

SIMULATION OF STELLAR OBJECTS IN SDSS COLOR SPACE

XIAOHUI FAN

Princeton University Observatory, Princeton, NJ 08544; fan@astro.princeton.edu

Received 1998 November 25; accepted 1999 February 3

ABSTRACT

We present a simulation of the spatial, luminosity and spectral distributions of four types of stellar objects. We simulate (1) Galactic stars, based on a Galactic structure model, a stellar population synthesis model, stellar isochrones, and stellar spectral libraries; (2) white dwarfs, based on model atmospheres, the observed luminosity function, mass distribution, and Galactic distribution of white dwarfs; (3) quasars, based on their observed luminosity function and its evolution, and models of emission and absorption spectra of quasars; and (4) compact emission-line galaxies, based on the observed distribution of their spectral properties and sizes. The results are presented in the color system of the Sloan Digital Sky Survey (SDSS), with realistic photometric error and Galactic extinction. The simulated colors of stars and quasars are compared with observations in the SDSS system and show good agreement. The stellar simulation can be used as a tool to analyze star counts and constrain models of Galactic structure, as well as to identify stars with unusual colors. The simulation can also be used to establish the quasar target selection algorithm for the SDSS.

Key words: galaxies: compact — Galaxy: structure — quasars: general — surveys — white dwarfs

1. INTRODUCTION

The Sloan Digital Sky Survey (SDSS; see, e.g., Gunn & Knapp 1993 and Gunn & Weinberg 1995) will obtain deep CCD images of about π steradians ($10,000 \text{ deg}^2$) of sky centered on the north Galactic cap in five bands, as well as spectra of approximately 10^6 galaxies, 100,000 quasars, and numerous stars selected from the photometric observations. It will detect about 10^8 stellar objects and 10^8 galaxies down to its limiting magnitude ($r' \sim 23.1$).

The five photometric bands (u' , g' , r' , i' , and z') of the SDSS cover the entire optical range from the UV atmospheric cutoff at about 3000 Å to the red silicon sensitivity cutoff at about 10000 Å (Fukugita et al. 1996, hereafter F96). The SDSS photometric system is based on the AB_{95} system defined in F96. The zero point of each passband is defined by the observations of Oke & Gunn (1983) spectrophotometric standard stars. Details of the photometric system are given in F96 and in Appendix A. The sensitivity curves of the five bands (including the contributions from the CCD, filter, telescope optics, and average atmosphere transmission; see also Appendix A) are shown in Figure 1. The spectra of a typical $z = 2.8$ quasar and an F5 V star are also plotted on Figure 1. The striking similarity of the overall shapes of these two spectra reveals the well-known difficulty of separating quasars at $z \sim 2.5\text{--}3.0$ from normal stars using only broadband colors (§ 6).

The scientific goals of the SDSS require a detailed simulation of the distribution of objects in color space, as a function of position on the sky. In this paper, we present simulations of four kinds of stellar objects that SDSS will observe: normal Galactic stars, white dwarfs, quasars, and compact emission-line galaxies (CELGs) the software will classify as point sources. The work is motivated by two important scientific projects of the SDSS:

1. *Quasar selection.*—SDSS will select quasar candidates according to both their morphology (point source) and

colors. It will generate the biggest and most homogeneous quasar sample to date, covering redshift z from 0 to > 5 . Quasars will be selected as the outliers from the dense stellar locus in color space by the target selection pipeline (Newberg & Yanny 1997). However, objects such as white dwarfs, CELGs, and very blue or red stars also have colors different from normal stars and are located outside of the stellar locus. Selecting them as quasar candidates will lower the success rate of the quasar spectroscopic identifications. The quasar selection criteria need (1) to produce a high success rate for quasars, i.e., to exclude possible nonquasar contaminants as much as possible (§ 6.3); (2) to select quasars as uniformly as possible, as a function of both sky position and redshift; and (3) to select the quasar sample to be as complete as possible. The selection criteria will be determined using a limited amount of data during the commissioning period of the survey. We therefore need detailed modeling of the distribution of stellar objects in SDSS color space, especially in the following senses: (1) modeling of the quasar colors as a function of redshift; (2) modeling of the distributions of possible contaminants of quasar candidates: white dwarfs, CELGs, and stars, both in color space and in different directions on the sky; and (3) modeling of the photometric errors of the SDSS—this is especially important for selecting faint quasars and high redshift quasars.

2. *Galactic structure.*—Half of the sources that the SDSS will detect will be stars in our Galaxy. The SDSS covers the whole northern Galactic cap, roughly speaking, with Galactic latitude $b > 30^\circ$. Compared to previous data used for star count studies, it covers a much larger part of the sky, with five well-calibrated bands (see Appendix A), and high photometric accuracy. SDSS is especially suitable for constraining Galaxy models, in particular the properties of the halo and thick disk, finding distinct groups or low surface brightness dwarf galaxies (such as the Sagittarius dwarf) in the halo, and determining the distributions of interesting

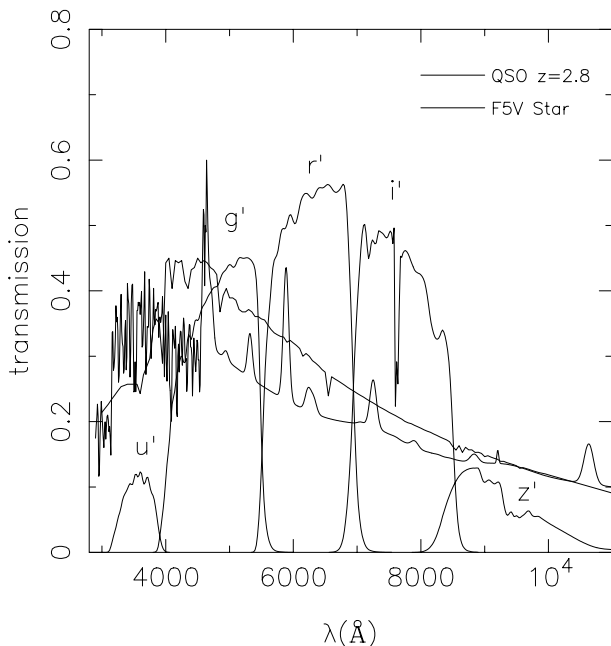


FIG. 1.—Response function of the SDSS photometric system, including the telescope efficiency, CCD quantum efficiency, filter transmission, and the atmospheric transmission at 1.2 air mass at the altitude of Apache Point Observatory (adapted from Fukugita et al. 1996). Also plotted are the spectral energy distribution of a F5 V star and a quasar at $z = 2.8$. Note the striking similarity of the two spectra, which illustrates the difficulty in separating quasars at that redshift range with stars from the broad-band colors.

classes of stars like white dwarfs and carbon stars. All these studies require detailed modeling of the colors and spatial distributions of stars.

The overall structure of the simulation presented in this paper is illustrated in Figure 2. It has five independent parts: one general part to calculate and apply Galactic extinction and random photometric errors in the SDSS system, and four special parts to simulate each of the four kind of stellar objects we consider in this paper.

A point-source model (SKY) that predicts number counts of Galactic and extragalactic sources has been developed by Cohen (1993; see also Wainscoat et al. 1992, Cohen 1994). SKY predicts number counts over a wide range of wavelengths from 1400 Å to 35 μm based on synthetic spectral library of 87 categories of sources. While the simulation presented in this paper has been fine-tuned to the SDSS observations as well as to the studies of Galactic structure and quasar selection from SDSS data, SKY will further provide valuable information on the classes of objects we do not attempt to simulation here (§ 2.4).

The paper is organized as follows: we discuss the simulations of stars, white dwarfs, quasars, and CELGs in §§ 2–5, respectively, following the structure shown in Figure 2. In each of these sections, we first discuss the spectral models of those objects, then discuss the population models (the spatial and luminosity distributions), and finally show the simulation results and compare them with test observations from the literature. In § 6, we first describe the distributions of stellar objects in the SDSS color space in general, based on the simulations presented in the previous sections; then we discuss the application of the simulation to quasar selection. We summarize the major conclusions of the paper in §

7. In the appendices, we provide more technical details of the simulation tools, including the realization of photometric errors (Appendix A), and simulation of quasar absorption systems (Appendix B).

2. SIMULATION OF GALACTIC STARS

Photometric surveys have been used as an important tool to study Galactic structure and the evolutionary history of our Galaxy (for reviews, see Bahcall 1986 and Majewski 1993). Many computer models have been developed to interpret those data. Among them are the classic star count models by Bahcall & Soneira (1981, 1984, hereafter BS), as well as those by Gilmore (1984) and Reid & Majewski (1993). These simulations are based on a Galactic model (disk + halo, sometimes with thick disk), and the observed luminosity function and color-magnitude diagram of each population. Works by Ng and collaborators based their simulated catalog on population synthesis models assuming the star formation history of our Galaxy (Ng 1994; Ng et al. 1997). Haywood, Robin, & Crézé (1997a, 1997b; see also Robin & Crézé 1986 and Haywood 1994) based their models on the synthesis of both the evolution of stellar population and the dynamical evolution of vertical structure of the disk.

Our Galactic simulation is based on the following:

1. The hybrid model stellar atmosphere libraries compiled by Lejeune, Buser, & Cuisinier (1997a, 1997b). We use them to calculate the SDSS magnitudes of stars given the effective temperature T_{eff} , metallicity $[\text{Fe}/\text{H}]$, and surface gravity $\log g$.
2. A simple stellar population synthesis model based on the Padova stellar evolutionary tracks and isochrones (Bertelli et al. 1994). We use it to generate the simulated catalog of a given stellar population.
3. A BS-type Galactic structure model (with some modifications), including a Galactic disk with scale height that grows in time and a flattened Galactic halo. We use it to generate the properties of stellar populations as a function of Galactic position.

2.1. Stellar Libraries

We use the grids of theoretical stellar atmospheres compiled by Lejeune et al. (1997a, 1997b). The grids are based on three different sets of models covering different parts of parameter space: (1) R. L. Kurucz (1995, private communication) models in the temperature range 3500–50,000 K; (2) Allard & Hauschildt (1995) M-dwarf models in the temperature range 2000–3500 K and $\log g > 3$; and (3) Fluks et al. (1994) synthetic spectra of M giants in the temperature range 2500–3500 K and $\log g < 3$. The whole library covers the following parameter range: T_{eff} : 2000–50,000 K; $\log g$: -1 to $+5.5$; $[\text{Fe}/\text{H}]$: -3.5 to $+1.0$. We then calculate the SDSS magnitudes of those grid spectra following Appendix A. We interpolate between the grids linearly in $\log T_{\text{eff}}$, $\log g$, and $[\text{Fe}/\text{H}]$. For a compilation of SDSS magnitudes based on R. L. Kurucz (1995, private communication) models, see Lenz et al. (1998).

Atmospheres of late-type stars are difficult to model. There is a noticeable discontinuity in the model stellar colors at $T_{\text{eff}} \sim 3500$ K, the transition region from Kurucz models to those of Allard & Hauschildt (see, e.g., Fig. 4c). Lejeune et al. (1997a, 1997b) further applied an empirical

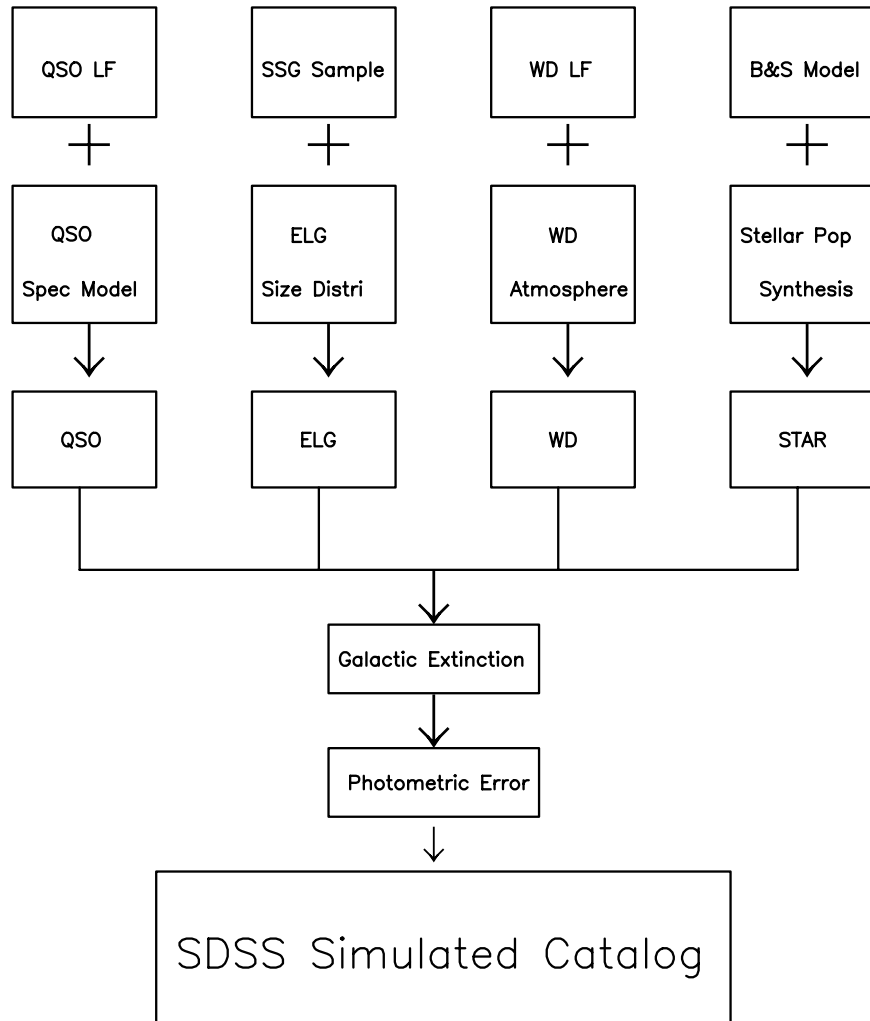


FIG. 2.—Schematic chart of the structure of the color space simulation programs

correction to the original atmosphere grids, so that the synthetic *UBVRI* colors matched the empirical color-temperature calibrations from observed cool stars. This smooths the transition between models. However, the smoothed models produce a redder $g' - r'$ color for very cool stars than the observations of Newberg et al. (1999), possibly owing to the differences between the SDSS bands and the *UBVRI* bands, where those empirical corrections were derived. We therefore retain the original Kurucz and Allard & Hauschildt grids and leave the problem of calibrating the cool star color-temperature relation to further SDSS observations.

We use the Padova library of isochrones (§ 5.2 of Leitherer et al. 1996). The isochrones are calculated from the stellar evolution models of the Padova group (for $M > 0.6 M_{\odot}$; Bertelli et al. 1994) with the aid of the Vandenberg (1985) models for $0.15 M_{\odot} < M < 0.6 M_{\odot}$. The grids cover the metallicities of $Z = 0.0004, 0.004, 0.008, 0.02, 0.05$, and 0.1 . They correspond to $[\text{Fe}/\text{H}] = -1.7$ to $+0.7$ assuming $Z_{\odot} = 0.02$. Each evolutionary sequence is followed from the zero-age main sequence (ZAMS) to the beginning of the thermally pulsing region of the AGB phase for low- and intermediate-mass stars. We interpolate between the grid points linearly in Z and $\log(\text{age})$ of the isochrones. For star

with a given age, in the mass direction, we interpolate linearly in $M_0/M_{0,\text{TO}}$, where M_0 is the ZAMS mass of the star, and $M_{0,\text{TO}}$ is the ZAMS mass of the main-sequence turnoff for that age.

2.2. Single Population Synthesis

In principle, one can simulate a stellar population from either of two approaches. BS (see also Reid & Majewski 1993) simulate stars based on observed luminosity functions (LFs) and color-magnitude diagrams (CMDs). This approach is sufficient for star count studies, especially those based on two or three colors. The SDSS observations, however, cover five optical bands, with u' color sensitive to the changes in metallicity and luminosity class (or $\log g$; see also Lenz et al. 1998). The LF does not contain this information.

Therefore, we adopt an approach based on evolutionary stellar population synthesis, similar to the models described in Haywood et al. (1997a, 1997b) and Ng et al. (1997), using the observed properties of the initial mass function (IMF), star formation rate (SFR), and the age-metallicity relation (AMR). Our simulation consists of two distinct populations: a Galactic disk with scale height growing in age, and a Galactic halo. A third population, a thick disk, can be

added easily if needed. The physical parameters characterizing each population are summarized in Table 1.

