSAT survey (Section), we will have soft X-ray fluxes, or limits thereon, for all the spectroscopic targets, allowing detailed comparison of the X-ray and optical properties of AGN. The presence of a galaxy with spectral lines indicative of an AGN within the positional error box of a ROSAT source should make the identification of that source unambigous, of course!

By assembling a uniform sample of AGNs with emission-line strengths and photometric and morphological data for each, it will be possible to explore the full range of AGN phenomena, and in particular, to find correlations between properties which point to physical processes which unify the plethora of AGN types. This database will become particularly powerful when combined with the surveys in other wavebands mentioned above. With such a complete accounting of the AGN sample, we will be able to calculate accurate luminosity functions in many wavebands, which is tremendously important for modeling the X-ray background (e.g., Fabian and Barcons 1992) and the evolution of quasars (e.g., Crampton 1992), among other things.

(f) Evolution of the Galaxy Population

It has become increasingly clear in the last decade that the faint number counts of galaxies are inconsistent with no-evolution or even passive evolution models (*e.g.* Tyson 1988), and it has been suggested from galaxy number counts that the evolution is evident even at $b_J = 18$ (Maddox *et al.* 1990b). That is, number evolution of galaxies may indeed be apparent within the redshift survey sample (*cf.* Figure 3.1.4). Redshift surveys of faint galaxies have failed to detect the signature of evolution in the *shape* of the redshift histogram, although the *amplitude*, and emission-line strength, does show signs of evolution (Colless *et al.* 1990). Models to explain this have invoked starbursts, dramatic merging, and a positive cosmological constant (*cf.* Koo and Kron 1992); it is now becoming clear that irregular galaxies have undergone much more dramatic evolution recently than have ordinary spiral and elliptical galaxies (Glazebrook *et al.* 1995; Fukugita *et al.* 1996). The SDSS photometric and redshift survey will of course not go as deep as one would like to probe evolutionary effects at the highest redshifts; this is really the domain of surveys on much larger telescopes, such as DEEP (Mould 1993; *cf.* Appendix), and of course imaging surveys from HST. However, there are a number of aspects of the evolution problem that the SDSS is well suited to tackle, especially in the Southern survey.

With detailed number counts in several bands to the photometric limits of the survey, one can put firm constraints on models for galaxy evolution, at least at relatively low redshifts. The Southern photometric survey will probe to redshifts of order unity, and will give the best data for evolutionary probes. Matching the photometric catalog with catalogs drawn from sky surveys in other wavebands, from X-ray to radio (*cf.* Appendix) will allow models of the evolution of flux in these wavebands to be developed and tested.

The faint galaxy count data have been interpreted as implying the existence of a new population of blue galaxies appearing at $B = 21$ or so (*e.g.* Lilly *et al.* 1991), which seems to show weaker clustering than do the red galaxies, at least at fainter magnitudes (Efstathiou *et al.* 1991). This will be testable directly from the southern photometric data; although it does not go as faint as does the data of Efstathiou *et al.*, it covers a *much* greater area and number of galaxies, and we will be able to measure correlations in narrow magnitude and color slices. However, Koo *et al.* (1993) have argued that our knowledge of the luminosity function as a function of passband and morphological type at the present is so uncertain that one can fit all existing data to models with very little evolution, and without needing to invoke a new population. The SDSS photometric and redshift surveys will tie down the luminosity function of nearby galaxies once and for all (see above), and will allow detailed modeling along the lines of Koo *et al.*, but at a far more sophisticated level.

References

Bahcall, J. N., Kirhakos, S., Saxe, D., and Schneider, D. P. 1996, *Ap. J.*, in press.

Baldwin, J. A., Phillips, M. M., and Terlevich, R. 1981, *Pub. A. S. P.*, **93**, 5.

Balick, B., and Heckman, T. M. 1982, *Ann. Rev. Astron. Astrophys.*, **20**, 431.

Becker, R., White, R., and Helfand, D. 1995, *Ap. J.*, **450**, 559.

Binggeli, B., Sandage, A., and Tammann, G. A. 1988, *Ann. Rev. Astron. Astrophys.*, **26**, 509.

Blandford, R. D., Netzer, H., and Woltjer, L. 1990, *Active Galactic Nuclei*, Saas-Fee Advanced Course 20 (New York: Springer-Verlag).

Bothun, G. D., Beers, T. C., Mould, J. R., and Huchra, J. P. 1986, *Ap. J.*, **308**, 510.

Bothun, G. D., Impey, C. D., and Malin, D. F. 1991, *Ap. J.*, **376**, 404.

Bothun, G. D., Schombert, J. M., Impey, C. D., Sprayberry, D., and McGaugh, S. S. 1993,*A. J.*, **106**, 530.

Bothun, G.D., Schommer, R.A. and Sullivan, W.T. 1982, *A.J.* **87**, 731.

Bruzual, A.G. 1983, *Ap.J. (Suppl.)* **53**, 497.

Carignan, C., and Beaulieu, S. 1989, *Ap. J.*, **347**, 760.

Cen, R., and Ostriker, J. P. 1993, *Ap. J.*, **417**, 415.

Coleman, G.D., Wu. C.-C., and Weedman, D.W. 1980, *Ap. J. Suppl.*, **43**, 393.

Colless, M., Ellis, R., Taylor, K., and Hook, R. 1990, *M. N. R. A. S.*, **244**, 408.

Collins, C.A., Heydon-Dumbleton, N.H., and MacGillivray, H.T. 1989, *M. N. R. A. S.*, **236**, 7P.

Condon, J. J. 1992, *Ann. Rev. Astron. Astrophys.*, **30**, 575.

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

Crampton, D., editor 1992, *The Space Distribution of Quasars*, Astronomical Society of the Pacific Conference Series # 21.

Crawford, T., Marr, J., Partridge, B., and Strauss, M. A. 1996, *Ap. J.*, **460**, 225.

Dalcanton, J. J. 1995, PhD Thesis, Princeton University.

Dalcanton, J. J. 1996, *Ap. J.*, **466**, 92.

Davidson, K., and Netzer, H. 1979, *Rev. Mod. Phys.* **51**, 715.

de Vaucouleurs, G., de Vaucouleurs, A. and Corwin, H.C. 1975, *Second Reference Catalogue of Bright Galaxies*, University of Texas Press, Austin.

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

de Vaucouleurs, G., de Vaucouleurs, A. Corwin, H.C., Buta, R.J., Paturel, G., and Fouqué, P. 1991, *Third Reference Catalogue of Bright Galaxies Vols I-III*, Springer-Verlag Co., New York.

Disney, M.J. 1976, *Nature*, **263**, 573.

Dressler, A. 1980, *Ap.J.* **236**, 351.

Dressler, A. 1984, *Ann. Rev. Astron. Astrophys.*, **22**, 185.

Eder, J. A., Schombert, J. M., Dekel, A., and Oemler, A. 1989, *Ap. J.*, **340**, 29.

Efstathiou, G., Bernstein, G., Katz, N., Tyson, J.A., and Guhathakurta, P. 1991, *Ap. J. (letters)*, **380**, L47.

Efstathiou, G., Ellis, R.S., and Peterson, B.A. 1988, *M. N. R. A. S.*, **232**, 431.

Efstathiou, G., and Fall, S. M. 1984, *M. N. R. A. S.*, **206**, 453.

Fabbiano, G. 1989, *Ann. Rev. Astron. Astrophys.*, **27**, 87.

Faber, S.M., Wegner, G., Burstein, D., Davies, R.L., Dressler, A., Lynden-Bell, D., and Terlevich, R.J. 1989, *Ap.J. (Suppl.)* **69**, 763.

Fabian, A.C., and Barcons, X. 1992, *Ann. Rev. Astron. Astrophys.*, **30**, 429.

Filippenko, A. V. editor, 1992, *Relationships between Active Galactic Nuclei and Starburst Galaxies*, Astronomical Society of the Pacific Conference Series # 31.

Filippenko, A.V. and Sargent, W.L.W. 1985, *Ap. J. Suppl.*, **57**, 503.

Filippenko, A.V. and Sargent, W.L.W. 1988, *Ap. J.*, **324**, 134.

Francis, P. J., Hewett, P. C., Foltz, C. B., Chaffee, F. H., Weymann, R. J., and Morris, S. L., 1991, *Ap. J.*, **373**, 465.

Francis, P. J., Hewett, P. C., Foltz, C. B., and Chaffee, F. H., 1992, *Ap. J.*, **398**, 476.

Freeman, K.C. 1975, in *Galaxies and the Universe*, ed. A. Sandage, M. Sandage and J. Kristian, University of Chicago Press, 409.

Frei, Z. 1994, PhD. Thesis, Princeton University.

Fukugita, M., Hogan, C.J., & Peebles, P.J.E. 1996, *Nature*, **381**, 489.

Giovanelli, R., and Haynes, M.P. 1983, *A.J.* **88**, 881.

Glazebrook, K., Ellis, R., Santiago, B., and Griffiths, R. 1995, *M. N. R. A. S.*, **275**, 19.

Graham, J.R., Carico, D.P., Matthews, K., Neugebauer, G., Soifer, B.T., and Wilson, T.D. 1990, *Ap. J. (letters)*, **354**, L5.

Gronwall, C., and Koo, D. C. 1995, *Ap. J. (letters)*, **440**, L1.