Disk.—Neither the IMF nor the SFR of the disk is well determined observationally (Majewski 1993; Scalo 1998). We require our adopted IMF and SFR to generate a theoretical LF that is consistent with the observed LF (as used by BS). We adopt the IMF and SFR of Haywood et al. (1997a, 1997b). The age of the disk is 10 Gyr, and the SFR is assumed to be constant in time for the past 10 Gyr. The IMF is a broken power law with

$$dN/dM \propto M^{-(1+x)}, \quad x = \begin{cases} 0.7 & M < 1 M_{\odot} \\ 1.5 & 1 M_{\odot} < M < 3 M_{\odot} \\ 2.0 & M > 3 M_{\odot}. \end{cases} \quad (1)$$

This form is consistent with the observed IMF of Kroupa, Tout, & Gilmore (1993). Note that for the old disk stars, the detailed slope at the high-mass end is not important.

We adopt the relation between metallicity and age of Rana (1991):

$$\begin{aligned} [\text{Fe}/\text{H}]_{\text{mean}} &= 0.68 - 11.2/(20 - t), \\ \sigma([\text{Fe}/\text{H}]) &= 0.20, \end{aligned} \quad (2)$$

where t is the age of the star in units of Gyr, and $[\text{Fe}/\text{H}]_{\text{mean}}$ is the average metallicity of that age, with a Gaussian scatter of 0.20. The disk has an average $[\text{Fe}/\text{H}]$ of -0.44 when it was formed (10 Gyr ago) and 0.12 at the present epoch. This relation is consistent with the recent observational results by Carraro, Ng, & Portinari (1998).

Halo.—The IMF of the halo is assumed to have the same form as that of the disk, and the SFR is assumed to be constant from $t_{\text{max}} = 16$ Gyr to $t_{\text{min}} = 10$ Gyr (Ng et al. 1997).

We use the halo AMR following Ng et al. (1997):

$$\frac{Z - Z_{\text{min}}}{t - t_{\text{min}}} = \frac{Z(t_{\text{max}}) - Z(t_{\text{min}})}{t_{\text{max}} - t_{\text{min}}}, \quad (3)$$

where $Z(t_{\text{max}}) = 0.0004$ and $Z(t_{\text{min}}) = 0.003$ are the metallicities at t_{max} and t_{min} , respectively. Stars in the halo will have $[\text{Fe}/\text{H}]$ ranging from -1.7 to -0.8 . The average observed $[\text{Fe}/\text{H}]$ in the halo is about -1.4 , with some halo stars having $[\text{Fe}/\text{H}]$ considerably smaller than -1.7 (Larid et al. 1988). On the other hand, the Padova evolutionary tracks are not available for stars with $[\text{Fe}/\text{H}] < -1.7$. The difference in colors between stars with $[\text{Fe}/\text{H}] = -2.0$ and $[\text{Fe}/\text{H}] = -3.0$, at a given temperature and gravity, is usually smaller than 0.05 mag. So using this AMR will not introduce large shifts in the color distribution we simulate.

2.3. Galactic Model

The model consists of a halo and a disk with an exponential scale height growing in time. The shape of the halo is a

deprojected de Vaucouleurs profile with a flattening factor q (Bahcall 1986, Table 1):

$$n_{\text{halo}} \propto (R/R_0)^{7/8} \times \exp[-10.093(R/R_0)^{1/4} + 10.093] \times [1 - 0.08669/(R/R_0)^{1/4}], \quad (4)$$

where $R_0 = 8.5$ kpc is the distance from the Sun to the Galactic center, $R = (x^2 + z^2/q^2)^{1/2}$, x is the projected distance from the star to the Galactic center in the Galactic plane, z is the vertical distance to the plane, and q is the axis ratio of the halo, set to be 0.8 (Reid & Majewski 1993).

The disk is exponential in both radial and vertical directions:

$$n_{\text{disk}} \propto \exp[-z/z_0(t)] \exp[-(x - R_0)/h] \quad (5)$$

with a radial scale length $h = 3.5$ kpc, and scale height $z_0(t)$ that grows with time t

$$z_0(t) = z_0(1 + t/\tau_0)^{2/3}, \quad (6)$$

where $\tau_0 = 0.5$ Gyr and $z_0 = 95$ pc (Rana & Basu 1992). The average scale height for all stars younger than the Sun is ~ 310 pc, similar to the scale height of the old disk in the BS model. This evolution of disk scale height is consistent with the results from solving the exact dynamical evolution equations of the disk (Rana 1991). The oldest disk stars will have a scale height of ~ 700 pc. In addition, we assume that the Sun is 15 pc above the Galactic plane ($z_{\odot} = 15$ pc; Cohen 1995 and Ng et al. 1997). The normalization of the halo and disk population is derived from matching the existing star counts toward the Galactic poles (see below). The disk is assumed to be perfectly axisymmetric without spiral structure.

For the Galactic extinction, we assume the dust is distributed in a uniform layer 200 pc thick along the Galactic plane. We use the Burstein-Heiles (1982) extinction map to get A_V^0 , the total extinction in V . The extinctions in SDSS bands A_i/A_V are listed in Table 2 (see Appendix A). The extinction in band i for a star at distance d and galactic latitude b is

$$A_i = \begin{cases} A_V^0 \frac{A_i}{A_V} \frac{d \sin b + z_{\odot}}{100 \text{ pc}}, & d \sin b + z_{\odot} < 100 \text{ pc}; \\ A_V^0 \frac{A_i}{A_V}, & d \sin b + z_{\odot} \geq 100 \text{ pc}. \end{cases} \quad (7)$$

2.4. Results of Stellar Simulations

A simulated stellar catalog is generated as follows: for a certain direction on the sky,

1. The total number of stars of each population is calculated according to the Galactic model.

TABLE 1
STELLAR POPULATION PARAMETERS

Population	Age (Gyr)	SFR	Z	Density Profile	Normalization at $b = 90^\circ$ for $M > 0.15 M_{\odot}$
Halo	16–10	Const.	0.0004–0.003	de Vaucouleurs $q = 0.8$ Exponential, $h = 3.5$ kpc	$14,000 \text{ star deg}^{-2}$
Disk	10–0.5	Const.	0.007–0.03	$z_0 = 95(1 + t/0.5 \text{ Gyr})^{2/3} \text{ pc}$	$3300 \text{ star deg}^{-2}$

2. For each of the stars, its distance, age, metallicity, and mass are generated according to the population synthesis model described in § 2.2.

3. The SDSS colors as well as the $UBVRI$ colors of each star are then calculated from the stellar libraries described in § 2.1.

4. Galactic extinction and random photometric errors are added according to the prescription in Appendix A.

The normalizations of the halo and disk populations are determined by fitting the total predicted star counts to the observations of the ESO Imaging Survey (EIS; Prandoni et al. 1998). The EIS includes CCD observations of 1.3 deg^2 area toward the south Galactic pole in B , V , and I bands. The comparison of star counts is shown in Figure 3. The $B-V$ distribution at $18 < V < 20$ (Fig. 3d) shows the bimodal distribution of halo and disk (see also Fig. 4). The simulation predicts $\sim 20\%$ fewer disk stars than the observation in this magnitude range. A much larger area is needed to fine-tune the model parameters.

We show the results of the simulation in the SDSS system toward two directions in the sky. Figure 4 shows the color-color diagrams and CMD for a 1 deg^2 region toward the

north Galactic pole (NGP). The halo and disk stars are represented by different symbols. From the figure, we notice the following:

1. In the $u'-g'$ versus $g'-r'$ diagram, halo and disk stars are separated at the blue end of the stellar locus. Halo stars are more metal poor, and thus bluer. There is a well-defined blue edge of the stellar locus. The location of this edge is related to the age and metallicity of halo stars as well as the interstellar extinction. The halo blue horizontal branch stars (BHBs) are well separated from the main-sequence locus. Their location on the diagram is consistent with those observed by Krisciunas, Margon, & Szkody (1998). They have very blue $g'-r'$ colors. Those stars have very similar colors to quasars at $2.5 < z < 3.0$ (see § 6). The big scatter toward the red end in this diagram is due to the increasing photometric errors in u' band for faint red stars.

2. In the $g'-r'$ versus $r'-i'$ diagram, the BHBs again stand out, as they are bluer than the main stellar locus. Note also that $g'-r'$ remains roughly constant for stars with $r'-i' > 0.5$, or $T_{\text{eff}} < 3500 \text{ K}$. For cool stars, absorption from molecular bands, especially from TiO, dominate the g' and r' bands. Thus $g'-r'$ color is not sensitive to the

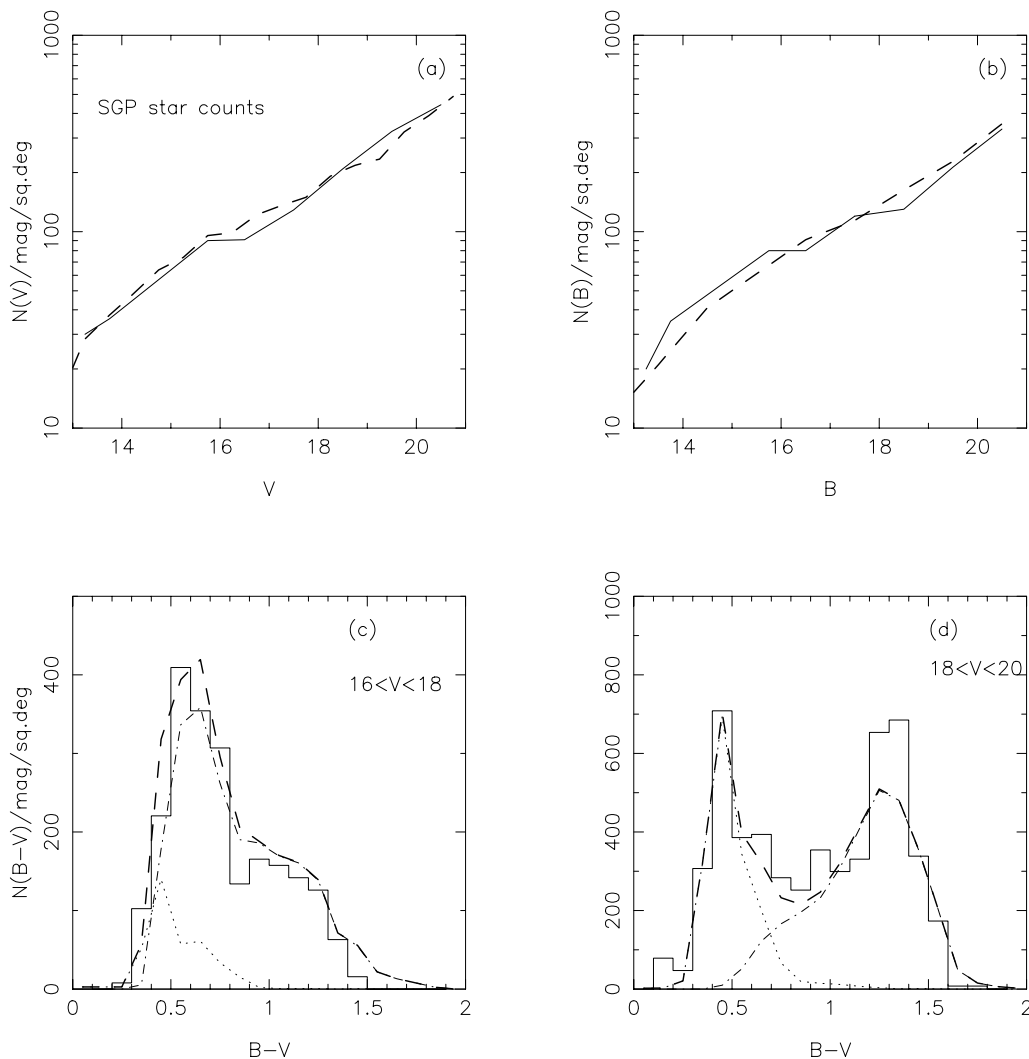


FIG. 3.—Comparison of the star counts predicted by the simulation with the observation of ESO Imaging Survey Patch B toward the south Galactic pole (Prandoni et al. 1998). The solid curves are the observed star counts. The thick dashed lines are the predicted total star counts. The thin dotted lines and dash-dotted lines are the contributions from the halo and disk populations, respectively.

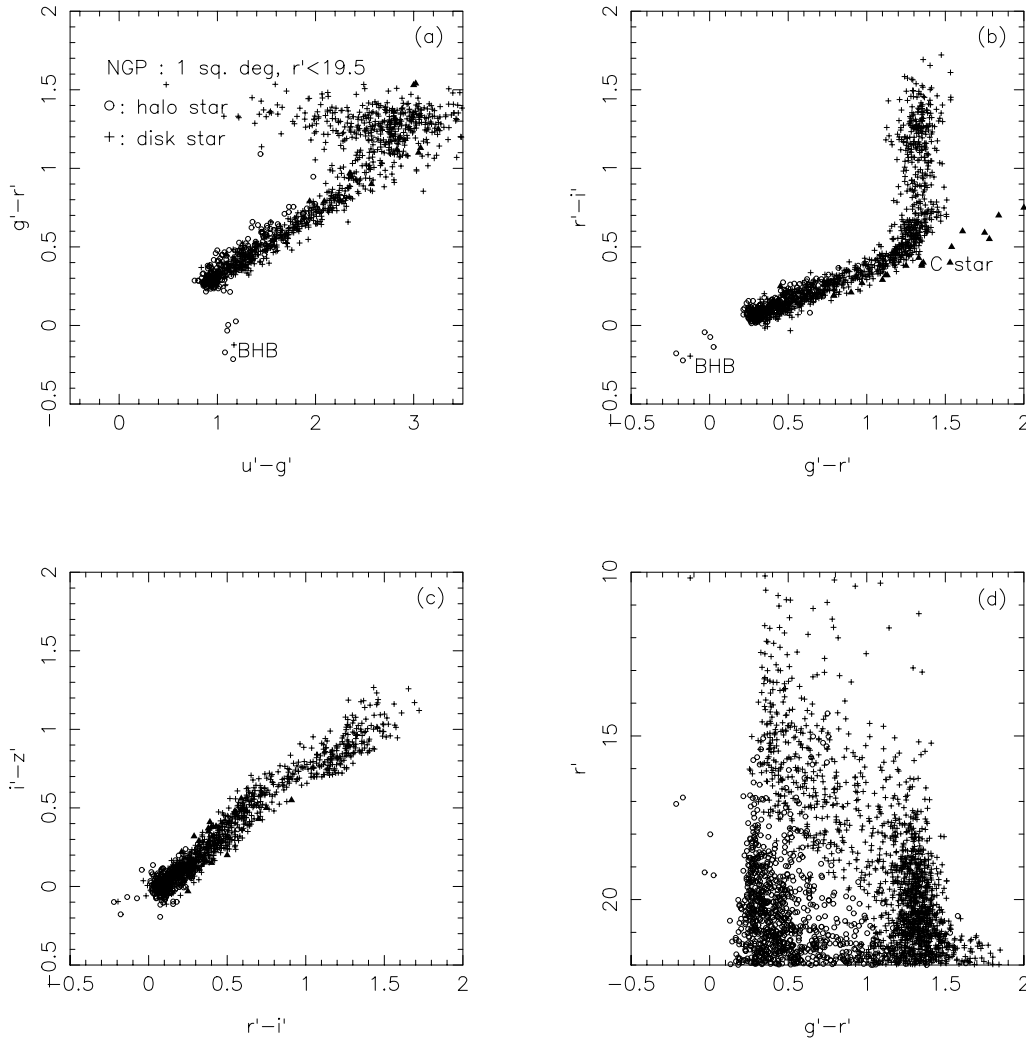


FIG. 4.—Color-color diagrams and color-magnitude diagram of a 1 deg^2 simulation toward the north Galactic pole, including the photometric errors for the SDSS observation and interstellar extinction. The solid triangles represent colors of carbon stars observed by Krisciunas et al. (1998).

change of temperature of the star. However, stars with different chemical composition (such as carbon stars), or with dusty envelopes that cause internal reddening, will show up with redder $g' - r'$ colors than normal cool stars. We also plot the high Galactic latitude carbon stars observed by Krisciunas et al. (1998) on this figure. They are easily distinguishable from normal stars. This diagram will be useful to pick out those peculiar stars.

3. The colors, especially the IR colors, are difficult to model in cool star atmospheres. We use different models for hot and cool stars (§ 2.1), which causes the slight artificial discontinuity in the $r' - i'$ versus $i' - z'$ diagram at $r' - i' \sim 0.7$ and $i' - z' \sim 0.7$.

4. The CMD (Fig. 4d) shows that at the bright end, most stars are disk stars, while at the faint end, the stellar colors show the familiar bimodal distribution: the blue peak comes from the halo stars near the main-sequence turnoff, while the red peak comes from the disk stars. Analysis of those diagrams from real SDSS observations will provide important information on Galactic structure.

Figure 5 presents the color-color diagrams and CMD toward the direction ($l = 0^\circ$, $b = 35^\circ$). This direction points toward the bulge and the Galactic plane. It has almost the

highest stellar density (for both disk and halo) and interstellar extinction in the whole region to be surveyed by the SDSS. Both the blue edge of the stellar locus and the position of the BHB are shifted to the red by ~ 0.15 mag in $u' - g'$, owing to the large extinction.

Krisciunas et al. (1998) obtained photometry of more than 1000 stars at high Galactic latitude using SDSS filters with the 40 inch (1 m) USNO telescope. In Figure 6 we compare their observations with the stellar locus from our simulations. The solid line encloses the stellar locus from the NGP simulation. The stellar colors predicted by the simulation agree well with the observations. The Krisciunas et al. (1998) data do not go as deep as the SDSS and thus do not include very red and faint stars.

The simulation results can be used to compare directly with the SDSS observations and to constrain Galactic models (as we did in Fig. 3). The star count data will be especially effective in constraining Galactic halo and old/thick disk models. The distribution of the blue edge in $u' - g'$ colors depends on the age and metallicity of the halo; the color distribution at fainter magnitudes is sensitive to the relative contribution from the halo and the disk (Fig. 3d), and the star counts in different directions (Figs. 4 and 5) will

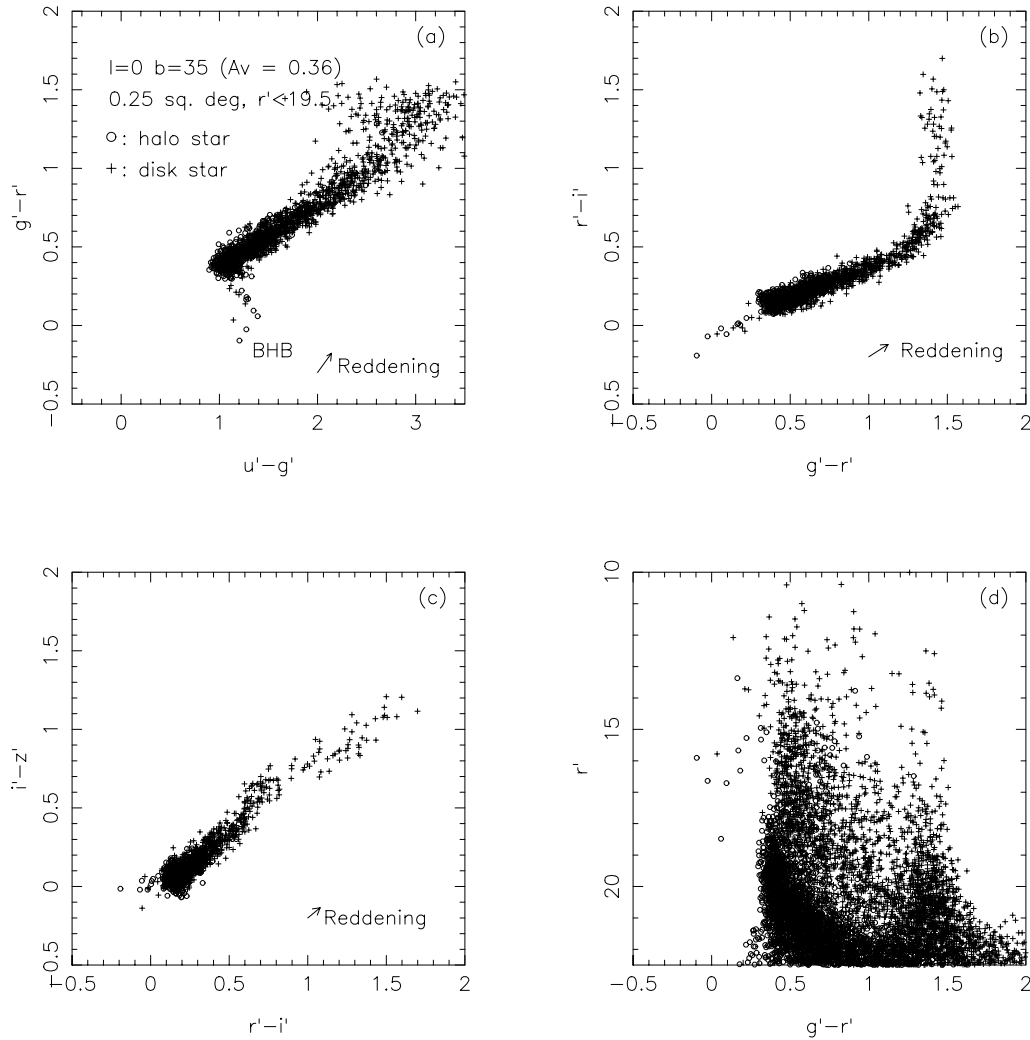


FIG. 5.—Color-color diagrams and color-magnitude diagram of a 0.25 deg^2 simulation toward $l = 0^\circ$, $b = 35^\circ$. This direction has the highest stellar density and extinction ($A_V = 0.36$) of the whole SDSS survey region.

be used to constrain the shape of the halo and the scale height of the disk. If gradients in halo/disk metallicity or age exists, this will also be apparent in comparing the simulation in different directions with observations.

We do not simulate the colors of various types of stars in their very early and late stage of evolution, or stars with peculiar spectral properties (such as T Tau stars, late stage of AGB stars, central stars of planetary nebulae, and Ae/Be stars). Nor do we simulate very low mass stars ($M < 0.15 M_\odot$) or brown dwarfs, which may present in SDSS data especially with the inclusion of the reddest z' band. Many of those objects will have peculiar SDSS colors and are interesting in their own rights. However, most of them are very rare, especially at high Galactic latitude. They are not likely to affect the studies of Galactic structure or to be important contaminants of quasar selection. For the number counts and Galactic distribution of some peculiar stars in various bands, see Cohen (1993).

3. SIMULATION OF WHITE DWARFS

Hot white dwarfs have blue colors and are one of the major contaminants in color surveys for low-redshift quasars, especially at the bright end (see, e.g., Schmidt &

Green 1983). But by using more than two colors, one can separate the majority of very hot white dwarfs from low-redshift quasars (Koo, Kron, & Cudworth 1986).

The spatial distribution, luminosity function, and mass distribution of white dwarfs, especially the bright, hot white dwarfs, have been widely studied (Weidemann 1995, and references therein). We simulate the colors of white dwarfs in the SDSS system, based on those statistics, in order to determine the optimal way to minimize the white dwarf contamination in quasar selection.

White dwarf models.—The majority of white dwarfs, DA white dwarfs, have strong hydrogen lines; their colors can be modeled by a pure hydrogen atmosphere. Most of non-DA white dwarfs are DB or DO white dwarfs with He line spectra, and their colors can be modeled by a pure helium atmosphere. There are very few white dwarfs belonging to other spectral types. In this paper, we regard all non-DA white dwarfs as DB/DO white dwarfs with helium atmospheres.

Bergeron, Wesemael, & Beauchamp (1995) have calculated H and He white dwarf atmosphere models. Combining these with a detailed evolutionary cooling sequence (Wood 1995), they obtain for each model the absolute magnitude, mass and age. P. Bergeron kindly calculated the

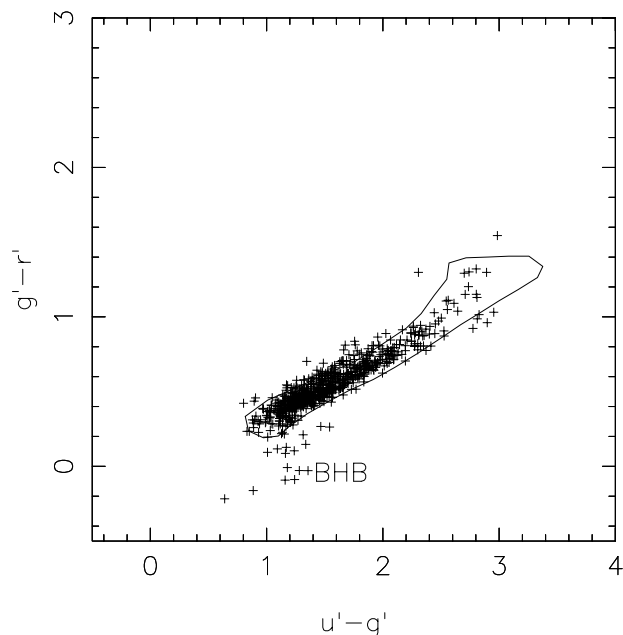


FIG. 6.—Comparison of the observations (*plus signs*) of stellar colors by Krisciunas et al. (1998) with our simulation. The solid line is the boundary of stellar locus estimated from the simulation in Fig. 4.

SDSS colors of those models, which we use here as the basis of our simulation.

Spatial distribution.—We consider only disk white dwarfs, as the density of halo white dwarfs is about 2 orders of magnitude smaller (Liebert, Dahn, & Monet 1988) and they are much farther away. The disk white dwarfs have a spatial distribution similar to that of old disk stars, with a scale height of about 275 pc (Boyle 1989). Most of the white dwarfs bright enough to be observed by the SDSS are very nearby, so the disk can be regarded as plane parallel; the density of white dwarfs thus does not depend on the Galactic longitude.

Luminosity function and mass distribution.—Liebert et al. (1988) have determined the local luminosity function of DA and non-DA white dwarfs (see also Oswalt et al. 1996). The LF is determined down to $M_{\text{bol}} = 15.75$. White dwarfs with $M_{\text{bol}} > 14$ have $T_{\text{eff}} < 6000$ K, with colors similar to normal stars. They are not important for quasar selection. Bergeron et al. (1992) have determined the mass distribution of hot white dwarfs, $M = 0.56 \pm 0.14 M_{\odot}$. We use those results in our simulation, assuming that the mass distribution is Gaussian and is independent of the luminosity.

To generate the simulated white dwarf catalog for a specific direction on the sky, we (1) calculate the total number of DA and non-DA white dwarfs based on their local LF and spatial distribution; (2) for each white dwarf, randomly choose its luminosity and mass from the LF and mass distribution; (3) knowing the mass and luminosity, determine the SDSS magnitudes from Bergeron's atmosphere models; and (4) finally, add the Galactic extinction and photometric errors as was done for stars (§ 2.2).

Figure 7 shows the number-magnitude relation of white dwarfs toward $b = 60^\circ$ in g' . In the figure, we also plot the number-magnitude relation for quasars. For $g' < 17$, the density of white dwarfs is higher than quasars at $b = 60^\circ$. White dwarfs are major contaminants of the bright quasars. In Figure 8, we plot the white dwarf density as a function of

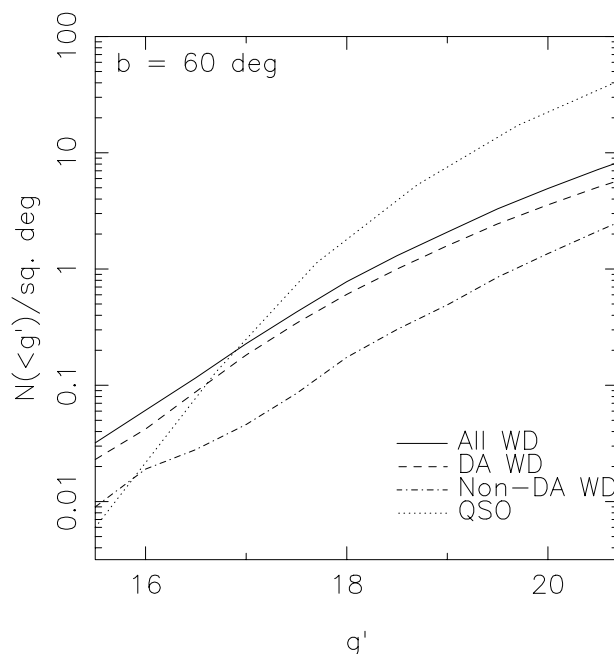


FIG. 7.—Number-magnitude relation for DA and non-DA white dwarfs at $b = 60^\circ$. We also compare it with that of quasars. White dwarfs outnumber quasars for $g' < 17$.

Galactic latitude (not including the effect of Galactic extinction). The number density of white dwarfs is much higher toward the plane.

Figure 9 plots the color-color diagrams of a simulation of white dwarfs (*filled circles*) for an area of 20 deg^2 centered at $b = 60^\circ$. Also shown are the synthetic SDSS colors (*open circles*) from the spectrophotometric atlas of white dwarfs of Greenstein & Liebert (1990), calculated by Lenz et al. (1998). The atlas consists of white dwarfs of all spectral types and temperatures. Their relative density does not represent reality. It is evident that the model colors of white dwarfs agree with the synthetic colors of real observations

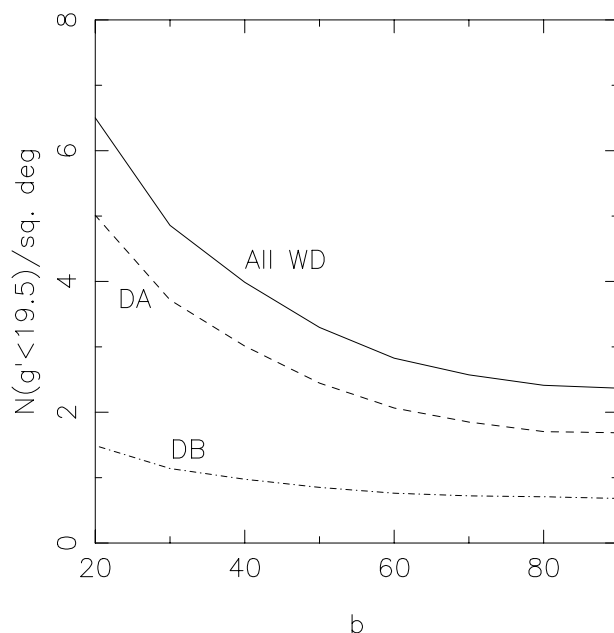


FIG. 8.—Number density of white dwarfs with $g' < 19.5$ as a function of Galactic latitude.

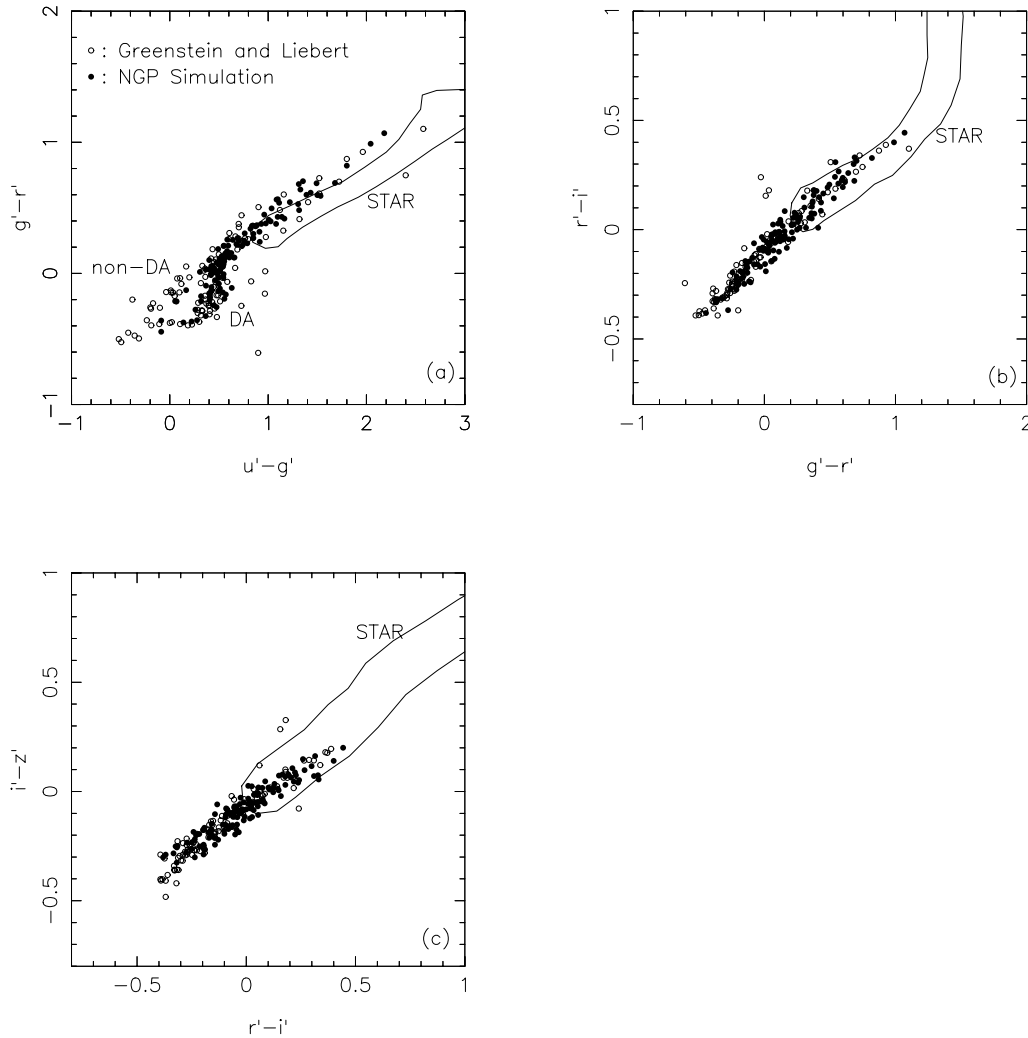


FIG. 9.—White dwarf distribution in the color-color diagrams in the SDSS system. Filled circles are from a simulation of 20 deg² toward $b = 60^\circ$, with $g' < 22.5$; open circles are the synthetic colors of white dwarfs from the spectrophotometric atlas of Greenstein & Liebert (1989), adapted from Lenz et al. (1998). Regions labeled as STAR are the approximate boundaries of the stellar locus (see Figs. 4 and 18).

quite well. For hotter white dwarfs, DAs and non-DAs follow slightly different cooling tracks on the color-color diagram, especially in the $u' - g'$ versus $g' - r'$ plot, since DA white dwarfs have a Balmer jump (although weaker than normal stars due to the very high gravity). When the white dwarf gets cooler, the colors of DA and non-DA white dwarfs become more similar to each other and to normal stars.