Guhathakurta, P., and Tyson, J.A. 1989, *Ap. J.*, **346**, 773.

Han, M. 1995, *Ap. J.*, **442**, 504.

Hausdorff, F. 1919, *Mathematische Annaler* **79**, 157.

Ho, L. 1996, *Pub. A. S. P.*, **108**, 637.

Ho, L., Filippenko, A. V., and Sargent, W. L. W. 1995, *Ap. J. Suppl.*, **98**, 477.

Hubble, E.P. 1926, *Ap.J.* **64**, 321.

Humason, M.L., Mayall, N.U. and Sandage, A.R. 1956, *A.J.* **61**, 97.

IRAS Point Source Catalog, version 2. NASA Reference Publications.

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

Impey, C., Bothun, G., Malin, D., and Staveley-Smith, L. 1990, *Ap. J. (letters)*, **351**, L33.

Impey, C. D., Sprayberry, D., Irwin, M. J., and Bothun, G. D. 1996, *Ap. J. Suppl.*, , 105, 209.

Jacoby, G.H., Hunter, D.A. and Christian, C.A. 1984, *Ap.J. (Suppl.)* **56**, 257.

Kennicutt, R. C. 1992, *Ap. J. Suppl.*, **79**, 255.

Kleinmann, S. G. 1992, in *Robotic Telescopes for the 1990's*, edited by A. Filippenko, Astronomical Society of the Pacific Conference Series # 34, 203.

Koo, D.C., Gronwall, C., and Bruzual, G.A. 1993, *Ap. J. (letters)*, **415**, L21.

Koo, D., and Kron, R. 1992, *Ann. Rev. Astron. Astrophys.*, **30**, 613.

Kormendy, J., and Djorgovski, S. 1989, *Ann. Rev. Astron. Astrophys.*, **27**, 235.

Kron, R.G. 1980, *Ap.J. (Suppl.)* **43**, 305.

Kron, R.G. 1982, *Vistas in Astronomy* **26**, 37.

Lauberts, A. 1982, ESO/Uppsala Catalogue of the ESO-B Sky Survey (Munich: European Southern Observatory).

Lauberts, A., and Valentijn, E.A. 1989, The Surface Photometry Catalogue of the ESO-Uppsala Galaxies, (Munich: European Southern Observatory).

Lauer, T. R., Ajhar, E. A., Byun, Y.-I., Dressler, A., and Faber, S. M. 1995, *A. J.*, **110**, 2622.

Lawrence, C., and McCarthy, P. 1996, *Ap.J.*, submitted.

Lees, J.F., Knapp, G.R., Rupen, M.P., and Phillips, T.G. 1991, *Ap.J.* **379**, 177.

Lilly, S.J., Cowie, L.L., and Gardner, J.P. 1991, *Ap. J.*, **369**, 79.

Loveday, J., Peterson, B.A., Efstathiou, G., and Maddox, S.J. 1992, *Ap. J.*, **390**, 338.

Lupton, R.H. 1986, Ph.D. Thesis, Princeton University.

Maddox, S.J., Efstathiou, G., Sutherland, W.J., and Loveday, J. 1990a, *M. N. R. A. S.*, **243**, 692.

Maddox, S.J., Sutherland, W.J., Efstathiou, G., Loveday, J., and Peterson, B.A. 1990b, *M. N. R. A. S.*, **247**, 1P.

Maoz, D., Filippenko, A. V., Ho, L. C., Rix, H.-W., Bahcall, J. N., Schneider, D. P., and Macchetto, F. D. 1995, *Ap. J.*, **440**, 91.

McGaugh, S. 1994, *Nature*, **367**, 538.

McGaugh, S. 1996, *M. N. R. A. S.*, **280**, 337.

McGaugh, S., Bothun, G. D., and Schombert, J. M. 1995, *A. J.*, **110**, 573.

McGaugh, S., Schombert, J. M., and Bothun, G. D. 1995, *A. J.*, **109**, 2019.

Mo, H. J., McGaugh, S. S., and Bothun, G. D. 1994, *M. N. R. A. S.*, **267**, 129.

Mould, J.R. 1993, ASP Conference Series 43, 281.

Murtagh, F., and Heck, A. 1987, *Multivariate Data Analysis* (Dordrecht: Reidel).

Netzer, H., and Ferland, G. J. 1984, *Pub. A. S. P.*, **96**, 593.

Nilson, P. 1973, *Nova Acta Roy. Soc. Uppsala*, Ser 5A, Vol **1A**.

Okamura, S., Kodaira, K., and Watanabe, M. 1989, in *The World of Galaxies*, eds. H. G. Corwin and L. Bottinelli (New York: Springer-Verlag), 75.

Osterbrock, D.E., Tran, H.D., and Veilleux, S. 1992, *Ap. J.*, **389**, 207.

Picard, A. 1991, *Ap. J. (letters)*, **368**, L7.

Rix, H.-W., and White, S.D.M. 1992, *M. N. R. A. S.*, **254**, 389.

Roberts, M. S., and Haynes, M. P. 1994, *Ann. Rev. Astron. Astrophys.*, **32**, 115.

Sandage, A., Binggeli, B. and Tammann, G.A. 1985, *A.J.* **90**, 1759.

Sandage, A., and Tammann, G.A. 1981, *A Revised Shapley-Ames Catalogue of Bright Galaxies* (Washington DC: Carnegie Institute of Washington) (RSA).

Sargent, W.L.W., Schechter, P.L., Boksenberg, A., and Shortridge, K. 1977, *Ap.J.* **212**, 326.

Schechter, P. 1976, *Ap.J.* **203**, 297.

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

Schombert, J.M. and Bothun, G.D. 1988, *A. J.*, **95**, 1389.

Schombert, J. M., Bothun, G. D., Impey, C. D., and Mundy, L. G. 1990, *A. J.*, **100**, 1523.

Schombert, J. M., Bothun, G. D., Schneider, S. E., and McGaugh, S. 1992, *A. J.*, **103**, 1107.

Sodré, L., and Lahav, O. 1993, *M. N. R. A. S.*, **260**, 285.

Sprayberry, D., Impey, C. D., Bothun, G. D., and Irwin, M. J. 1995, *A. J.*, **109**, 558.

Sprayberry, D., Impey, C. D., and Irwin, M. J. 1996, *Ap. J.*, **463**, 535.

Statler, T. 1995, *A. J.*, **109**, 1371.

Stetson, P. 1987, *P.A.S.P.* **99**, 191.

Tegmark, M., Taylor, A., and Heavens, A. 1996, *Ap.J.*, submitted (astro-ph/9603021).

Thuan, T. X., Gott, J. R., and Schneider, S. E. 1987, *Ap. J. (letters)*, **315**, L93.

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

Trevese, D., Pitella, G., Kron, R.G., Koo, D.C., and Bershady, M. 1989, *A. J.*, **98**, 108.

Tyson, J.A. 1988, *A.J.* **96**, 1.

Ulmer, M. P., Bernstein, G. M., Martin, D. R., Nichol, R. C., Pendleton, J. L., and Tyson, J. A. 1996, *AJ*, in press (astro-ph/9610040).

Veilleux, S., and Osterbrock, D.E. 1987, *Ap. J. Suppl.*, **63**, 295.

Vogeley, M. S., and Szalay, A. S. 1996, *Ap. J.*, **465**, 34.

Whitmore, B. 1984, *Ap. J.*, **278**, 61.

Williams, T., and Schommer, R. 1993, *Ap. J. (letters)*, **419**, L53.

Yoshii, Y., and Takahara, F. 1988, *Ap.J.* **326**, 1.

Zaritsky, D., Nelson, A.E., Dalcanton, J.J., & Gonzalez, A.H. 1996, preprint (astro-ph/9612021).

Zwicky, F., Herzog, E., Wild, P., Karpowicz, M. and Kowal, C.T. 1960-1968, *Catalog of Galaxies and Clusters of Galaxies* **1-6**, Carnegie Institute of Washington.

4. Overview and Project Status

4.1. Summary

This chapter describes the status of the many components of the Sloan Digital Sky Survey (SDSS) as of Thanksgiving 1996; much more detail on the design, construction and performance of the individual components is given in the following chapters. An overview of the entire project in flow-diagram form, from light entering the telescope to data archived for science analysis, is given in Figure 4.1. The photometric data are collected at the Apache Point Observatory (APO) by the 2.5 meter telescope with its camera, and shipped by express carrier to Fermi National Accelerator Laboratory (FermiLab) for reduction, together with photometric calibration data from the Monitor Telescope (MT). The reduced photometric and astrometric data with their calibration information are written to the Operational Data Base (ODB) and to the Science Data Base (SDB); the former runs the survey and the latter is a read-only archive accessible by the project participants for science analysis. The target selection pipeline then selects astronomical objects for spectroscopy, as described in Chapter 3, and designs a fiber plug plate for each spectroscopic field. The plug plates are drilled, plugged at APO, and the spectroscopic fields observed. The spectroscopic data are reduced at FermiLab and the spectra and redshifts recorded in the data bases.