We will examine the distribution of white dwarfs in color space and how to separate them from quasars in more detail in § 6.

4. SIMULATION OF QUASARS

Simulated catalogs of quasars in the SDSS system will be used to determine the color selection criteria of quasars (§ 6). The quasar selection function, critical to the determination of the quasar LF, also requires detailed modeling of the distribution of quasars in color space (see, e.g., Warren, Hewett, & Osmer 1994). We generate the simulated quasar catalog based on the colors calculated from a quasar spectral model (§ 4.1) and the observed evolution of the quasar LF (§ 4.2).

4.1. Quasar Spectral Model

A quasar spectrum consists of three components: (1) continuum, which for a limited range in optical and UV, can be approximated by a power law, $f_\nu \propto \nu^{-\alpha}$; (2) emission lines; and (3) line and continuum absorptions.

In the simulation, we assume that the power-law index α has a Gaussian distribution $\alpha = 0.5 \pm 0.3$ (Francis 1996). The distribution of the indices may in fact not be Gaussian and may also be a function of redshift (or wavelength). For example, Sargent, Steidel, & Boksenberg (1989) determined from a sample of high-redshift quasars $\alpha = 0.78 \pm 0.20$. A difference of 0.4 in the power-law index will result in a difference in broadband colors of ~ 0.1 mag (Fan & Chen 1994). Small changes in the power-law indices will therefore not significantly change quasar colors.

The emission-line flux ratios of 16 major lines in quasar spectra are taken from the statistics of the sample of Wilkes (1986). This sample consists of radio-selected quasars and therefore does not have an a priori bias in the optical emission line statistics. We also include 20 weaker lines and Fe II features (those not measured in Wilkes 1986), using the line ratios from Francis et al. (1991) based on the large

bright quasar survey (LBQS) sample (Hewett et al. 1991, and references therein). The equivalent width (EW) of Ly α is assumed to follow a Gaussian distribution of 65 ± 34 Å (Wilkes 1986). The line ratios are assumed to be the same for all quasars, with only the strength of Ly α varying. For broad lines in quasar spectra, we assume the FWHM is 5000 km s^{-1} with a Gaussian shape. For narrow forbidden lines, FWHM = 1000 km s^{-1} . The width of Fe II features are taken from Francis et al. (1991). Different assumptions on the width of the lines have little effect on the broadband colors.

High-redshift quasars have complicated absorption systems. Those systems are the major factor for the color evolution of quasars with redshift. Various people have simulated the absorption spectrum and its effects on broadband colors at high redshift (e.g., Giallongo & Trevese 1990; Møller & Jakobsen 1990; Fan & Chen 1993; Warren et al. 1994). We follow the prescriptions by Fan & Chen (1993) (similar to those used by Warren et al. 1994) to simulate the absorption spectrum caused by intervening H I absorbers (Ly α forest, damped Ly α system, and Lyman limit systems, LLSs). The metal absorption lines have very small equivalent width and do not affect colors significantly. Details of the absorption simulation and parameters used are summarized in Appendix B.

In Figure 1, we give an example of a simulated quasar spectrum at $z = 2.8$, overplotted on the filter response curves of the SDSS system. The spectrum has a resolution of 10 Å. Note that at this redshift, Ly α emission is in the g' band, and the Ly α forest lines absorb a large fraction of the light in the u' band.

4.2. Quasar Evolution

The quasar LF has very strong redshift evolution. Several analytical models for the quasar LF have been proposed (see, e.g., Schmidt & Green 1983; Boyle, Shanks, & Peterson 1988; Warren et al. 1994; Pei 1995). Pei (1995) fitted the evolution of quasars over the range $0 < z < 4.5$ with a pure luminosity evolution model. He used the empirically determined quasar luminosity function from the samples of Hartwick & Schade (1990) ($z < 2.2$) and Warren et al. (1994) ($2.2 < z < 4.5$). The shape of the luminosity function is described as a double power law or an exponential $L^{1/4}$ law, while the characteristic luminosity of the quasar evolves with redshift. For the double power-law model,

$$\Phi(M, z) = \frac{\Phi^*}{10^{0.4[M - M^*(z)](\beta_l + 1)} + 10^{0.4[M - M^*(z)](\beta_h + 1)}}, \quad (8)$$

where β_l and β_h are the indices of the power laws and $M^*(z)$ describes the luminosity evolution:

$$M^*(z) = M_0^* + 2.5(1 - \alpha) \log(1 + z) + 1.086 \frac{(z - z^*)^2}{2\sigma_z^2}, \quad (9)$$

where α is the average power-law index of the quasar continuum. In this form of luminosity evolution, the characteristic luminosity evolves as a power law of $\sim(1 + z)^3$ at low redshift and reaches its maximum at $\sim z^*$ and then decays as a Gaussian toward higher redshift. The absolute magnitude is evaluated in M_B (including K -correction), and Φ is in units of $\text{Mpc}^{-3} \text{ mag}^{-1}$. We use $(h, q_0, \alpha) = (0.5, 0.5, 0.5)$ throughout the paper. For these parameters, the best-fit luminosity function is given by $\beta_l = -1.64$, $\beta_h = -3.52$,

$$M_0^* = -27.10, \quad z^* = 2.75, \quad \sigma_z = 0.93, \quad \text{and} \quad \Phi^* = 8.22 \times 10^{-7} \text{ Mpc}^{-3} \text{ mag}^{-1}.$$

The sample in Warren et al. (1994) includes only one quasar with $z > 4$, and the slope of decay for the fit in Pei (1995) is not well determined. Indeed, using this fit predicts many fewer quasars at $z > 4$ than are found in the surveys of Schneider, Schmidt, & Gunn (1991a, 1994, hereafter SSG), Kennefick, Djorgovski, & de Carvalho (1997), and Irwin, McMahon, & Hazard (1991). So in the simulation we use the fit of Schmidt, Schneider, & Gunn (1995) for the quasar LF at $z > 4.0$:

$$\log \Phi(z, < M_B) = -6.835 - 0.43(z - 3) + 0.748(M_B + 26) \quad (10)$$

or

$$\Phi_{hz}(M, z) = \frac{\Phi_{hz}^*}{10^{0.4[M - M_{hz}^*(z)](\beta_{hz} + 1)}}, \quad (11)$$

where $M_{hz}^*(z) = -27.72 + 0.57z$, $\beta_{hz} = -2.87$, and $\Phi_{hz}^* = 2.42 \times 10^{-7} \text{ Mpc}^{-3} \text{ mag}^{-1}$.

The simulated quasar catalog is generated as follows:

1. For a specific area on the sky, the total numbers of quasars for the low- ($z < 4$) and high-redshift ($z > 4$) subsamples are calculated from equations (9) and (11), respectively, and the redshift and absolute magnitude values (in M_B) of each quasar are derived from their distribution functions.
2. For each quasar, the continuum and emission+absorption spectra are determined. The SDSS colors are calculated from the spectrum.
3. Finally, the Galactic extinction and photometric errors are added.

The number-magnitude relation of quasars in the i' band is shown in Figure 10. We compare the predicted numbers of quasars at low and high redshift with observed numbers (normalized to the survey area of $10,000 \text{ deg}^2$). For $z < 4.0$, our numbers agree well with the results of Crampton, Cowley, & Hartwick (1987; *dotted line*). To transfer B to SDSS i' , we assume $i' = B - 0.5$, appropriate for a quasars with $z < 2.5$. For $z > 4.0$, we compare our predictions with the following observations:

1. SSG, who found 10 quasars with $z > 4$ in 61 deg^2 . We assume the limiting magnitude of SSG is $r \sim 21$, and $i' \sim r - 0.3$ for quasars with $z \sim 4.3$ (Fig. 12c).
2. The DPOSSII survey (Kennefick et al. 1995), which found nine quasars with $z > 4$ in 681 deg^2 with limiting magnitude $r < 19.6$, assuming an average efficiency of 50% as estimated in the paper.
3. The APM survey (Irwin et al. 1991; Storrie-Lombardi et al. 1996), which found 27 quasars with $z > 4$ and $R < 19$ in 2500 deg^2 . Our predicted numbers for $z > 4$ agree well with those observations.

In the whole SDSS survey one expects to find $\sim 100,000$ quasars to the spectroscopic survey limit of $i' \sim 19$. Figure 11 shows the predicted numbers of quasars as a function of redshift for different magnitude limits.

4.3. Evolution of Quasar Colors with Redshift

Figure 12 shows the evolution of the median and scatter of quasar colors with redshift in the SDSS bands, calculated

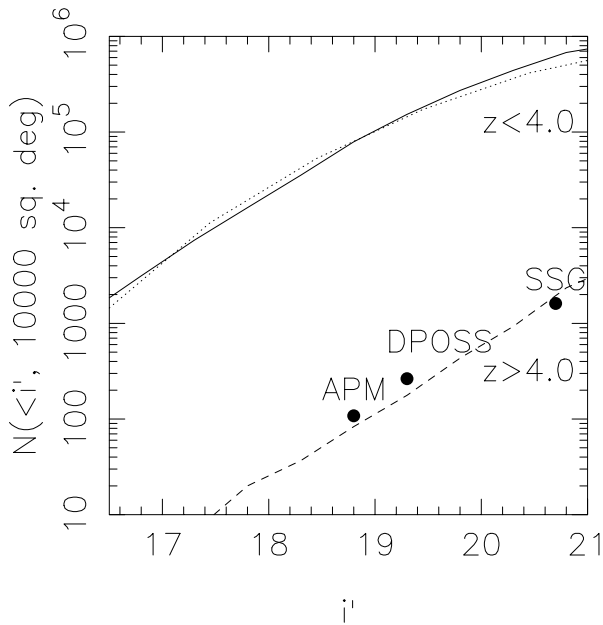


FIG. 10.—Predicted number-magnitude distributions of low- ($z < 4$) and high- ($z > 4$) redshift quasars. They are compared with observed counts: the low-redshift results of Crampton et al. (1987; dotted line), and the SSG, DPOSS, and APM surveys for $z > 4.0$.

from 100 Monte Carlo realizations in each bin of 0.1 in redshift. From this plot, we notice the following:

1. For $z < 2.0$, the $u'-g'$ color (Fig. 12a) of quasars remains smaller than 0.3. In this redshift range, quasar colors are dominated by the power-law continuum. The colors of a pure power-law spectrum are independent of redshift for a given index α .

All quasar emission lines other than $\text{Ly}\alpha$ have equivalent widths smaller than 100 Å and thus have only small effects on broadband colors. The small change in the mean colors at low redshift is caused by those emission lines passing

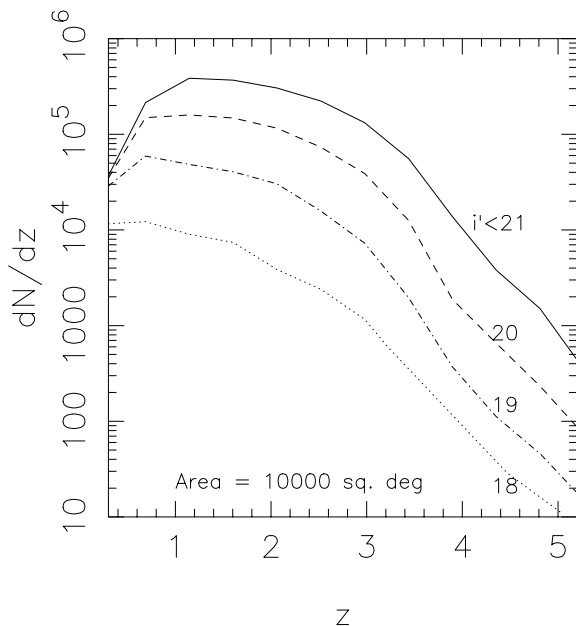


FIG. 11.—Predicted quasar number counts as a function of redshift for several different magnitude cuts for the SDSS survey.

through the u' and g' bands. At $z \sim 2.0$, $\text{Ly}\alpha$ emission enters the u' band, making the $u'-g'$ color bluer. When $\text{Ly}\alpha$ emission moves from u' to g' at $z \sim 2.5$, the $u'-g'$ color gets redder.

2. For $z > 2.5$, the absorption systems, first the $\text{Ly}\alpha$ forest, then the LLSs, enter the u' band. They absorb most of the continuum radiation in the u' band, and the $u'-g'$ color reddens quickly with redshift. Meanwhile, the random distribution of the absorption systems, especially the presence or absence of LLS, optically thick to continuum, in the u' band, causes very large scatter of color for a given redshift. The number density of LLS grows with redshift (see Appendix B); for $z > 3.7$, very few quasars still have detectable u' flux.

3. The evolution in redder colors ($g'-r'$, $r'-i'$, $i'-z'$; Figs. 12b–12d) has similar behavior to that of $u'-g'$, only shifted to higher redshift. LLSs dominate the g' , r' , and i' bands from $z = 4.0, 5.0$, and 6.0 , respectively. The major difference of the quasar color evolution of those colors from that of $u'-g'$ is that there is no obvious dip at the redshift when $\text{Ly}\alpha$ emission is in the bluer band, as it is when the emission is in u' at $z \sim 2.0$. This is because for quasars with $z > 4$, $\text{Ly}\alpha$ forest lines absorb more than half of the continuum blueward of $\text{Ly}\alpha$ emission, so the increase of flux from $\text{Ly}\alpha$ emission in a certain band is cancelled by absorption immediately blueward of it.

Figure 13 compares the simulation results with observations of quasars in SDSS colors. The simulation is for an area of 10 deg^2 ; only quasars with $i' < 19.5$ are plotted. There are very few quasars at $z > 3.5$. To generate enough simulated quasars for comparison, we assume the quasar LF to be constant at $z > 3.5$, instead of the evolving LF described above. The observations are taken from test observations of known quasars in the SDSS bands by Richards et al. (1997), Newberg et al. (1998), and Fan et al. (1999). The simulated colors agree with the observations very well, with few exceptions, where the observed quasars have broad absorption lines (BALs), not being simulated in this work. In Figure 13d, we plot the median and scatter of i' magnitude for a quasar with $M_B = -27$, a very bright quasar similar to 3C 273, as a function of redshift. SDSS could easily detect such a quasar at $z > 5.0$, if there were any in the universe. Discovery of such quasars would have profound impact on the studies of the structure formation and early galaxy evolution.

In § 6, we will compare the quasar colors with other types of stellar objects, and investigate how to separate quasars from other objects in color space in detail.

5. SIMULATION OF COMPACT EMISSION-LINE GALAXIES

Some compact emission-line galaxies (CELGs) have power-law continua and strong emission lines, and thus their colors are similar to those of low-redshift quasars. Some CELGs have very strong narrow emission lines, with equivalent width exceeding 1000 Å. Those strong emission lines affect the colors greatly, and this effect can change with redshift as the emission lines pass through different bands. The emission spectra of CELGs differ greatly in ionization and excitation levels. Their continua can be powered by starburst activity or a central black hole. Their luminosity function is not well established (Salzer 1989), and little is known about their redshift evolution (Koo et al. 1994). Furthermore, the possibility that those objects will be identified

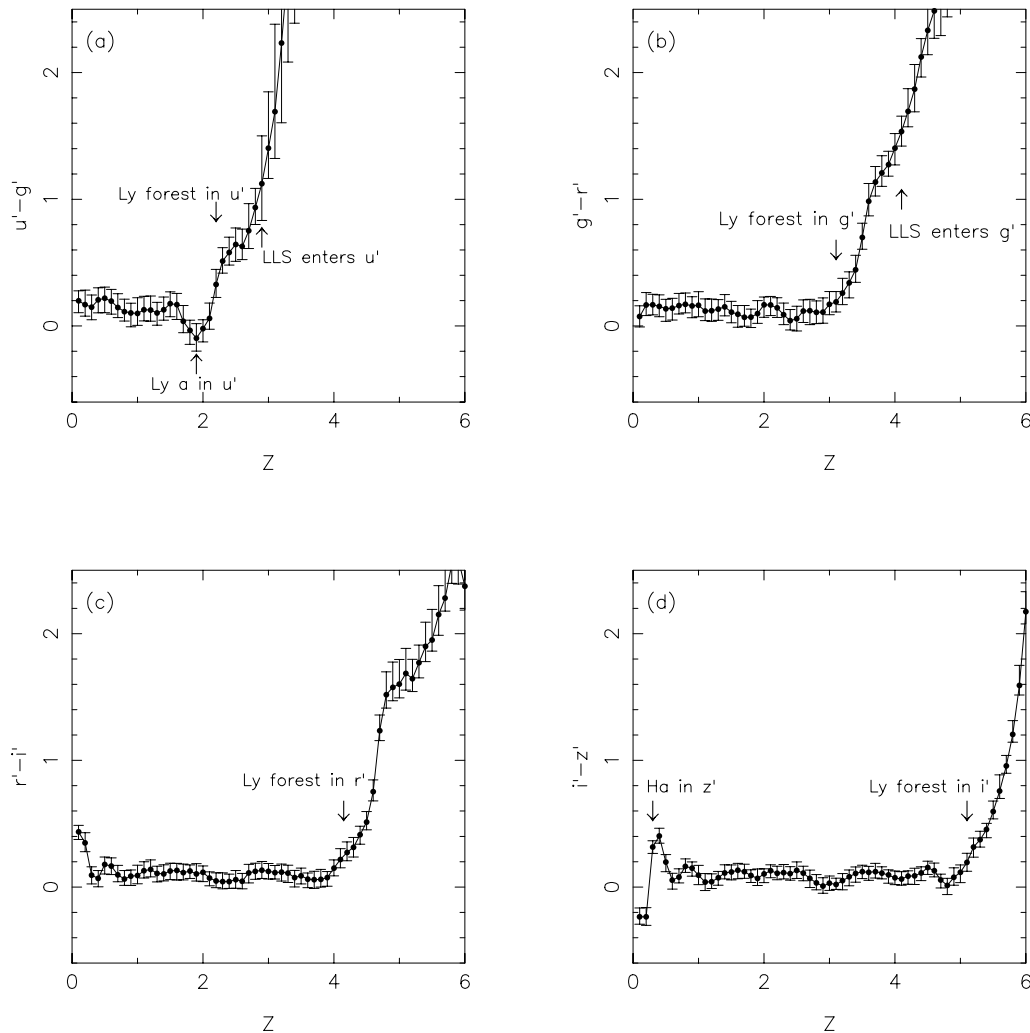


FIG. 12.—Color evolution of quasars as a function of redshift. The figures show the median and 68% scatter calculated from simulations of 100 quasars at each redshift.

as stellar objects in the SDSS is a function of both the observing condition and the surface brightness profiles of those objects. It is therefore very difficult to simulate the distribution of CELGs in the SDSS color space. Instead of attempting to develop a population synthesis model for CELGs (as we did for stars and quasars), we construct the simulated catalog by directly using the spectroscopic data of the SSG survey. We first calculate the SDSS magnitudes of the SSG galaxies according to their line fluxes (Horowitz 1994) and the continuum slope distribution of McQuade, Calzetti, & Kinney (1995). In constructing the catalog, we randomly select galaxies from the SSG sample (for $r' < 19$, where they are complete, and some extrapolation beyond that) and assign the angular size of each object according to the size distribution of Salzer, MacAlpine, & Boroson (1989).

5.1. CELG Model

SSG sample.—We first want to define a flux-limited, uniformly selected subsample from the SSG sample. The SSG survey selects emission-line object candidates based on Palomar 4-Shooter grism spectra. Their emission-line galaxy sample goes deep (close to the magnitude limit of the SDSS spectroscopic survey for quasars, $i' \sim 19$) and has

large sky coverage (61.4 deg^2). Horowitz (1994) analyzed the spectral properties of the SSG sample. He measured the line flux and EW of $\text{H}\beta + [\text{O III}]$, as well as line ratios of $\text{H}\alpha$, $[\text{N II}]$, $[\text{S II}]$, and $[\text{O II}]$ for all the SSG galaxies that were originally selected by their $\text{H}\beta + [\text{O III}]$ emission on the grism spectra. Horowitz (1994) also classified the spectral type of the CELG to be one of five types: Seyfert 1, Seyfert 2, starburst, H II galaxy and LINER. This information is the basis of our simulation.

Horowitz (1994) does not include galaxies detected only by $\text{H}\alpha$ from the grism. They make up about 30% of all the SSG galaxies. But they are mostly of very low excitation (otherwise, they would have been detected by their $\text{H}\beta + [\text{O III}]$ emission), an indication of low activity, thus unlikely to have nonstellar colors, and therefore not a serious problem for the quasar selection. Furthermore, they are at low redshift (in order for $\text{H}\alpha$ to be within the range of grism spectra) and thus have bigger sizes, which makes them unlikely to be classified as stellar objects by the SDSS. We ignore those galaxies in the current simulations.

SSG select objects according to two criteria: (1) equivalent width of $\text{H}\beta + [\text{O III}]$ greater than 50 \AA , and (2) line flux greater than about $10^{-14.1} \text{ ergs s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1}$. Transferring these constraints to continuum flux, this sample is

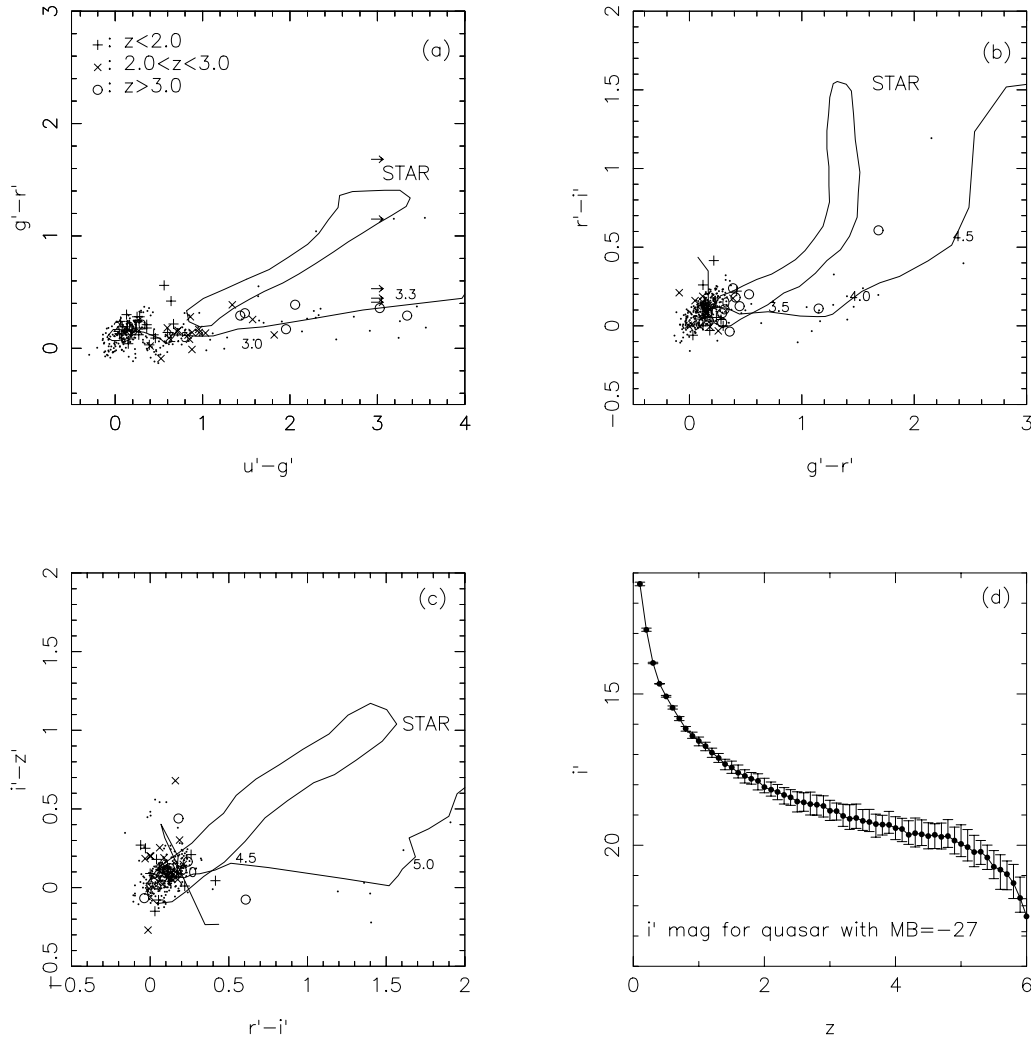


FIG. 13.—(a)–(c) Color-color diagrams of quasars, from a simulation of 10 deg^2 with $i' < 19.5$ (small dots), compared with observations in SDSS filters of known quasars (larger symbols) at various redshifts. Arrows represent lower limits in $u'-g'$ due to nondetections in u' observations. The solid lines labeled with redshift are the median tracks of quasar colors calculated from Fig. 12. Regions labeled as STAR are the approximate boundaries of the stellar locus. (d) The median and 68% scatter of the i' magnitude of a quasar with $M_B = -27$ (similar to 3C 273), calculated from 100 realizations at each redshift.

“complete” for all galaxies with equivalent width greater than 50 \AA and $m_{\text{AB}}^{\text{con}} < 19$. For fainter galaxies, this sample is complete only for galaxies with even stronger lines. Therefore, we select a “complete” subsample of 167 galaxies with $r' < 19$ and refer it as SSG19.

The redshift distribution of the SSG sample is fitted by a Γ distribution, $N(z) \propto z \exp(-z/z_0)$ where $z_0 = 0.08$, and will be used in the simulation to generate fainter CELGs. The number count versus magnitude relation is consistent with $N(m) \propto 10^{0.4m}$. For $r' = 19$, it has a surface density of 2.7 deg^{-2} .

McQuade et al. sample.—McQuade et al. (1995) provide UV-to-optical spectrophotometry of a sample of 31 emission-line galaxies. From this sample, we calculate the continuum slope of different types of CELGs. We assume the continuum of CELGs to be power law in the relevant wavelength range; the power-law indices of each type are as follows: Seyfert 1: $\alpha = -1 \pm 0.5$; Seyfert 2: $\alpha = -2 \pm 0.5$; starburst: $\alpha = -1.5 \pm 0.5$; H II galaxy: $\alpha = -1 \pm 0.5$; LINER: $\alpha = -2 \pm 0.5$.

The spectrum of an emission-line galaxy is assumed to be a power-law continuum plus 16 major UV and optical emission lines. For each galaxy in the SSG19 sample, we ran-

domly select a continuum slope based on the statistics above, assuming Gaussian distribution. If a line ratio is not measured in Horowitz (1994), we will assume a flux ratio (relative to H β) based on the measurements of McQuade et al., according to its type: for Seyfert, we use the flux ratios of NGC 1068; for starburst/LINER, NGC 6052; and for H II galaxy, UGC 9560. The SDSS magnitudes of all the SSG19 galaxies are calculated.

Size distribution.—Salzer et al. (1989) studied the properties of a deep, complete sample of emission-line galaxies from the University of Michigan survey. The isophotal magnitude and isophotal radius are determined for each galaxy. Using their sample, we get a relation between the absolute magnitude and isophotal radius:

$$\log D_{25} = (-2.424 \pm 0.093) - (0.183 \pm 0.005)M_B. \quad (12)$$

The rms scatter of $\log D_{25}$ around this relation is 0.119. This relation is plotted in Figure 14. Different spectral types are represented by different symbols. It shows no strong selection effect against low surface brightness galaxies toward fainter magnitude. Assuming an exponential disk, one can derive the scale length and characteristic surface brightness R_e and μ_0 . The distribution of μ_0 is roughly a

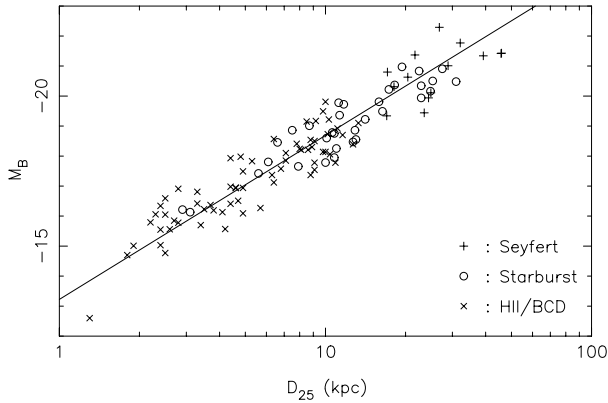


FIG. 14.—Relation between M_B and isophotal diameter D_{25} from the emission-line galaxy sample of Salzer et al. (1989). Different types of galaxies are indicated by different symbols. They follow the same best-fit relation (eq. [12]).

Gaussian, peaking at $\sim 20.6 \text{ mag arcsec}^{-2}$, 1 mag brighter than the Freeman (1979) central surface brightness of disks. In the simulation, the D_{25} of a CELG is calculated from equation (12), with a Gaussian scatter of 0.119 in logarithm.

R_e is then calculated assuming a pure exponential disk. The SDSS imaging observations will be taken only under seeing $\lesssim 1''$. Under such observing conditions, the photometric pipeline will be able to resolve galaxies with $R_e > 0.5''$ at $r' < 19$. Therefore, the simulated catalog only includes those galaxies with $R_e < 0.5''$. About 50% of the CELGs at $r' < 19$ in the simulated catalog have $R_e < 0.5''$.

5.2. Colors of CELGs

The simulated catalog is constructed separately for two subsamples: the brighter sample, where SSG is roughly complete (SSG19, $r' < 19$), and the fainter sample:

1. *Brighter sample*.—After specifying the size of the area, (1) the average number of galaxies with $r' < 19$ in this area is calculated. (2) The actual number N of galaxies is selected randomly from a Poisson distribution. (3) N galaxies from the SSG19 sample are randomly chosen and the SDSS magnitudes are calculated according to Horowitz's measurement and the spectral model described above. Note that this brighter sample is based on the SSG sample; in particular, there is no assumption about the evolution or luminosity function of these objects. The brighter sample is as deep as the SDSS spectroscopic survey for quasars.