4.2. Site Infrastructure

The original installation of utilities and infrastructure at APO, which began in 1984, was planned to accommodate three large telescopes, of which the first was the Astrophysical Research Corporation (ARC) 3.5 meter. The site, and its suitability for this project, are described in detail in Chapter 4.2. The long-term plans filed with the National Forest Service, which controls the permit for use of the site, included plans for additional dormitories, roads and office space as later expansion items. The details of this construction were allowed to be driven by the requirements of the new projects. It was thus straightforward to accommodate the needs of the SDSS project as an expansion of the site within these plans.

The SDSS telescope enclosure was erected on its site in Summer 1994, and the necessary utility connections and hookups were made at this time. At the same time, the main operations building for the 3.5 meter telescope was expanded by 3000 square feet to provide

additional laboratory, office, operations and meeting space. The original APO darkroom was rebuilt as a laboratory for handling SDSS and has now been upgraded to clean-room status in readiness for the arrival of the photometric camera in Spring 1997. A new dormitory was built. The enclosure for the monitor telescope (MT) was erected ENE of the 2.5 meter telescope enclosure and a building ESE of the 2.5 m telescope was erected for the handling of the fiber plug plates. Additional trenches were dug for cabling and utilities for these new buildings.

The loading dock to the operations building was improved to allow easy movement of instruments back and forth between the 2.5 meter telescope and the laboratories. Provision for a large SDSS control room was added to the operations building, and enough new lab space added to provide adequate support for both telescopes. Extra kitchen, laundry and office space is now in place, and has allowed the average steady state personnel presence to grow from 5 to 10. The extra personnel load during the integration and commissioning of the SDSS hardware will be accommodated by off-site living space and on-site trailers to serve as temporary office space.

The site is served by a leased T1 communication channel to New Mexico State University (NMSU) and from there via Internet to the ARC/SDSS institutions. This link will be shared by SDSS for project personnel communications.

Figure 4.2 shows the site from the south. The closest buildings are the white 2.5 m telescope enclosure and, to its right, the plugplate support building. The upper part of the 2.5 m telescope itself can be seen to the left. The small silver dome behind the plug plate building houses the Monitor Telescope, and the larger white dome housing the ARC 3.5 meter telescope is seen to the upper right. Figure 4.3 shows the 2.5 m telescope on its pier and looks west over the White Sands National Park.

Located under the site are an electric ground plane, a sewage drainage facility and a water storage and distribution system for fire safety. The APO staff participate in the local volunteer fire departments and safety organizations which are essential for site safety in remote regions. Drinking water is purchased from the National Solar Observatory (NSO) in Sunspot, NM, at local rates. All other utilities are provided by contract between local prime suppliers and NMSU/ARC.

In general, staff at APO are shared between the SDSS and the 3.5 m telescope under the supervision of the Site Manager. The on-site staff at APO provide management, computing, engineering and observing support. Services for plumbing, carpentry, electrical work, machine shop work and snow removal are purchased from NSO. Local contractors are available for large machining jobs (at High Rolls, NM, 40 minutes from the site) and for standard building and construction work (at Cloudcroft, 30 minutes away, or Alamogordo, 60 minutes away). Additional shop work and some optics work can be obtained from the laboratories of the U.S. Army at the White Sands Missile Range, 75 minutes away. APO and NSO have an arrangement for mutual backup of computing systems administration.

The overall responsibility for staff allocation, personnel issues, human safety and site resource management rests with the APO site manager. ARC provides funds to NMSU for all site expenses, which are managed by the Site Director, who is also the NMSU Principal Investigator for the ARC/NMSU operations agreements. Purchasing, site insurance, bookkeeping and related items are provided by NMSU on a for-fee basis. ARC and NMSU re-negotiate the site agreements every three years.

In summary, the physical infrastructure and management are in place at the APO site to support the SDSS. The process of hiring operators for the survey is now underway.

Figure 4.1 Data flow through the SDSS.

4.3. Hardware

4.3.1 *The 2.5 Meter Telescope*

The design of the telescope is described in detail in Chapter 6. The main telescope construction contracts are complete. The telescope structure was installed in October 1995, and is shown in Figures 4.4, 4.5, and 4.6. The exterior wind baffle is substantially complete, and is in storage at the vendor's shop, about 40 minutes from the site. It will shortly be fit-checked,

Figure 4.2 View of the SDSS 2.5 m and ARC 3.5 m telescopes from the south.

then stored until the optics have been installed. The interior light baffles are in fabrication and are scheduled to arrive at APO by the end of 1996.

The telescope controls are still in the detailed design stage. The design is based on programmable off-the-shelf components and on the 3.5 meter telescope control system, with improvements suggested by experience with the 3.5 meter. The safety interlock system is in the detailed design stage, and will be the last item installed; in the meantime, we expect

Figure 4.3 The SDSS 2.5 m telescope from the east.

the first tests of the telescope control system to use safety check lists for the operators. The telescope control computer, a duplicate of that used on the 3.5 meter telescope, has been installed. Since the operating system and design of the two telescopes are very similar, much of the experience gained in operating the 3.5 meter telescope can be applied to the 2.5 meter. For example, drift scanning has been tested and verified on the 3.5 meter telescope, and the pointing correction algorithms used by the 3.5 meter have been installed on the 2.5 m.

Careful attention was paid at all stages to the performance of the telescope components, by

Figure 4.4 The 2.5 m telescope enclosure. The monitor telescope dome is seen to the left.

Figure 4.5 The 2.5 m telescope. This view shows the azimuth bearing and the location of the spectrograph mounting. The rolled-off enclosure can be seen to the right.

Figure 4.6 The 2.5 m telescope.

measuring the precision of each component before and after installation. As a result, the contribution of the telescope structure to the cumulative tracking error is < 0.1 arcseconds. The telescope design will allow tracking precision three times better than this, but this implementation must await tests with a working telescope.

The telescope has three large optical elements: the primary and secondary mirrors and the common corrector lens. In addition, there are correcting lenses for the instruments. The primary mirror is at the site, has been aluminized and is being prepared for installation. The secondary mirror is at the site and will be aluminized once its mounting fixtures are attached. The figuring of the common corrector is almost complete, and the element will be sent to the vendor for anti-reflection coating at the beginning of December 1996. The figuring of the spectrographic corrector is done, and this element also awaits coating. The photometric corrector is finished and coated; it doubles as the mounting structural element for the camera, whose status is discussed in more detail below.

Still to be done are the installation and commissioning of the telescope control system, the assembly of the ventilation systems for the primary and secondary mirrors, the installation of the optics and the alignment of the optics. We currently expect first light (with the photometric camera mounted) in Spring 1997.

4.3.2 *The Monitor Telescope*

The monitor telescope (MT) enclosure (see Figure 4.7) was erected in 1994 and the 0.6 meter MT installed in June 1995. The CCD camera and filter wheel were installed in late 1995. The design and purpose of this telescope are described in detail in Chapter .

Commissioning of the telescope is nearing completion. The telescope optics have been recoated to match the spectral response of the photometric camera quite closely (the photometric camera uses both thinned and unthinned CCDs, while the MT camera uses a thinned chip). Good progress is being made towards fully automated operation over a whole night, and the monitor telescope pipeline (which reduces the MT data at FermiLab) has been installed at APO to aid in the commissioning. The MT camera electronics will be replaced in early December 1996 to fix a problem with excess readout noise. The MT should become operational in December 1996 - January 1997, and at this point the telescope will be finally aligned and collimated. It is hoped that the MT will be ready for routine operation in early 1997 to set up the primary standard star network; when routine operation is achieved the MT will begin setting up the calibration patches.

The SDSS has installed a 10μm cloud-cover monitor at the site; this instrument has proven very useful for 3.5 meter telescope operations. We are in the process of comparing its images with determinations of the photometricity of given nights, from observations of standard stars.

4.3.3 *The Photometric/Astrometric Camera*

The camera is described in Chapter 8. The photometric array consists of 30 photometric CCDs, in six columns of five rows, each row to observe through one of the five survey filters. The 24 astrometric CCDs form two arrays at the leading and trailing edges of the camera, and include two CCDs to measure the focus and tracking. The photometric corrector lens is a structural element of the camera; the filters are glued to it and the six photometric and two astrometric dewars, which contain the CCDs, the readout and control electronics and the cooling system, are mounted directly to the corrector via cemented threaded inserts and attachment bars. The camera electronics include the FOXI chips, supplied by FermiLab, which transfer the data from the CCDs to the data acquisition system.

Figure 4.7 The enclosure for the 0.6m monitor telescope.

Figure 4.8 The monitor telescope with its CCD camera.

The camera is now in an advanced stage of assembly. All electronic design is complete, and all electronic fabrication work has been completed and delivered. All of the electronic components have been thoroughly tested before and after assembly.

All optical fabrication is complete and the filters and spacers are bonded to the corrector. The corrector has been placed in its support. All mechanical fabrication is complete and the assembly tests are done. The dewars are now being assembled and tested. The shutter is fabricated and is being tested. The cable harness, mounts for the electronic and cooling systems and the cart which transports the camera between the telescope and the storage facility at APO are all assembled. All specifications on the optics figuring, the optics transmission and the mechanical stability have been tested, met and verified.

Figures 4.9, 4.10 and 4.11 show the camera at various stages of assembly.

The CCDs plus spares are all in hand, and the quantum efficiency, noise figure and cosmetic quality of all of the CCDs have been measured, tested and characterized. The CCDs have all been bonded to the mountings which allow their orientation and tilt to be aligned and adjusted.