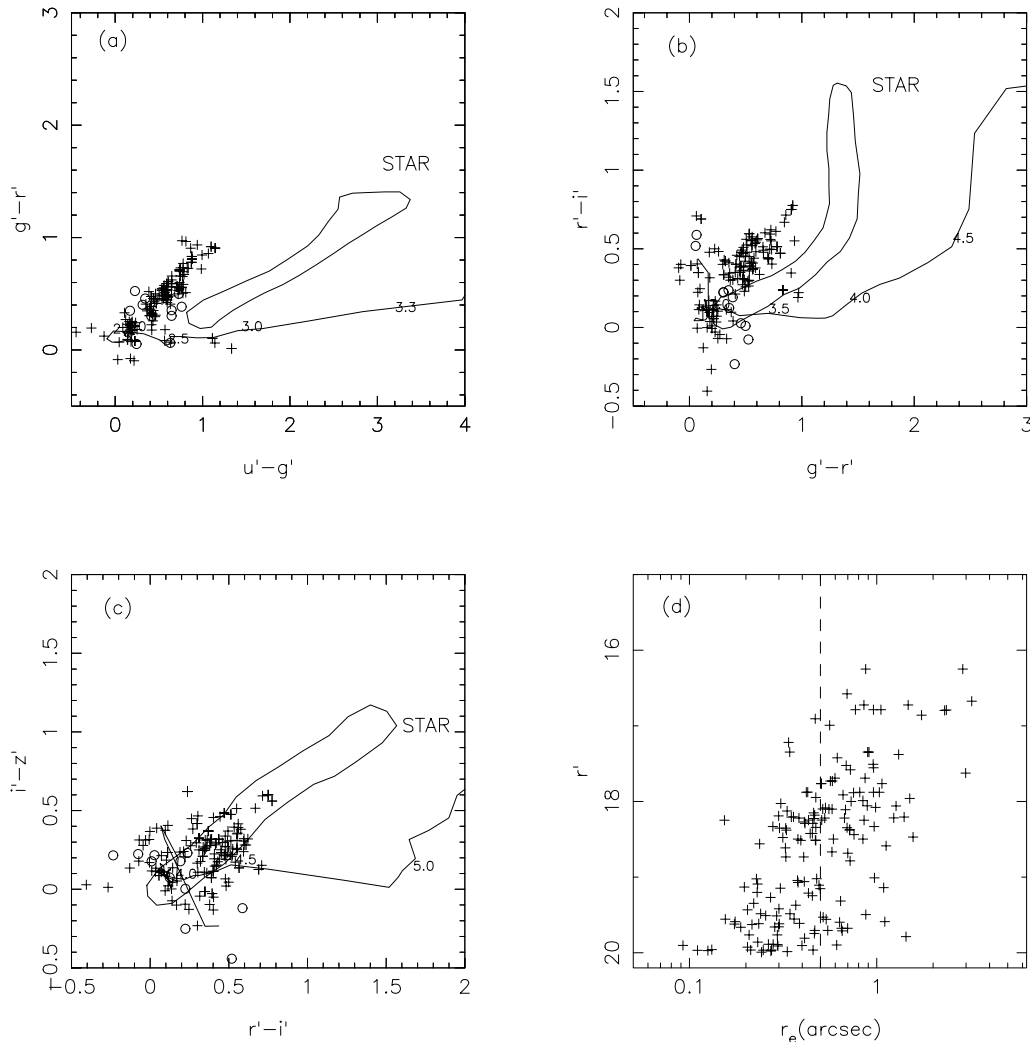


FIG. 15.—Color-color diagrams and size distribution of simulated CELGs (*plus signs*). The observed colors of CELGs from Hall et al. (1996, transferred to SDSS bands) are plotted as open circles. We also show the approximate stellar loci and the quasar tracks at different redshifts in the diagrams.

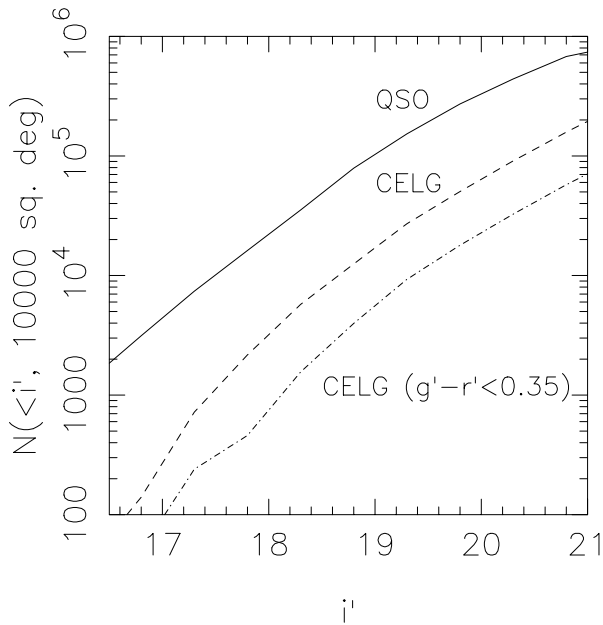


FIG. 16.—Predicted number-magnitude relation of CELGs. It is compared with quasar number counts as well as the number counts of blue CELGs ($g' - r' < 0.35$; see § 6.3).

2. *Fainter sample.*—After specifying the size of the area, (1) the average number of galaxies with $19 < r' < 24$ is calculated assuming $N(m) \propto 10^{0.4m}$ and the normalization of SSG19. (2) The actual number N in the fainter sample is selected randomly from a Poisson distribution. (3) For each of the galaxies, the r' magnitude and redshift are drawn randomly from the distribution of SSG galaxies in § 5.1. (4) We randomly pick N galaxies from the SSG sample. For each of the galaxies, we use its type, the EW of $H\beta + O\text{ III}$, and line ratios, and determine the continuum slope and line ratios of weaker lines from the spectral model described above according to its spectral type. We calculate the SDSS magnitudes from the spectrum.

We calculate the angular size of the galaxy as described above, rejecting objects with $R_e < 0''.5$. Finally, we add the Galactic extinction and photometric errors to each of the galaxies.

In Figure 15, we present the color-color diagrams and size distribution of a simulation of CELGs in 10 deg^2 . We also compare the simulation with the colors of CELGs found by Hall et al. (1996; catalogs given in Osmer et al. 1998), transferred to the SDSS colors using the relations in F96. We plot only the CELGs with $z < 0.25$ in Hall et al.'s sample (their survey is much deeper than the SDSS quasar survey).

From the figure, we note the following:

1. The SDSS colors of most of the emission-line galaxies are dominated by the continuum. This is simply due to the large bandwidth of SDSS filters. An emission-line EW of 100 \AA will have only about 0.1 mag effect on the magnitude at most (if the line is at the center of the filter). Half of the galaxies in SSG19 have $\text{EW}(H\beta + [O\text{ III}]) < 100 \text{ \AA}$. The obvious locus of emission-line galaxies in Figure 15 is the color of this power-law continuum.

2. The colors of CELGs are on average redder than those of quasars, owing to their redder continua. The accurate

modeling of the locations of those galaxies in color space critically depends the choice of power-law distribution. Hall et al. (1996) conclude that putting a limit on $B - V < 0.6$ will eliminate half of the CELGs from the quasar candidates. Our simulation also includes some galaxies redder than those of Hall et al. They have colors similar to normal late-type galaxies and will not be picked up by Hall et al.'s color selection criteria.

3. Some CELGs have very strong emission lines and almost no continuum, which leads to enormous EWs (some of the SSG galaxies have emission lines with EW exceeding 6000 \AA), which results in a color change of several magnitudes from the power law. The redshift effect will make their colors even more peculiar. Those CELGs are located in parts of color space very far from normal stars and quasars.

The number-magnitude relation of CELGs in the i' band is presented in Figure 16. We compare it with the number count of quasars (Fig. 10). The number density of CELGs is 20%–25% of the number density of quasars for a large magnitude range (see also Koo et al. 1994). However, since colors of CELGs are redder than those of quasars, rejecting sources with $g' - r' > 0.35$ will eliminate about 60% of CELGs while retaining most of the quasars. We will discuss the separation of quasars and CELGs further in § 6.

6. DISCUSSION

6.1. Distribution of Stellar Objects in the Color Space

In this subsection, we summarize the results of the simulations described in the previous sections and show how different types of stellar objects occupy different regions of the color space.¹

Figure 17 shows a 10 deg^2 simulation of all four kinds of stellar objects toward the NGP, with $i' < 19.5$, in $u' - g'$, $g' - r'$, and $r' - i'$ color space. In Figure 18, we show the approximate locus of each type of object on color-color diagrams, estimated from Figure 17. From Figures 17 and 18, we can study the relative distribution of stellar objects in color space:

1. Bright and hot white dwarfs occupy the bluest corner of color space. First, they are hotter and bluer than normal stars. Second, they have similar $u' - g'$ colors to quasars, but with bluer $g' - r'$, $r' - i'$, and $i' - z'$ colors. For hot white dwarfs ($T_{\text{eff}} > 10,000 \text{ K}$), their spectra are already in the Rayleigh-Jeans part of the blackbody curve at $\lambda > 5000 \text{ \AA}$, i.e., $f_{\nu} \propto \nu^2$, much bluer than those of quasars or AGNs with $f_{\nu} \propto \nu^{-0.5}$. This property can be used to separate hot white dwarfs from quasars.

2. Low-redshift quasars ($z < 2.0$) have blue $u' - g'$ colors. They are well separated from stars and white dwarfs in color space and may be confused only with some CELGs. The position of quasars at $z < 2.0$ changes little with redshift, owing to the dominant power-law continuum. For $z > 2$, the $u' - g'$ color becomes increasingly red.

3. Quasars with $2.5 < z < 3.0$, white dwarfs with $T_{\text{eff}} \sim 7000 \text{ K}$, blue halo stars and some CELGs intersect in the region of color space with $u' - g' \sim 0.8$, $g' - r' \sim 0.3$, and $r' - i'$, $i' - z' \sim 0$. For quasar at this redshift range, (1) the $\text{Ly}\alpha$ emission is redshifted to $\sim 4500 \text{ \AA}$; (2) the continuum radiation blueward of $\text{Ly}\alpha$ emission is absorbed by $\text{Ly}\alpha$

¹ A collection of the simulated catalogs, for quasars, CELGs, and for stars and white dwarfs toward several directions on the sky, is available at http://www.astro.princeton.edu/~fan/sdss_simu.html.

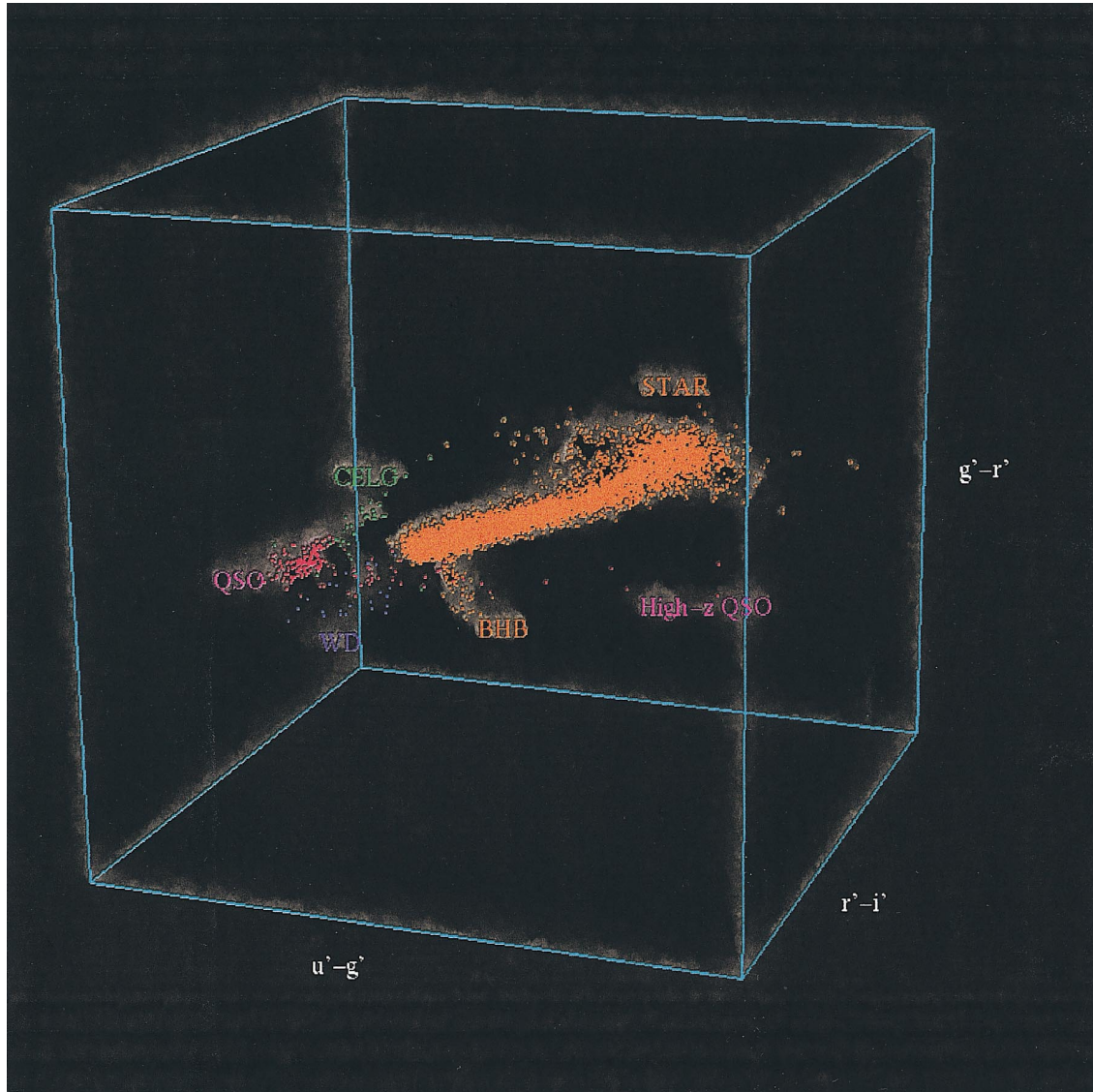


FIG. 17.—Distribution of stellar objects in $u'-g'$, $g'-r'$, and $r'-i'$ color space in a simulation of stars, white dwarfs, quasars, and CELGs for a 10 deg^2 region toward the north Galactic pole.

forest lines and LLSs, which causes a break at $\sim 4000 \text{ \AA}$ in the SED; and (3) the continuum redward of $\text{Ly}\alpha$ emission gradually declines toward longer wavelength. These three features have very similar wavelengths to the features in a stellar spectrum of $T_{\text{eff}} \sim 7000 \text{ K}$: (1) the flux peak at $\sim 4500 \text{ \AA}$, (2) the Balmer jump and Balmer absorption lines at $3500\text{--}4000 \text{ \AA}$, and (3) the gradual decline of flux at longer wavelengths. These coincidences cause the broadband colors of these objects to be very similar (see also Fig. 1). This is the most difficult redshift range for quasar selection. Furthermore, since both the stellar populations and Galactic extinction change with Galactic position, the detailed distribution of stars and white dwarfs in color space also change, which makes the quasar selection even more complicated.

4. For $z > 3$, quasars have very red colors because of the intervening absorption systems in their spectra. They are well separated from other kinds of stellar objects in color space.

5. CELGs have a big scatter in their distribution in color space. Most of them have redder colors than quasars due to

their redder continua. Those with AGNs have similar colors to quasars. Some H II galaxies, with very strong optical emission lines dominate their spectra, can have very peculiar colors.

6.2. The “Fundamental Plane” in Color Space

Newberg et al. (1999) point out that in their observations, stars, galaxies and low redshift quasars are distributed approximately in the same “fundamental plane” in SDSS color space. The stellar locus, in particular, forms a ribbon-like structure in color space. Our simulations further demonstrate this point. Using the algorithm developed in Newberg & Yanny (1997), we fit a set of stellar locus points to our stellar distribution in the simulated color space of $(u'-g', g'-r', r'-i')$. This locus is shown in Figure 18. The distribution of stars perpendicular to this locus is then fitted by an ellipse. We choose a set of axes:

$$\begin{cases} c1 &= 0.95(u' - g') + 0.31(g' - r') + 0.11(r' - i') \\ c2 &= 0.07(u' - g') - 0.49(g' - r') + 0.87(r' - i') \\ c3 &= -0.39(u' - g') + 0.79(g' - r') + 0.47(r' - i') \end{cases} \quad (13)$$

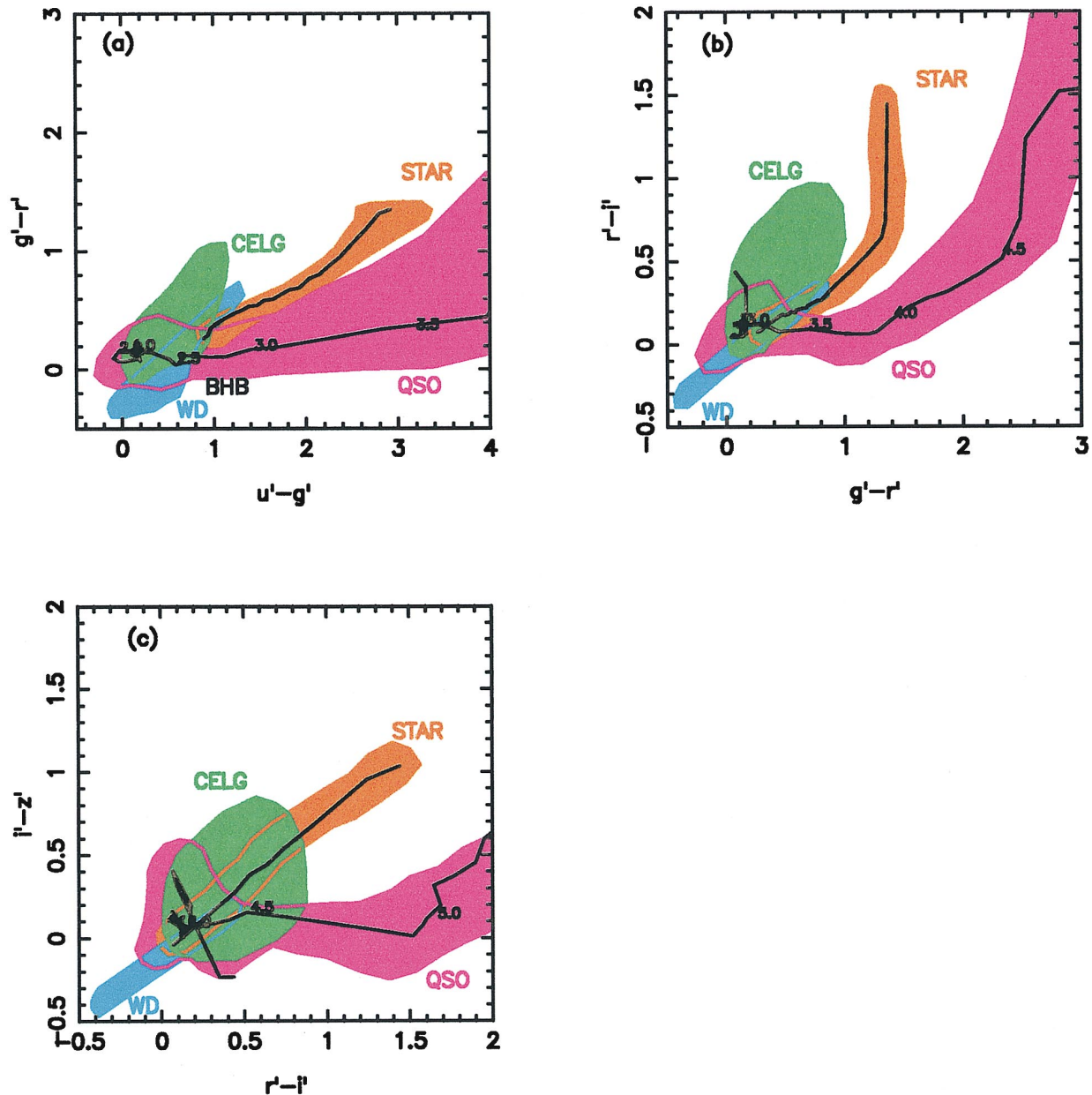


FIG. 18.—Approximate locations of each kind of object on the color-color diagrams estimated from the simulation shown in Fig. 17. The solid lines going through the stellar regions are the best-fit stellar locus points using the algorithm of Newberg & Yanny (1997). The lines going through the quasar regions are the median quasar tracks as a function of redshift.

in which $c1$ is along the average direction of the stellar locus for $T_{\text{eff}} > 4000$ K (in Fig. 18), $c2$ is along the major axis of the fitted ellipse perpendicular to the stellar locus, and $c3$ is along the minor axis of the ellipse.