Scientists from the Japan Promotion Group (JPG) have built a device to measure the quantum efficiency as a function of wavelength with a resolution of 10 Å , and baseline measurements have been made of all of the CCDs. These will be repeated when the camera is assembled and periodically during the operation of the camera at APO.

After full assembly and testing at Princeton, the camera will be partly dis-assembled, shipped, then re-assembled in the clean room at APO. The camera is currently scheduled to be shipped in early 1997.

4.3.4 *Spectrographs*

The two double spectrographs, described in Chapter 11, were designed, constructed, assembled, and tested at the Johns Hopkins University (JHU). and the four CCD cameras were made at Princeton. The cameras and camera electronics have been completed and shipped to JHU. The CCD electronics have been successfully integrated with the data acquisition system (DAQ) at APO. The first of the two spectrographs has arrived at APO, while the second remains at JHU for final optical testing. The spectroscopic mounting has been successfully fit-tested on the 2.5 meter telescope. Figure 4.12 and 4.13 show two views of the spectrographs in the labs at JHU. The dewars and mounting rings are anodized red and blue to correspond to the red and blue camera ports.

Throughout the construction of the spectrographs, all optical efficiencies have been maintained at better than specification, as have all measured tolerances on the gratings. The mechanical machining and assembly specifications have been met and verified. The performance of the CCD cameras is at, or exceeds, specification.

4.3.5 *Spectroscopic Fiber System*

Spectra are taken by positioning fibers in the telescope focal plane using plug plates. The plate drilling plan is finished, and all required tolerances for hole diameter, shape, tilt, cleanliness and placement have been verified as achievable by two commercial vendors as well as in the University of Washington (UW) shops. The fixtures used to bend the plates in the drilling machine and to bend the plates (in the opposite direction) on the telescope have been demonstrated to give accurate positioning of the fibers in the focal plane. Test plates corresponding to real positions of galaxies have been made, and simulations done

Figure 4.9 The photometric corrector with the mounting bars for the CCD dewars. Note the anti-reflection coating strips; the coatings are different for each row of filters.

Figure 4.10 Closeup of the mounting sockets for the photometric and astrometric dewars.

Figure 4.11 The photometric corrector with the filters glued in place. The "crenellated" rows of filters at either end of the photometric array are the r' filters for the astrometric CCDs. The photometric filters from left to right are g', z', u', i', and r'.

Figure 4.12 One of the two double spectrographs in the lab at JHU. The red and blue sides of the spectrograph are color-coded.

to determine that the fibers as installed in their cartridges have adequate reach. Tests and investigations are underway to ascertain the completeness level of the plugging and the effects of tangling.

The fibers and cartridges (which restrict the flexure to an amount which will not damage the fiber) have been ordered. The first set arrived in October 1996 and the sets will continue to arrive at a rate of about one per month for the next year or so. The bid followed

Figure 4.13 Blue side of the SDSS spectrograph with the CCD camera.

extensive testing to be sure that multiple use would not degrade the fibers and that the tips, polishing, beam divergence and transmission were up to specification. The fibers are delivered individually polished and mounted in packages of 20 attached to slits, with tips and cladding on the plug plate end. The fibers are mounted to slit plates at UW, affixed to slit heads and threaded into position on the cartridges.

4.3.6 *Data Acquisition Hardware*

The computer hardware and software required to acquire the data and write it to tape have been installed at APO and were successfully integrated with the CCD camera electronics (including verification of simultaneous readout of five photometric CCDs) in October 1996. The equipment includes the VME microcomputers, redundant DLT 200 drives, scrolling monitors to show the real-time output of each photometric CCD, enough disk space to hold 40 minutes' worth of photometric data (about 12 GB) for quality control and monitoring purposes, and cabling, power and racks.

The software includes rudimentary command control of the instruments, a quick-look toolkit for simple manipulation of the data, and programs to output the data stream to disk, tape etc. The cabling and racking of the DA system were completed in October 1996. It is likely that most of the testing of the camera will be done at the site, using the DA system, before the camera is mounted on the telescope.

The only uncompleted task at this point is the final version of the operations software, which allows the operator to issue commands to start and stop the instruments, etc. Operations and DA software installation is complete for the MT.

4.4. Software

4.4.1 *Organization*

The software which reduces and calibrates the SDSS data, carries out target selection and archives the data is written in modular form as a series of pipelines which are integrated and run together under the SHIVA operating system (Chapter). The software has been developed and tested by scientists at the SDSS institutions and is organized under CVS (concurrent versioning system).

The development of each software component is organized in four stages: prototype, Level 0, Level 1 and Level 2. Level 1 is the code containing the entire functionality and extensively tested on detailed simulations which is taken into commissioning and the test year. During the test year, the algorithms are refined and optimized to handle real data, and are then "frozen" as Level 2 for the start of the survey proper.

An intensive effort has been made to produce simulations of the survey to aid integration and testing of the algorithms before real data from the 2.5 meter telescope becomes available. The software also includes a detailed data model, to keep track of the changing outputs of the pipelines.

The pipelines are typically developed by one group and tested by another. A large computer has been acquired at FermiLab to run the survey (this is known as the production machine) and the pipelines have been successfully ported to this machine. At this point (November 1996) the software is with a few small exceptions ready to receive data from the 2.5 meter.

4.4.2 *The Operating System*

The operating environment, SHEBA, integrates the data reduction pipelines and also contains an extensive toolkit for display, analysis and debugging. It has been operated at Level 1 for about a year.

4.4.3 *Astrometry Pipeline (Astrom)*

This pipeline calibrates the positions of stars between 14^m and 15^m with reference to the astrometric standard network. It is written and tested by two groups at the United States Naval Observatory (USNO) and is complete to Level 1.

4.4.4 *Monitor Telescope Pipeline (MTpipe)*

This pipeline reduces the data from the MT and produces photometric calibrations for the data from the photometric camera. It is written at FermiLab and Princeton and tested by the JPG. It is complete to Level 1.

4.4.5 *Photometric Pipeline (Photo)*

The largest of the pipelines, Photo reduces the data from the photometric camera and produces reduced images and object catalogs. With the exception of one remaining task (deblending overlapping images) Photo is complete to Level 1. An example of Photo's output is shown in Figure 4.14. It is written at Princeton and tested by the JPG. Astrom, MTpipe and Photo have been successfully integrated for more than a year.

4.4.6 *Target Selection (TS)*

Based on Photo's outputs, TS selects objects for spectroscopy. The code is done at Level 0 and many of the algorithms are in place at Level 1. It is written at FermiLab and tested by David Weinberg (Ohio State).

Figure 4.14 Example of automated photometric data reduction by Photo. Top: simulated CCD image of a field containing a bright star. Bottom: Image with bright star subtracted.

4.4.7 *Spectroscopic Pipeline (Spectro)*

Spectro reduces the spectroscopic data and calculates calibrated spectra, with redshifts and other parameters. Spectro is at Level 0, some of the Level 1 code is in place, and the code will shortly be ready for testing and integration. Spectro is written at Chicago and is to be tested at Princeton.

4.4.8 *Survey Strategy*

Strategy designs the plug plates, generates a drilling plan, does bookkeeping for the survey and designs and updates observing plans to maximize the rate at which the sky is covered. Some parts have been finished for about a year, others are still under development.

4.4.9 *Operational Data Base (ODB)*

The ODB contains the outputs from the pipelines in a commercial object oriented data base. It resides in a Silicon Graphics Challenge at FermiLab and maintains the database and data model, applies calibrations, runs target selection and maintains a history of calibration and target selection. The ODB is at Level 0 and has recently been successfully used to run TS.

4.4.10 *Science Data Base (SDB)*

The science data base contains the calibrated SDSS data under an object oriented data base with rapid, efficient access and query and visualization tools. It is written at JHU and is expected to reach Level 1 in early 1997. The SDB will develop into the tool for public distribution of the SDSS data.

4.5. Installation and Commissioning

The schedule for the delivery of the remaining components and their installation at APO is shown in Figure 4.15. The telescope and camera are expected to be functioning together and ready to take data by April 1997. An extensive period of commissioning will follow, with a well-defined set of performance goals to be met before the survey can begin. Staffing at the full operations level will be necessary beginning early 1997.

Figure 4.15 Overall schedule for the SDSS.

5. 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.

5.1. 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 million galaxies with $r' < 18.15$ Petrosian magnitudes will constitute the spectroscopic survey (see Chapter 3.1.2 for a full discussion); 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 5×10^7 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.

5.1.1 *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 2048×2048 CCDs we have chosen, 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 we have 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 6 and 8) 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 6, 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 2048×2048 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.

5.1.2 *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.

5.1.3 *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 π 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 $z \sim 0.2$ (see Chapter 3.1).

The geometric boundaries of the survey region need to be considered in some detail. The cone $b > 30^\circ$ 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 24.6° at transit ($\delta = -32.6^\circ$). 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

Figure 5.1 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.

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^h20^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^h6^m to about 17^h34^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 colums densities measured by Stark *et al.* (1992) in Figure 5.1. 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).

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 6 and 8 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 5.1, 5.2 and 5.3).

The meridian which passes through the center of the survey area, 12^h20^m, defines the *central meridian* of the scan. The great circle perpendicular to it passing through the survey center at $\delta = 32.8°$ is the *survey equator*. A point in the survey region is defined by a *survey latitude* η 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* λ 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 $\delta = 0$, $\alpha = 18^h20^m$ and 6^h20^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 6.2) 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^h40^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

Figure 5.2 Projection on the sky of SDSS survey area. 'N' is the North Celestial Pole. The stripe longitude λ is measured from the survey central meridian positive to the east along the great circles perpendicular to that meridian; the latitude η 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 $7^{\rm th}$ magnitude. The largest symbols represent $0^{\rm m}$ stars and the smallest stars of $5^{\rm m}$ and fainter. The Monitor Telescope patches are shown as squares and are placed 6 per $15^{\rm o}$ of λ to avoid bright stars.

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 12) to have three such growing regions,

Figure 5.3 Right ascension-declination grid superposed on the survey area. The projection is the same as in Figure 5.2.

but it is probably necessary for the sake of the homogeneity of the survey.

5.1.4 *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, which we discussed in more detail in Chapter 3.9, 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 $\mu_{24} = 21.5$ reliably (compact ellipticals like M32 have $\mu_{24} \leq 21.0$).

We have done fairly extensive simulations of the survey imaging data (*cf.* Chapters 3.9 and 13) to investigate questions like these; Galaxy 4 in the first mosaic (Figure 3.9.1) is 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 20σ level, and a magnitude fainter and 20% smaller, at $z = 0.288$, still at 10σ. 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 4.5σ away from a star. Indeed, at the survey limit the difference between a de Vaucouleurs law classification and

an exponential is 3.2σ, 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.

At the other extreme, that of low surface brightness galaxies, we should be able to do very well indeed, as we discuss more fully in Chapter 3.9; 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; *cf.* Section 3.1.4); 4) 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). Many of these issues have been discussed at length in Chapter 3.

5.2. 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 (*cf.* Chapter 3.1). 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 (Chapter 3.1.2).

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 π 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.

5.2.1 *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 τ needed to reach some fiducial signal-to-noise ratio varies as $\tau \sim s^{-1}$ (if the sky background is negligible), but the number of galaxies varies as $N \sim s^{-1.25}$. Hence the number of galaxies that can be measured in a fixed length of time varies as $N \sim s^{-0.25}$. 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 $18^{\rm th}$ in r' the flux fraction grows something like $s^{-1/2}$, so a $\tau \sim s^{-1}$ 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. As we described in Chapter 3.1.2, 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 what 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).

5.2.2 *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 (*cf.* Chapter 3.1.2 for a complete discussion). 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.

5.2.3 *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 λ5180, Ca λ5270, and the Na I doublet λ5890. In emission, Hα 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 sec^{-1}, and we will take 125 km sec^{-1} 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 11) 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 (Fig. 3.1.4), so the upper wavelength limit should be at least 8250Å for Hα. 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 11, 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 Hα 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. The simulated galaxy and quasar spectra in Chapter 3 and the sensitivity estimates in Chapter 11 indicate that we can in fact reach this goal with exposures of somewhat less than an hour with our telescope and spectrographs.

5.2.4 *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 34.5° ($\pm 30^m$ 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 λ_c is the central wavelength in the sense that the displacement at the ends is symmetric with respect to it; λ_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 λ3727, having only lately entered the range, is in a place where the losses are about 24% of those at the ends. Hα 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 Hα 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 arcseconds, 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 arcseconds (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.

5.2.5 *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, as discussed in Appendix , 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 11).

5.3. 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:

1. Details of the performance of the camera system, which include tracking algorithms, the scale-determining algorithms discussed in Chapters and 11, ghost images, the

effects of stars bright enough to saturate the CCDs, the focus servo, and others which will arise as engineering is completed.

2. 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). 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.

3. All the detailed algorithms which make up the pipelines with which we are reducing the data need to be tested with real data. We are carrying out extensive simulations of the SDSS data with the express purpose of testing these algorithms (*cf.* Chapter 13), but there are sure to be any number of senses in which these simulations are not realistic enough to test the pipelines thoroughly.

4. 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.

5. 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.

6. 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.

5.4. 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 12).

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. The disposition of this time and the time in the developing gaps between the South and the North is discussed in Chapter and Appendix .

5.5. 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^h40^m to about 5^h0^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 $\alpha = 0^h,\ \delta = +15^\circ$ and $\alpha = 0^h,\ \delta = -10^\circ$. 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 (cf., Chapter 3.1).

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 might go two magnitudes deeper yet.

The principal scientific drivers for a deep, multi-color imaging survey have been discussed in Chapters 3.1, 3.6, and 3.9: 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. As discussed in Chapter 3.9, 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, as discussed in Chapter 3.1.2. 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 $\sim 1/2$ 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, *A. J.*, **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, *Ap. J. Suppl.*, **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

6. The 2.5-meter Telescope

6.1. Optical Design

6.1.1 *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 Scientific Imaging Technology (SITe), the independent entity created from 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 characteristics. 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 Irenee 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.

6.1.2 *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 $1/n^{\text{th}}$ 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 6.1; the rays in that figure are for an object at the extreme edge of the 3° 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 8), 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 100μ RMS, which at f/5 corresponds to an image degradation of about 20μ. 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 11.6.

Figure 6.1 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 6.1a. The Optical Design for the DSS 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 6.1b. The Optical Design for the DSS 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 6.1. Optical Design of the SDSS Telescope

In these 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 6.1a, and the spectroscopic design in Table 6.1b. 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, which we are in the final stages of figuring and polishing at the University of Arizona's Optical Sciences Center, we are using a null lens which will soon be verified using a technique developed by J. Burge (Steward Observatory Mirror Lab) which uses a computer generated hologram. Interferometric testing of the secondary, presently being aspherized at the Mirror Lab, 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, and will soon be 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 has recently been completed (by Loomis Custom Optics) with excellent results. Custom Optics has also finished drilling the camera mounting holes (see Chapter 8) in the flat side of this corrector, and the corrector is now undergoing application of a custom striped antireflection coating which matches the camera photometric bands. The second spectrographic corrector has been aspherized under contract with Tinsley Laboratories.

6.1.3 *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 6.2a. 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 ϵ 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 6.2) 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 6.2b. δ 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 4μ) 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 describe briefly here and summarize the results. The optical layout of the camera focal plane is sketched in Figure 6.2, which shows the locations of the 30 2048x2048 "photometric" CCDs, the 22 2048x400 "astrometric" chips, and the two 2048x400 focus sensors. The filters u′ (λ_{eff} = 3540Å), g′ (λ_{eff} = 4760Å), r′ (λ_{eff} = 6280Å), i′ (λ_{eff} = 7690Å), and z′ (λ_{eff} = 9250Å), with which each field is identified, are fully described

in Chapter 8. (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.

The final images (five per CCD) are composed of the five individual monochromatic images and, because TDI integrates along a column, 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 defocussed 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 6.3. Here each row of images is the model of TDI output for the array as labeled in Figure 6.2; 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 5). 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 6.2c, where each row lists the properties of one field (detector). The identifying numbers are as in Figure 6.2. 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 6.2b, 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

Figure 6.2 The field of the SDSS Camera. The photometric CCDs are arranged in 6 identical columns of 5 2048 × 2048 chips, each with one of the five primary filter bands of the survey. Leading and trailing these columns are arrays of 12 2048 × 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.

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 8 for a fuller discussion of the focussing 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.

6.1.4 *The Performance of the Spectrographic Design*

In Table 6.3 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 11). 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 Ds are longitudinal deviations from best focus at that wavelength, and the ϵs 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 11.

Table 6.2a. Parameters for best focus at each color at each radius

Angle arcmin	λ Å	f_b mm	h mm	ϵ 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 6.2b. Focal surface, lam=4760A, scale=3.61519 mm/arcmin

Angle arcmin	δ mm	ht mm	lindev 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 6.2c. 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	49.2 × 49.2	u'	1.32	4.3	−3.6311	3	0.2	17	40
2	136.5	0.0	49.2 × 49.2	u'	1.32	4.3	−3.6326	10	−0.5	19	29
3	227.5	0.0	49.2 × 49.2	u'	1.32	4.3	−3.6329	30	−1.5	30	41
4	45.5	65.0	49.2 × 49.2	i'	1.40	4.3	−3.6350	4	0.2	18	21
5	136.5	65.0	49.2 × 49.2	i'	1.40	4.3	−3.6353	13	−0.4	20	24
6	227.5	65.0	49.2 × 49.2	i'	1.40	4.3	−3.6339	30	−1.3	25	35
7	45.5	−65.0	49.2 × 49.2	z'	1.40	4.3	−3.6350	4	0.1	18	21
8	136.5	−65.0	49.2 × 49.2	z'	1.40	4.3	−3.6351	13	−0.5	20	26
9	227.5	−65.0	49.2 × 49.2	z'	1.50	4.3	−3.6348	26	−1.1	28	39
10	45.5	130.0	49.2 × 49.2	r'	1.30	4.3	−3.6342	12	−0.5	18	19
11	136.5	130.0	49.2 × 49.2	r'	1.30	4.3	−3.6338	25	−1.0	20	23
12	227.5	130.0	49.2 × 49.2	r'	1.50	4.3	−3.6348	37	−1.4	27	32
13	45.5	−130.0	49.2 × 49.2	g'	1.30	4.3	−3.6347	11	−0.4	19	21
14	136.5	−130.0	49.2 × 49.2	g'	1.30	4.3	−3.6344	24	−0.9	21	23
15	227.5	−130.0	49.2 × 49.2	g'	1.40	4.3	−3.6345	40	−1.6	27	32
16	45.5	204.5	49.2 × 9.6	r'	1.50	4.3	−3.6351	2	−0.1	19	21
17	136.5	204.5	49.2 × 9.6	r'	1.50	4.3	−3.6347	10	−0.8	22	26
18	227.5	204.5	49.2 × 9.6	r'	1.80	4.3	−3.6358	20	−1.4	30	40
19	0.0	220.0	49.2 × 9.6	r'	1.50	4.3	−3.6350	1	−0.0	20	21
20	91.0	220.0	49.2 × 9.6	r'	1.50	4.3	−3.6349	6	−0.4	21	24
21	182.0	220.0	49.2 × 9.6	r'	1.60	4.3	−3.6351	17	−1.2	26	33
22	0.0	235.5	49.2 × 9.6	r'	1.50	4.3	−3.6350	1	−0.2	22	23

As will be discussed further in the strategy chapter (Chapter 5), 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.