In Figure 19, we plot the NGP simulation in the projections of $c1$ versus $c2$ and $c1$ versus $c3$, i.e., the edge-on and face-on view of the “fundamental plane” on which most of the stars are distributed. We show the reddening vectors on these projections. We also plot the tracks of blackbody spectra (dotted line, from 2000 to 30,000 K) and power-law spectra (solid line, $\alpha = 0.0$ –2.0) on those projections.

Figure 19a demonstrates the following:

1. Normal stars, white dwarfs, as well as quasars and most of the CELGs, are distributed in a plane, with $c2 = 0.07(u' - g') - 0.49(g' - r') + 0.87(r' - i') \sim 0.0 \pm 0.1$.

Note that $c2$ has almost no contribution from u' . For a blackbody spectrum, $c2 \sim 0$ for all temperatures. The fact that all but the coolest stars are distributed on this plane merely reflects the fact that their SEDs are close to blackbody except in the u' band, which covers the Balmer jump for hotter stars and the metal line blending in UV for cooler stars.

2. For M stars ($T_{\text{eff}} < 3500$ K), the optical spectra are dominated by molecular absorption and thus have big deviations from a blackbody spectrum. They therefore depart from the “fundamental plane.”

3. The colors of quasars and CELGs are dominated by the power-law continuum. They have $c2 \sim 0.05$ –0.10.

4. The reddening vector points almost entirely in the $c1$ direction. Galactic extinction can only move objects within

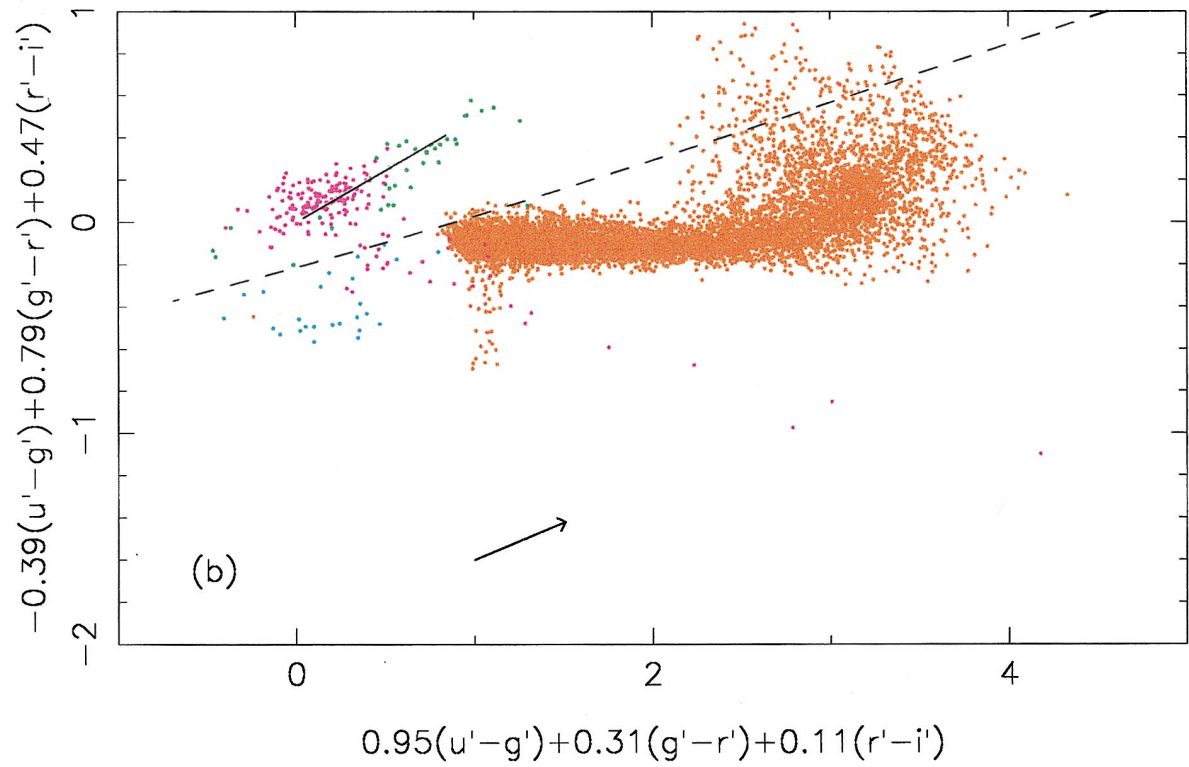
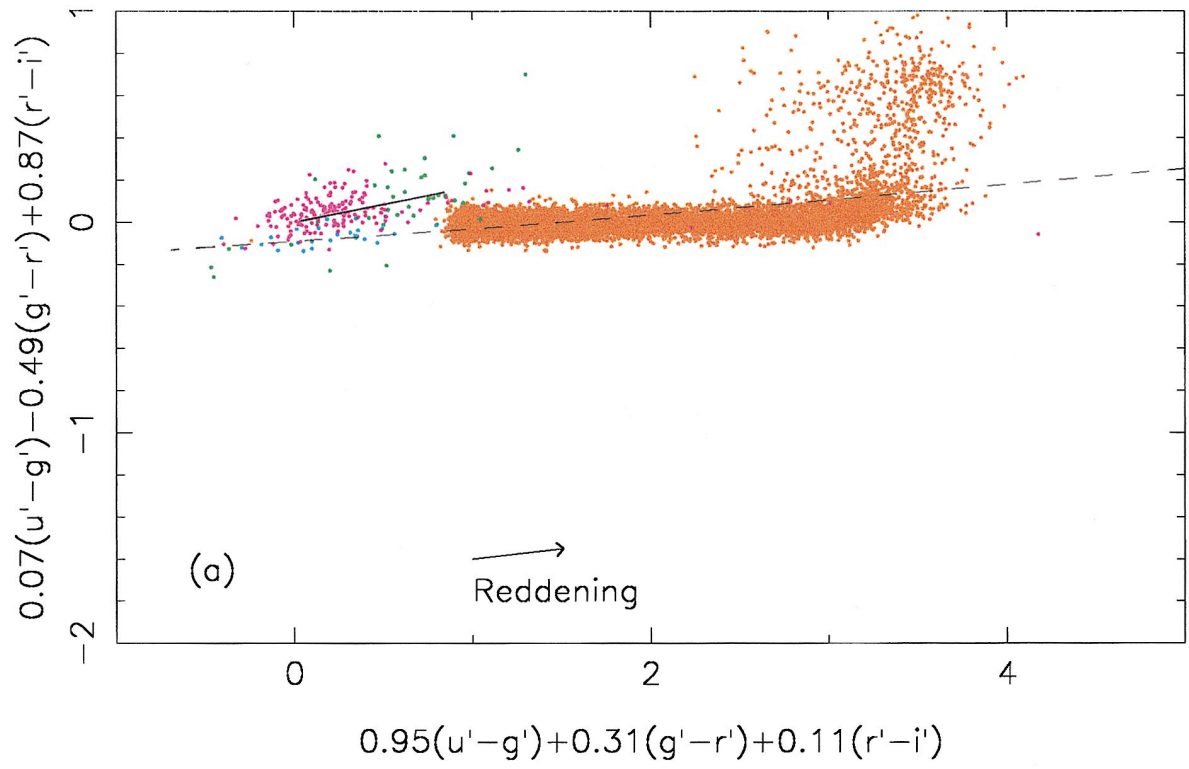


FIG. 19.—Projections of the stellar objects (from the simulation shown in Fig. 17) onto the $c1$ - $c2$ and $c1$ - $c3$ plane (see text). Normal stars (*orange points*), white dwarfs (*blue points*), CELGs (*green points*), and quasars (*magenta points*) distribute near to a fundamental plane $c2 \sim 0$ in SDSS color space. The dashed lines represent the colors of blackbodies for $2000 < T < 30,000$ K, and solid lines represent the color of power-law spectra with $0 < \alpha < 2$.

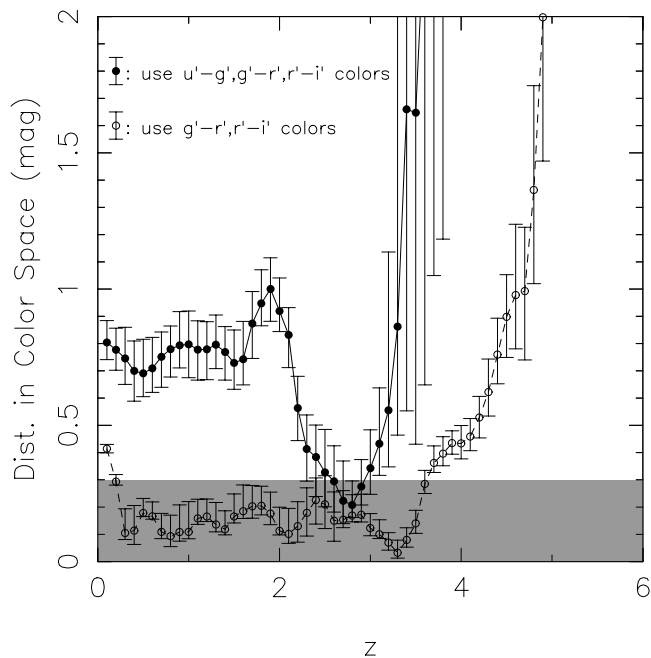


FIG. 20.—Distance from the quasar locus to the stellar locus based on three (filled circles) and two (open circles) SDSS colors. The shaded region marks the redshift range where the distance is smaller than 0.3 mag. Quasars overlap with the stellar locus in the redshift range $2.5 < z < 3.0$.

the plane. The blue edge of the stellar distribution in the $c1$ versus $c2$ projection depends on the stellar population as well as reddening.

The points above explain the existence of such a “fundamental plane” in color space.

Figure 19b shows the face-on view of the “fundamental plane.” On this plot, different types of objects are well separated. For a fixed temperature, changes in stellar metallicity, gravity, and reddening will result in different positions on the $c1$ - $c3$ plane. The BHBs stand out clearly from the stellar locus. They are also located in different positions from the white dwarfs with similar temperature ($c1 \sim 1.0$ for BHBs, and $c1 \sim 0.3$ for white dwarfs with similar temperature) owing to the large difference in their surface gravities and therefore different strength of the Balmer jump.

6.3. The Quasar Selection Problem

It is not the goal of the current paper to discuss the quasar selection algorithm of the SDSS in great detail. Aspects of it are discussed in Newberg & Yanny (1997). The final algorithm will be established during the SDSS test year. However, we intend to show through simulations the separation of quasars and other objects in color space, following the discussion in § 6.1.

In Figure 20, we plot the median distance and its 68% scatter of quasars from the stellar locus (see § 6.1) in color space as a function of redshift. We show both the distances based on three colors ($u' - g'$, $g' - r'$, and $r' - i'$) and based on the two redder colors. It demonstrates two problems with the separation of quasars from stars with colors.

1. At redshift 2.5–3.0, quasars move very close to the stellar locus in color space. Some of the quasars have colors in all bands that are almost identical colors to blue halo

stars as we saw above. Any quasar selection based on color will miss a fraction of all quasars in this redshift range or will suffer a large contamination from stars. Note that this redshift range is also where the quasar LF peaks (§ 4.2). The evolution of the quasar population in this redshift range is important to cosmology. But we will unavoidably need a big selection correction here. Detailed modeling of the quasar population, as we described in § 4, is crucial in determining the quasar selection function (see also Warren et al. 1994).

2. For $z > 3.3$, quasars become increasingly redder in $u' - g'$. They are well separated from the stellar locus. However, they will have very faint u' detections or, in most cases for $z > 4$, will be completely absent from the u' images. One has to use the lower limit in $u' - g'$ to separate quasars and stars. Quasars with $z < 4.2$ can have very similar $g' - r'$ and $r' - i'$ colors to stars. Quasars with very high redshift ($z > 4.2$) can be separated from stars based on $g' - r'$ and $r' - i'$ colors alone.

White dwarfs, BHBs, and CELGs are also rare (compared to normal stars) and are well separated from the stellar locus. Simply selecting objects as outliers from the stellar locus (see, e.g., Newberg & Yanny 1997) will include those contaminants in the quasar candidate list. White dwarfs will outnumber quasars for $r' < 17$, while CELGs have a number density about 25% of that of quasars at the faint end (Koo et al. 1994). In Figure 21, we show the separation of quasars in the redshift range $2.2 < z < 3.2$ and stars/white dwarfs/CELGs. As discussed in § 6.1, hot white dwarfs (with $T_{\text{eff}} > 8000$ K) have similar $u' - g'$ but bluer $g' - r'$ color than quasars; a cut at $g' - r' > -0.20$ will eliminate most of them. At $r' < 17$, most of the white dwarfs are hot and have $g' - r' < -0.20$. At $r' < 19.5$, there are more cooler white dwarfs, and $\sim 60\%$ of white dwarfs can be eliminated in this way. Most of the CELGs with H II or starburst spectra will have redder continua than those of quasars or Seyfert galaxies. By rejecting objects with $g' - r' > 0.35$, we exclude $\sim 60\%$ of CELGs (see also Fig. 16).

7. SUMMARY

We present in this paper a simulation of the color-magnitude distribution of the stellar sources in the SDSS filter system. We simulate the distributions of Galactic stars, white dwarfs, quasars, and compact emission-line galaxies that the SDSS photometric pipelines will pick as stellar sources by morphology. We do not simulate the colors of normal galaxies that may appear stellar for the SDSS. We also do not simulate various kinds of peculiar stars (see § 2.4). With these exceptions aside, we have covered the major classes of stellar objects that the SDSS will observe at high Galactic latitudes. In this paper, we present the models of the spatial and spectral distributions for each type of object. We also add realistic photometric errors and Galactic extinction in the simulation.²

We simulate the distribution of stars based on (1) evolutionary stellar population synthesis models of the disk and

² Simulated catalogs of SDSS observations toward several directions on the sky are available at http://www.astro.princeton.edu/~fan/sdss_simu.html.