Figure 6.3 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 6.2). The images in a set are spaced 3 arcseconds apart.

Figure 6.4 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.

Table 6.3a. Average focus for all colors at each radius (fibers)

Angle arcmin	λ Å	f_b mm	h/dh mm	D mm	ϵ 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 6.3b. Average focal surface: scale=3.62730 mean exit pupil at -5174 mm

Angle arcmin	Focus mm	ht mm	hlindev mm	yp rad	dyp 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

Greyscale PSFs for the field angles in the table are shown in Figure 6.4, 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.

6.2. Mechanical Design and Performance

6.2.1 *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.

6.2.2 *Structure*

The telescope optics support structure (OSS) consists of the primary support structure (PSS) and the secondary truss (Figure 6.5). 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.

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.

Figure 6.5 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 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.

6.2.3 *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 arc seconds 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.

6.2.4 *Optics: Support and Thermal control*

The mirror is a borosilicate honeycomb (Figure 6.6) 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 is under contract for the generation and figuring of the mirror. As of December 1996, the mirror is in the final stages of figuring, with RMS surface errors of less than 250 nm.

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 6.6 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 $2.8 \times 10^{-6}/°C$. 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 piezoelectric 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 thoughout 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.

6.2.5 *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.

6.2.6 *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, now being 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 to prevent direct illumination of the primary mirror by sources more than 27° from the boresight (Figure 6.1).

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 2×10^{-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.

6.2.7 *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.

6.2.8 *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 6.7 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 6.7 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.

6.3. 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 6.8 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 6.8). 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 6.9 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 6.9 and partially completed in Figure 6.8. 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., *Proc. of the S. P. I. E.*, **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., *Proc. of the S. P. I. E.*, **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., *Proc. of the S. P. I. E.*, **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., *Proc. of the S. P. I. E.*, , 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., *Proc. of the S. P. I. E.*, **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., *Proc. of the S. P. I. E.*, **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., *Proc. of the S. P. I. E.*, **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., *Proc. of the S. P. I. E.*, **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., *Proc. of the S. P. I. E.*, **1236**, 834.

Ulich B.L., 1988, "Overview of Acquisition, Tracking, and Pointing System Technologies", in *Acquisition, Tracking, and Pointing II*, *Proc. of the S. P. I. E.*, **887**, 22.

8. The Photometric Camera and the CCDs

8.1. The Photometric System

The photometric camera for the SDSS consists of two TDI scanning CCD arrays, one using 30 Tektronix/SITe 2048×2048 CCDs in a 5 by 6 array for five-color photometry, and the other using 24 2048×400 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 exists 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 our choice of coatings and the CCDs we will use are shown in Figure 8.1.

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 Ångstroms 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

Figure 8.1 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.

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 percent 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. We have calculated ghost images for the camera, and they will not represent a serious difficulty with 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_ν (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).

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 (*cf.* Figure 3.8.1.).

We had originally planned to devote two of the five chips in a give 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 (see Section 3.8).

Figure 8.2 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.

8.2. 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 8.2. Also shown there are the 22 smaller (2048×400) chips used to provide astrometric calibration, and the two 2048×400 focus monitors. The scanning is top-to-bottom in Figure 8.2, 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.)

The expected photometric performance of the camera is outlined in Table 8.1, which presents the assumed filter parameters, derived from the curves in Figure 8.1. The quantity q_t is the peak system quantum efficiency in the system, and $qtdl/l$ 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:

$$N_{el} = 1.96 \times 10^{11} t(qtdl/l) 10^{-0.4\mathrm{AB}_\nu}.$$

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 8.1c for stellar objects with 1 arcsecond FWHM images with a reasonably realistic seeing-dominated PSF (see Chapter 13). 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 3×10^5 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.

8.3. Mechanical Design of the Photometric Array

Since the telescope has been designed carefully to accommodate the CCDs we will use, 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 (see Figures 8.2 and 8.3) 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

SDSS Filter Set and Photometric Sensitivity

Table 8.1a. Filter Parameters

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

Table 8.1b. 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 8.1c. 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

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 8.2 shows the front view of the whole assembly as it would look with the shutters removed, through the final corrector. Figure 8.4 shows the assembled corrector plate, mounted in its support, with the filters and the dewar mounting rails.

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 (see Figures 8.3, 8.5, and 8.7). 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 8.5) 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 8.5) 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 1.3μ in the focus direction and 1.0μ 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

Figure 8.3 A disassembled photometric dewar. The CCDs are mounted on the invar optical bench (foreground).

Figure 8.4 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.

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.

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 8.8 below) with it.

8.4. Cooling the Photometric Array

The dark currents for the Tek/SITe CCDs range from 30 to about 200 pA/cm^2 at 20 °C. At -80 °C, this is reduced by about a factor of 3×10^5, which for a 24μ 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

Figure 8.5 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 8.6 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).

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 FPCs connecting the CCDs to their support circuitry, conductive losses to the force actuators on the kinematic

Figure 8.7 A central section side view of the photometric dewar. The locations of the liquid nitrogen reservoir and the electronics are shown.

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 8.7) 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 8.5, 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.

8.5. 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 $14^{\rm th}$ 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 8.2 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 with respect to performance, Chapter 5 with respect to the survey strategy and in Chapter 3 with respect to scientific issues; 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 2048×400 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 2048×400, 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 8.2; 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 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 2×10^{-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).

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 2×10^{-3} to 3×10^{-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.

8.6. 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 $1.5 \times 10^{-6}/°$C, and a one-degree temperature change induces a dimensional change over the 455 mm width of the array of 0.7μ, 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 8.2). 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 1 μ. 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.

8.7. 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 8.8 shows images at five positions on the focus sensor through focus with 150μ focus steps, as produced by the design optical system and convolved with 0.8-arcsecond Gaussian seeing.

We need to control the focus very accurately to maintain astrometric accuracy. If we want 2μ positional accuracy in the focal plane, we must control the focus to about 35μ, since the maximum angle which the central ray makes to the focal plane is about 0.055 radian. Focus errors this large contribute only 7μ 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.

8.8. 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

Figure 8.8 Focusing the camera. Images in the focus CCDs are shown at at five positions across the chip and at focal positions -300μ, -150μ, 0, +150μ, and +300μ referred to best focus. The left panel are the images as produced by the telescope, the right convolved with 0.8 arcsecond Gaussian seeing.

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 $\times$ 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 8.9). 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.

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 700μs, 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.

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 ± 1/4 LSB. The circuitry for this is illustrated in Figure 8.10, and the timing diagram in Figure 8.11.

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

Figure 8.9 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.

Figure 8.10 A schematic of the dual signal chain and A/D board.

Figure 8.11 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.

measurement with a set of serial-input 8-position analog switches. The circuit for this voltage generator and monitor is in Figure 8.12. 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.

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') 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

Figure 8.12 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.

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).

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 as part of 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 20×25 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.

8.9. The CCDs

This whole project hinged quite visibly on the availability of many (42, including spares) 2048 × 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 which has led to this proposal 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 × 1024 CCDs and tend to their electronics?) or extend the survey time-to-completion unconscionably.

Table 8.2. The SDSS CCD Complement: Requirements

The Photometric Imagers: 2048 × 2048, 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: 2048 × 2048, 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: 2048 × 2048, 24 micron pixels

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

The Astrometric Imagers: 2048 × 400, 24 micron pixels

28 unthinned frontside devices, noise < 15 electrons.

The situation improved very markedly in the intervening years. Tektronix produced large numbers of 2048×2048 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

Figure 8.13 Five CCds on an optical bench, with the gold plated heat shield in place.

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 8.13. 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 8.14. 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 percent for all our devices at -80 C at 3500 Å.

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.

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.

Figure 8.14 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.

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.

Figure 8.15 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.

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 8.15, 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 8.2. 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.

11. Spectrographs

11.1. 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 $r' \approx 18.2$ 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 5), 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 Hα to a redshift of $z = 0.2$ or more requires coverage to at least 8000 Å. Detector size suggests 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 (180μm)

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 (180μm). For galaxies near our faint limit, this diameter gives a galaxy flux comparable to the sky.

11.2. 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 2048×2048 CCDs with square 24μm 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).

11.3. 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.

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 11.1). 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 (180μm diameter) will just span three pixels ($3 \times 24\mu\text{m} = 72\mu\text{m}$) with a spectrograph demagnification of 2.5. Placing the individual fibers at 360μm intervals we might squeeze up to 340 spectra onto a single 2048×2048 CCD. In practice, packaging constraints limit us to 320 spectra per spectrograph and the fiber ends are spaced at 390μm 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

Figure 11.1 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.

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

11.4. Fiber Feed System

Fibers

We are using silica UV-enhanced step-index fiber with a core diameter of 180μm 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 11.3. 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.

At the slit, the fibers are separated by 390μm center to center. The fiber outside diameter is only 220μm, 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

Figure 11.3 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.

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 11.2).

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,

Figure 11.4 Detail showing v-groove blocks mounted to the slit.

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.

At the spectrograph slit, the twenty fibers of the harness terminate in one custom-made stainless steel v-groove block (Figure 11.4). 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 have received and tested prototype harnesses from a number of vendors. They have achieved adequate throughput and have survived lifetime tests. We are now preparing the invitation to bid for the final contract, which will be awarded to one of these vendors.

We will install the fiber optic harnesses ourselves. The harnesses are attached at two points:

Figure 11.5 Fiber tester

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 excess from contaminating 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 11.5. 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 are making two fiber testers. The manufacturer will use one to measure every fiber and reject those found to have inadequate throughput. We will use the other to verify the manufacturer's measurements for some fraction of the fibers.

11.5. Optical Design

The spectrograph optical layout is shown in Figure 11.6. 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 2048×2048 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, 7×17 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.

Figure 11.6 Spectrograph optical layout. The beamsplitter (C) transmits longward of 6000 Å.

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 do not exist, new masters have been ruled.

Camera

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 24μm 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 24μm 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 11.7. 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 × 2048 with square 24μm 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 11.9).

Optical Prescription

Table 11.1 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.

Figure 11.7 Spectrograph camera designed by Harland Epps. Diameters are in inches. The asphere is on the air surface of the FPL51Y lens.

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 11.8 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 11.9 shows the detail efficiencies of the spectrograph components.

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 11.2 shows rms spot diameters over the field of the blue CCD. The red performance is similar. The wavelengths

Table 11.1. 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	

(in Å) 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 $3900-6100$ Å 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.

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.

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 11.10; the top panel shows the blue half and the bottom the red. The same spectrum is shown flux-calibrated and fully reduced (and binned down by 2 pixels) in Figure 3.2. 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.

Figure 11.8 Spectrograph throughput.

11.6. Mechanical Design

Plug-plates

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

Figure 11.9 Spectrograph component efficiencies. The telescope was not included in Figure 11.8.

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 degradation in drilling accuracy was observed for a

Table 11.2. 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

range of slopes in the work-piece surface from 0 to 18 milliradians. A test to extend these results to 70 milliradians is in progress.

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 11.11). 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.

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

Figure 11.10 Simulated spectra. Spectra are shown for the red and blue detectors in detected electrons with the averaged sky emission from three fibers subtracted for a $g' = 19.8$ spiral galaxy at a redshift of 0.20.

Figure 11.11 Fiber Cartridges shown with spectrographs. The large circle is the outline of the instrument rotator.

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 illuminating 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 11.12 shows how the major optical components are attached to the optical bench, which is the large boxy piece in the center.

Figure 11.12 Parts of the spectrograph.

The final product is shown in Figure 11.13. The structure was welded, stress relieved, heat treated and painted.

Figure 11.13 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 11.14 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 11.14 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 11.14 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 11.6).

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 11.15 shows the concept.

We have assembled one camera (Figure 11.16). Initial tests show that its optical performance is exactly as predicted by the raytrace.

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.

Figure 11.15 Triplet element lens cell.

Figure 11.16 Spectrograph camera

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.

11.7. 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 (α, δ, 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.

11.8. 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 2×2 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.

12. 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.

12.1. 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 ...").

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 12.1); 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).

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

Figure 12.1 How the tiling problem maps onto a network flow. The left-most 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.

99.5% of targets by increasing the number of plates to 770, but this is an additional overhead of almost 20%; it represents a year of operation and roughly a million dollars.

12.2. 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 12.2. In practice, the observing strategy of interleaving photometry

Figure 12.2 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 $36 \times 36°$ area. The blank regions are holes drilled out around bright stars.

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 ($\sim 5\%$) by tiling in sub-regions, but this is still much more efficient than using a fixed tiling pattern.

12.3. 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 (see the discussion in Chapter 3.6).

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.

13. 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, *e.g.* statistical techniques like those described in section 3.1.3. 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 (see, for example. section 3.1.2).

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 will also allow us to examine the possible impact of photometry errors and uncertainty in Galactic extinction on measurements of galaxy clustering (see sections 3.1 and).

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 13.2 describes the other two tiers, in particular our techniques for creating artificial images.

13.1. 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 $380^3 \approx 55$ 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 Ω_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 $0.064L_*$. There are just over 8 million galaxy particles in the simulation volume.

The logic behind using the CDM+Λ 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 $0.064L_*$ (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 8.1. 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.

We have already illustrated a simulated redshift survey produced by these methods, in Figures 3.1.2 and 3.1.3. 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 earllier version of the target selection criteria; cf., the discussion in Chapter 3.1). 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 13.1 repeats the 6 degree by 130 degree redshift-space slice shown in Figure 3.1.3. 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 in this sort of picture.

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 (see section 3.8.2). Another important element still to be added is Galactic extinction; we will use the simulations to check our procedures for computing and correcting this effect, as well as for checking our algorithms for selecting stars whose colors can be used to measure the extinction (see section).

13.2. 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 13.2.2.

i) PSF

The simulations to date have used a PSF consisting of a double-Gaussian core with an r^{-3}

Figure 13.1 Slices from a simulation of the SDSS redshift survey. The upper panel (repeated from Figure 3.1.3) 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.

power-law tail. Thus, the intensity pattern produced by a point source of total flux F_0 is

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

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 α, β, σ_1, and σ_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 section). 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) will continue to characterise 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:

$$I_{\rm spider}(r,j) = spk_{\rm amp}(r) * spk_{\rm cross}(j) ,$$

where

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

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

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

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 $spk_{\rm amp}(r)$ represents the "raw" intensity pattern of the diffraction spike, and the convolution with $spk_{\rm cross}(j)$ 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:

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

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 $qfactor \times qt\,dl\,l$ electrons, where $qfactor = 3.02 \times 10^9$ and $qt\,dl\,l = 0.047$, 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 13.2.2 shows examples of the simulated images for fields from the northern and southern surveys. The simulated field has a star density appropriate to $(l, b) = (0^o, 40^o)$. The horizontal axis is the east-west (scanning) direction. Figures 13.2a and 13.2b show portions of the g' images, 512 × 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 13.2c and 13.2d show the corresponding r' images.

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 (unflattened) image frames (see chapter , "The Data System," for details). We use realistic values for atmospheric extinction, sky brightness, atmospheric refraction, and so forth, and

Figure 13.2a Simulated g' image for the Northern survey. The area shown is 512×512 pixels, and the exposure time is 55 sec.

Figure 13.2b Simulated g' image for the deep Southern survey. The area is 512×512 pixels, and the exposure time is 1650 sec.

Figure 13.2c Simulated r' image for the northern survey.

Figure 13.2d Simulated r' image for the southern survey.

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, *Ap. J. Suppl.*, **55**, 67.

Coleman, G.D., Wu, C.-C., and Weedman, D.W. 1980, *Ap. J. Suppl.*, **43**, 393.

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

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

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

Gunn, J.E., and Stryker, L.L. 1983, *Ap. J. Suppl.*, **52**, 121.

Moffat, A.J.F. 1969, *Astr. Ap.*, **3**, 455.

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

Postman, M., and Geller, M.J. 1984, *Ap. J.*, **281**, 95.

Racine, R. 1996, *Pub. A. S. P.*, **108**, 699.

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

van der Kruit, P.C., and Searle, L., 1981, *Astr. Ap.*, **95**, 105.

14. The SDSS Data System and Data Products

14.1. Introduction and Overview

The SDSS will collect, process and distribute vastly more data than has any previous astronomical survey. 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. 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 14.1 summarises the data acquisition rates and the total data storage requirements for the Survey. The meaning of some of the terms used in Table 14.1, and the assumptions used to calculate the data rates, are as follows. The data stream from each photometric CCD is cut into 'frames', 2048×1362 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 29×29 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

Table 14.1.

	Spectroscopic	Photometric	Postage Stamps	Quartiles	Astrometric
Bits/pixel	16	16	16	3×16	16
Number of CCDs	4	30	52	30	24
Number of pixels per CCD	2192×2068	2192×1354	40×29^2		2192×1354
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	

cost 29 ¢ 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 similar to 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 29×29 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 14.1 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.

Figure 4.1 (in Chapter 4, Project Status) shows an overall schematic of the data flow in the SDSS from the sky to published astronomical papers. The software necessary to achieve this falls into two main categories, data acquisition and data processing.

14.2. Data Acquisition Systems

Figure 14.1 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.

14.2.1 *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 8.2 (in Chapter 8 about the Camera). 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.

14.2.1.1 *The Photometric Array*

Figure 14.2 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 in order to allow rearranging data before they are written to the tape, as will be discussed anon, and 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 14.1 Top-Level Survey Operations

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 $\sim$ 2 Megabyte/second, more than

Figure 14.2 Online system for 15 photometric CCDs

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 calibration 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.

14.2.1.2 *The Astrometric Array*

The DA system for the astrometric CCDs is nearly a copy of that for the photometric CCDs. However, just 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.

14.2.2 *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 2048×2048 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 11.8), 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.

14.2.3 *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. 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.

14.2.4 *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'. Astroline can can also be used off line on simulated data. This prepares the simulations for analysis by the data processing pipelines; the simulated data look to those pipelines exactly as though they are produced by the imaging DA system. The monitor telescope has its own host computer that provides the automated control and online processing.

14.2.5 *Control Software*

All of the control and analysis software will be written using the TCL (Tool Command Language) based software framework that is described in Section 14.3.2. 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.

14.3. Data Processing

14.3.1 *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 12). 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 14.3.5). 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 14.3 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, spectroscopic, and monitor telescope pipelines are rather straightforward to design and implement. 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 smoothed 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 (cf., Section 14.3.5). 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 preliminary photometric and astrometric calibrations 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.

14.3.2 *The Management and Organization of the Science Software Pipelines*

The overall SDSS management and organization structure are described in Chapter , but those relevant to the software organization are briefly summarized here. 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 operating system, 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.

9. The science data base, which enables scientific access to and analysis of the survey data, is developed at Johns Hopkins.

10. The survey strategy software is written at Chicago and Fermilab.

The simulations are described in Chapter 13 and the science data base architecture in Appendices and . The survey strategy is discussed in Chapter 5. The data processing pipelines themselves are described in the next section. The development of the software pipelines is designed to proceed in four stages, as follows.

14.3.2.1 *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.

14.3.2.2 *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 14.3 Data flow through the imaging and spectroscopic pipelines.

14.3.2.3 *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.

14.3.2.4 *Level 2*

This is the software which will carry out the data reduction for the entire survey. During the test year (Section 5.3), the algorithms will be optimized for real data. Level 2 will be "frozen" at the beginning of survey operations proper.

14.3.2.5 *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 an operating system 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 14.3. 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.

14.3.3 *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 three times 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 $<$ 5 mas due to systematic effects and $\sim$ 20 mas due to noise. This is much better than the errors expected from classical seeing theory (Section) 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 14.4 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 — 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.

Figure 14.4 Architecture of the Astrometric Pipeline

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,ν) — 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 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 5μ, 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,ν — 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 8; 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.

14.3.4 *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). 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 5 and). The data flow through MT is shown in Figure 14.5. MT automatically reduces these observations and makes the measurements which the photometric pipeline needs to calibrate the imaging data.

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 14.7) 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.

14.3.5 *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 65×65 pixels) from the photometric data stream for input to the PSP, and orders these files in an appropriate format.

Figure 14.5 Data flow through the Monitor Telescope Pipeline.

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 swap in to each processor asynchronously a new set of frames 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.

14.3.5.1 *The Serial Stamp Collection Pipeline*

Figure 14.6 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 $14^{\rm m}$ (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 $7^{\rm m}$; 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.

14.3.5.2 *The Postage Stamp Pipeline*

The data flow through the PSP is shown in Figure 14.7. 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 not to be used in calculating the bias and flat-field vectors. The postage stamps from the SSC

Figure 14.6 Flow diagram for the Serial Stamp Collection Pipeline

Figure 14.7 Flow diagram for the Postage Stamp Pipeline

are filtered for unsuitable objects (saturated stars, bright galaxies) and the point spread functions calculated by fitting a double-Gaussian model to the image, as described above in Section 14.3.3. The composite PSF for each frame is then calculated from the weighted means of each of the PSFs 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 the entire run; bias drift values from both amplifiers in each frame; the sky value for each frame; and the sky gradient for each frame.

14.3.5.3 *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 14.8):

Start

First Loop: for each frame (u', g', r', i', z'):

Correct Frames

Find 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).

Figure 14.8 Flow diagram for the Frames Pipeline

Correcting Frames

First, the frames are corrected for defects and are then flatfielded. 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.

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 ± 4 columns on either side of the bad column, and, for bright stars containing a one-column defect, recovers the flux to about 1%.

In the opposite case, data with no signal, the optimal interpolated value becomes the mean of the data, and the variance is the noise in the data. The interpolation coefficients vary smoothly between the values for the extreme case of noiseless data to that of data dominated by noise. Figure 14.9 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). The frames are then flat-fielded using the flat field found from the quartiles in the PSP.

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 at high galactic latitudes 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.

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 60σ, 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

Figure 14.9 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.

Figure 14.10 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.

and unsaturated stars. Figure 14.10 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 $12^{\rm m}$.

Figure 4.14 in Chapter 4 (Current Status) shows the result of fitting and removing a bright star. 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.

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 will be peaks.

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 14.11 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. It does so using the marginal distribution in x and y separately. It takes the bounding box produced by the object finder, adds 3 pixels in each direction, rejects bad pixels, and projects the remaining pixels on the x and y axes. The center is then found by fitting the projected PSF to these distributions.

The image is then re-interpolated by sinc-shifting 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 reinterpolation is done only for the inner 5 pixels in radius; the outer pixels keep their original values.

We measure the azimuthally averaged radial profile by measuring the median of the data in a set of concentric logarithmically spaced annuli centered on the center pixel; this is done to

Figure 14.11 Object finding. A sample simulated r' frame is shown. Pixels deemed to belong to an object are colored yellow for bright objects and red for faint objects. 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.

reject, 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 uncorrected cosmic rays and 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 1σ 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 (see the discussion in Section 3.1.2); the flux within a $3''$ aperture (the spectroscopic fiber diameter) under a fiducial seeing, and an isophotal flux, 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" = const $\times$ $\sinh^{-1}$(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 (cf. Chapter 3.9) will likely await the development of off-line pipelines to further manipulate the data, as described below in Section 14.3.5. 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 14.12. The object classifications are shown as a function of their 'real', i.e. input, type. In Figure 14.12 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.

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 PSF variations 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 (Section 3.1.4). 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)

Figure 14.12 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.

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 detailed description of the algorithms will be published.

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 1σ threshold, plus a border of 10 pixels. 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; cf., Section), 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 swapped 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 14.13 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

Figure 14.13 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.

in place not only to keep up with the photometric data flow but to reduce it many times, if need be.

14.3.5.4 *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 (cf., the discussion in Section 3.9.2). For these objects PHOTO will find the centroid, cut atlas images, and tag them as a 'large' galaxy for automatic spectroscopic observation (Section 3.1.2). 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 5.5), 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 (point source or galaxy, position to a few arcseconds accuracy, magnitude to $0.1^{\rm m} - 0.5^{\rm m}$, colors to 0.1^m and, for galaxies, color redshift) we can build up two cubes describing the survey 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, as described in Chapter 3.9. This effort may allow a real comparison between the properties of nearby and distant galaxies.

14.3.5.5 *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.

14.3.5.6 *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 20^{th} 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.

14.3.6 *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 Chapters 3, 5 and , 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 ($\sim$15%) spectrophotometry, using the calibrated photometric images.

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 14.14) 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 199; 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(λ) 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

Figure 14.14 Spectra of brightest red galaxies in Abell clusters. The data are from Collins *et al.* (1995).

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 14.16 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 (*cf.* Figure 3.2), had no confident redshift determination.

14.3.7 *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 14.17. 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.

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.

Figure 14.16 Confidence estimates for the measured redshifts. The sample consists of 91 E/S0 galaxy spectra.

14.3.8 *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

Figure 14.17 The flow diagram for the Operational Data Base

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 14.3.5 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.). The detailed target selection criteria for each type of object are described in Chapter 3.

6. Tile the above list. Tiling is described in more detail in Chapter 12 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. As pointed out in Chapter , 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 Appendices and .

14.3.9 *The Data Model*

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 14.18. The model is under development for target selection and spectroscopy.

14.3.10 *Survey Strategy*

Chapters 5 and , Survey Strategy and Operations, describe 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).

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 14.19, 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

Figure 14.18 Data model for the SSC pipeline

Figure 14.19 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.

Figure 14.20 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 14.21 Completed part of the Northern imaging survey after four years. See caption for Figure 14.20.

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 14.20 and 14.21. 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 4.2 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.

14.4. 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.

14.4.1 *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.

14.4.2 *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 4×4 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 (Figure 3.9.1) 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 numbers in Table 3.9.1 and 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 13) 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 14.2.

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 14.2.

Table 14.2. 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	

In this table, the higher levels correspond to wider anticipated distribution, perhaps with levels ≤ 0 stored only in the master archive at Fermilab.

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.

14.5. 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 14.3.

14.5.1 *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.

14.5.2 *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.

14.5.3 *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.

14.5.4 *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

14.5.5 *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

14.6. 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.

14.6.1 *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.

14.6.2 *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.

14.6.3 *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.

14.6.4 *Reviews*

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

14.6.5 *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, *M. N. R. A. S.*, **274**, 1071.

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

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

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

Heavens, A.F. 1993, *M. N. R. A. S.*, **263**, 735.

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

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, *A. J.*, **84**, 1511.

Table of Contents