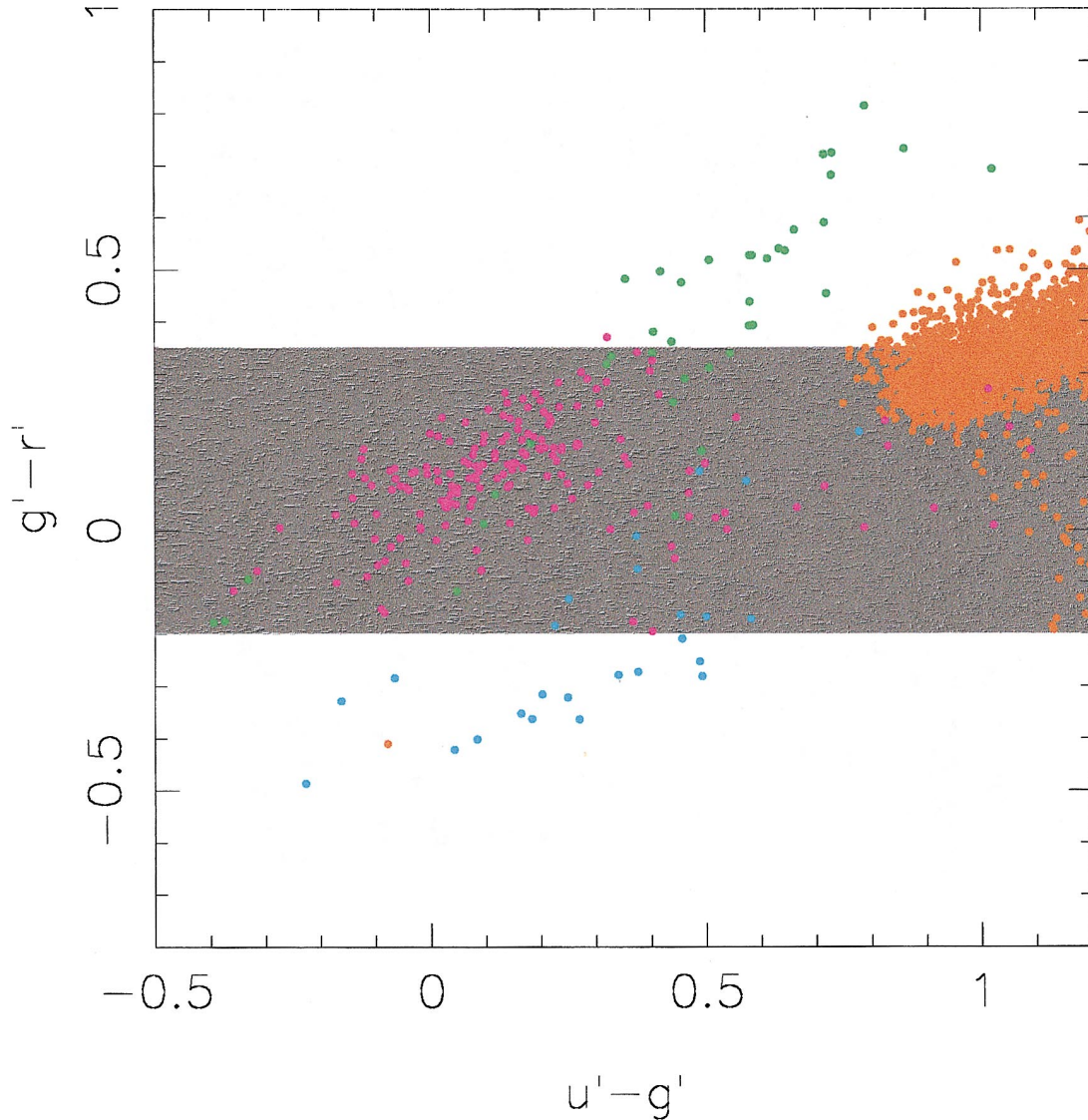


FIG. 21.—Region in the color-color diagram where the white dwarf (blue), star (orange), CELG (green), and quasar (magenta) loci intersect. Using a color cut $-0.20 < g' - r' < 0.35$, more than 50% of CELGs and white dwarfs are eliminated, but very few quasars are dropped.

halo populations, using the Padova stellar isochrones and Lejeune et al.'s (1997a, 1997b) stellar atmosphere libraries; (2) a Bahcall-Soneira (1984)–type Galactic model, including a disk whose height grows with stellar age, and a flattened stellar halo. The disk stars are separated from the halo stars in color-magnitude diagrams as well as in color-color space. The simulation tools will be very useful for constraining Galactic star count models as well as for finding peculiar stars.

We simulate the distribution of white dwarf stars based on (1) their observed spatial distribution (similar to old disk stars), luminosity function and mass distribution and (2) H and He white dwarf atmosphere models (Bergeron et al. 1995) that transfer white dwarf properties to the SDSS magnitude system. We show that white dwarfs outnumber quasars for $r' \lesssim 17$ at moderate Galactic latitude. Hot white dwarfs can be easily picked out (and separated from low-redshift quasars) from their positions in color space.

We simulate the distribution of compact emission-line galaxies based on (1) the luminosity and emission line strength distribution from SSG, (2) the continuum slope

distribution from McQuade et al. (1995), and (3) the size distribution of the UM sample (Salzer et al. 1989). We show that down to the SDSS spectroscopic survey limit of $i' \sim 19$, about half of the emission-line galaxies will be so compact that they cannot be distinguished from stars based on SDSS images. Those galaxies occupy a similar place in color space to that of quasars, but the majority of them are redder than low-redshift quasars.

We simulate the distribution of quasars based on (1) a quasar spectral model, which includes the observed distributions of power-law continuum, optical/UV emission lines, and a large number of absorption systems due to intervening neutral hydrogen; and (2) the evolution of the luminosity function of quasars. We show how the colors of quasars evolve with redshift. The low-redshift ($z < 2.2$) quasars can be easily separated from stars and hot white dwarfs when more than two colors are used in the selection. The colors of high-redshift quasars are dominated by absorption blueward of Ly α emission and are well separated from the stellar locus. The colors of quasars with $2.5 < z < 3.0$, however, are easily confused with hot halo

stars, especially halo blue horizontal-branch stars, and relative cool ($T_{\text{eff}} \sim 7000$ K) white dwarfs. We also discuss the separation of quasars from other types of objects using color in more detail and suggest that a color cut in $g' - r'$ will eliminate most of the emission-line galaxies and hot white dwarfs, which themselves are also well separated from the stellar locus, and could be selected as quasar candidates otherwise. For $2.5 < z < 3.0$, on the other hand, quasars fall very close to the stellar locus. This effect is also a function of Galactic position, with stellar population depending on direction. The simulation is used to establish and test the quasar selection criteria for the SDSS survey.

We show the relative distribution of different type of objects in color space. All four kinds of stellar objects are distributed basically on the same plane in color space, as previously pointed out by Newberg et al. (1999) from test observations using SDSS filters. This is due to the fact that (1) the red part of optical stellar spectra does not deviate much from a blackbody and (2) the power-law continuum that dominates the colors of quasars and CELGs also lies close to this plane. Changes in stellar metallicity, gravity and reddening mostly result in scatter of the stellar distribution *within* this plane.

The simulation developed in this paper will be thoroughly tested and improved when comparing with the SDSS survey data, which is to be available soon. These data will be a very useful tool both for the study of Galactic structure and the selection of peculiar objects from the SDSS survey.

We thank Jim Gunn, Zeljko Ivezic, Robert Lupton, Jill Knapp, Allyn Smith, Heidi Newberg, Yuen Ng, Jeff Pier, Neil Reid, Gordon Richards, Don Schneider, Michael Strauss, and Simone Zaggia for very helpful discussions. We thank Pierre Bergeron for providing theoretical SDSS colors of white dwarf models; Heidi Newberg and Brian Yanny for their stellar locus fitting code; Dawn Lenz for the synthetic colors of observed white dwarfs; Kevin Krisciunas for sending his catalog in digital form; Irwin Horowitz for sending his data in digital form; and Mike Blanton for providing his three-dimensional plotting software. We thank Yuen Ng for his help on stellar simulations. We especially thank Michael Strauss and Jill Knapp for carefully reading and commenting on an earlier manuscript. This research is supported in part by NSF grant AST96-16901 and the University Research Board of Princeton University.

APPENDIX A

SDSS PHOTOMETRIC SYSTEM AND PHOTOMETRIC ERRORS

The SDSS system comprises five bands (u' , g' , r' , i' , and z') that cover the entire range from the atmospheric cutoff in the ultraviolet to the sensitivity limit of silicon in the near-IR (F96; for a description of the SDSS photometric camera, see Gunn et al. 1998). The sensitivity curves of the five bands are shown in Figure 1. We plot the total system throughput including the quantum efficiency of the CCD, the throughput of the telescope and camera optics, and the average atmospheric extinction at an air mass of 1.2 for the altitude of Apache Point Observatory (2800 m). We also show the low-resolution spectra (in the units of f_λ , arbitrary normalization) of a quasar with redshift of 2.8 and a normal F5 V type star (for the star/quasar separation, see § 6).

The zero point of the SDSS photometric system is based on the AB₉₅ system defined in F96 (see also Oke & Gunn 1983). The magnitude is defined by

$$m = -2.5 \log \frac{\int d(\log \nu) f_\nu S_\nu}{\int d(\log \nu) S_\nu} - 48.60, \quad (\text{A1})$$

where S_ν is the system throughput. They are based on the observations of four F-subdwarf spectrophotometric standards of Oke & Gunn (1983). Their synthetic SDSS magnitudes are listed in F96. The SDSS photometric system will be defined by observations with the SDSS "Photometric Telescope," a 50 cm reflector located at Apache Point Observatory, during the commissioning phase of the survey.

The basic properties of each band are listed in Table 2. The effective wavelength λ_{eff} listed in Table 2 is defined by Schneider, Gunn, & Hoessel (1983):

$$\lambda_{\text{eff}} = \exp \frac{\int d(\ln \nu) S_\nu \ln \lambda}{\int d(\ln \nu) S_\nu}, \quad (\text{A2})$$

which is in some sense halfway between an effective wavelength and an effective frequency (see also F96).

TABLE 2
FILTER PROPERTIES

Filter	u'	g'	r'	i'	z'
λ_{eff} (Å)	3543	4770	6231	7625	9134
FWHM (Å)	567	1387	1373	1526	950
Q	0.0185	0.118	0.117	0.0874	0.0223
$t \times (\text{sky} + \text{dark}) \text{ pixel}^{-1}$	45	401	690	1190	1120
$m_{\text{lim}}(5 \sigma)$	22.3	23.3	23.1	22.7	20.8
A_i/A_V	1.593	1.199	0.858	0.639	0.459

We need to simulate the photometric errors as observed by the real survey in order to compare the simulation with the observations. The photometric errors have two contributions: photon noise (from the star, sky, dark current, and read noise) and calibration error (how well the magnitudes are tied to the primary standards; F96). For an object with a noiseless magnitude in band i as m_i^{real} , we apply the following procedures to get its observed magnitude m_i^{obs} and associated photometric error σ_i^{obs} :

1. The total mean number of electrons detected will be

$$\langle S_i \rangle = 1.96 \times 10^{11} \times t \times Q_i \times 10^{-0.4m_i^{\text{real}}}, \quad (\text{A3})$$

where t is the total exposure time (55 s); Q_i is the flux sensitivity quantity defined in F96 (see Table 2):

$$Q_i = \int d(\ln \nu) T_v^i; \quad (\text{A4})$$

and T_v^i is the total system throughput including CCD, telescope, and atmosphere.

2. The number of observed electrons S_i is drawn from a Poisson distribution $P[\langle S_i \rangle]$.
3. The total photon noise will be

$$N_i = [S_i + RN^2 + n_{\text{eff}} \times t \times (\text{sky} + \text{dark})_i]^{1/2}, \quad (\text{A5})$$

where RN is the read noise of the CCD (close to $7 e^-$; Gunn et al. 1998); $t \times (\text{sky} + \text{dark})_i$ is the total sky and dark current counts per pixel; and n_{eff} is the effective number of pixels used for PSF photometry (for the normal seeing condition in APO, $n_{\text{eff}} = 28$ pixel, 1 pixel = $0''.4$).

4. The magnitude and error without calibration error are

$$m_i^{\text{nocal}} = -2.5 \log [S_i / (1.96 \times 10^{11} \times t \times Q_i)] \quad (\text{A6})$$

$$\sigma_i^{\text{nocal}} = 1.086 N_i / S_i. \quad (\text{A7})$$

5. We assume a random calibration error $\sigma_i^{\text{cal}} = 0.02$ mag for all bands. A random $\Delta(m)$ from a Gaussian distribution is added to m_i^{cal} :

$$m_i^{\text{obs}} = m_i^{\text{nocal}} + \Delta(m) \quad (\text{A8})$$

$$\sigma_i^{\text{obs}} = [(\sigma_i^{\text{nocal}})^2 + (\sigma_i^{\text{cal}})^2]^{1/2}. \quad (\text{A9})$$

APPENDIX B

SIMULATION OF H I ABSORPTION SPECTRUM OF HIGH-REDSHIFT QUASAR

We create the synthetic spectrum of absorptions by intervening H I absorbers along the line of sight to high-redshift quasars, in the spectral range blueward of Ly α emission. The absorption systems we consider are Ly α forest systems, Lyman limit systems (LLSs), and damped Ly α systems. We simulate both the line absorption (up to Ly $_{10}$) and continuum absorption by those systems. The simulation follows procedures similar to those outlined by Møller & Jakobsen (1990), Fan & Chen (1993), and Warren et al. (1994).

An H I absorber is characterized by three parameters: its H I column density N_{HI} , redshift z , and Doppler width b . The number density of the absorption systems evolves with redshift. The evolution is usually fitted by a power law:

$$N(z) = N_0(1 + z)^\gamma. \quad (\text{B1})$$

And the H I column density distribution is also usually expressed as a power law:

$$f(N_{\text{HI}}) \propto N_{\text{HI}}^{-\beta} \quad (\text{B2})$$

The redshift evolution of the Ly α forest system is well established. On the other hand, the results on the evolution of LLSs and damped systems remain controversial. Sargent et al. (1989) found $\gamma = 0.68$, consistent with no evolution in comoving density for LLSs, while Lanzetta (1991) found strong evolution for $z > 2.5$ ($\gamma = 3.5$). Storrie-Lombardi et al. (1994) found an evolution rate somewhere in between ($\gamma = 1.55$). The evolution rate of LLSs is important for the quasar colors, since the optically thick continuum absorption has the biggest effect on the colors of high-redshift quasars. Storrie-Lombardi et al. (1994) used the APM sample, which includes a large number of quasars at $z > 4$, so we will use their statistics on γ and β for the LLSs and damped systems. The statistics we adopt in the simulation of H I absorbers are summarized in Table 3.

TABLE 3
PROPERTIES OF H I ABSORBERS

Parameter	$\log N_{\text{H I}}(\text{cm}^{-2})$	N_0	γ	β	b (km s $^{-1}$)	References
Ly α forest	13–17.3	50.3	2.3	1.41	30	1
LLS	17.3–20.5	0.27	1.55	1.25	70	2
Damped Ly α system	20.5–22	0.04	1.3	1.48	70	2

REFERENCES.—(1) Møller & Jacobsen 1990; (2) Storrie-Lombardi et al. 1994.

The absorption spectrum is simulated as follows:

1. For a quasar with emission redshift z_{em} , the total number N of each kind of absorber along the line of sight is calculated by drawing randomly from a Poisson distribution $P(N_{\text{ave}})$, where N_{ave} is the average number expected: $N_{\text{ave}} = \int_0^{z_{\text{em}}} N(z) dz$.
2. For each of the N absorbers, its redshift and column density are randomly selected from the distribution according to equations (B1) and (B2), respectively.
3. For each absorber, the opacity profile τ_λ of the first 10 lines in the Lyman series are calculated. A Voigt profile with width b and natural broadening is assumed. The Lyman continuum absorption of each line under the Lyman limit is also calculated. Then, the line and continuum absorptions are redshifted, with the contribution on opacity from all of the absorbers co-added, and finally the absorption is applied to the emission line+continuum spectrum of the quasar. In this process, two important parameters, D_A and D_B , the continuum depression between Ly α and Ly β , and between Ly β and the Lyman limit, are also calculated:

$$D_i = \langle 1 - e^{\tau_\lambda} \rangle_\lambda, \quad \begin{cases} 1050 \text{ \AA} < \lambda < 1170 \text{ \AA}, & i = A \\ 920 \text{ \AA} < \lambda < 1015 \text{ \AA}, & i = B. \end{cases} \quad (\text{B3})$$

D_A and D_B provide a good check on how well we are simulating the overall level of absorption. Figure 22 plots our simulated distributions of D_A and D_B against observations. Both the average and scatter of the parameters agree very well between simulations and observations. The simulation thus can reproduce the effect on quasar colors and their scatter caused by the intervening absorptions.

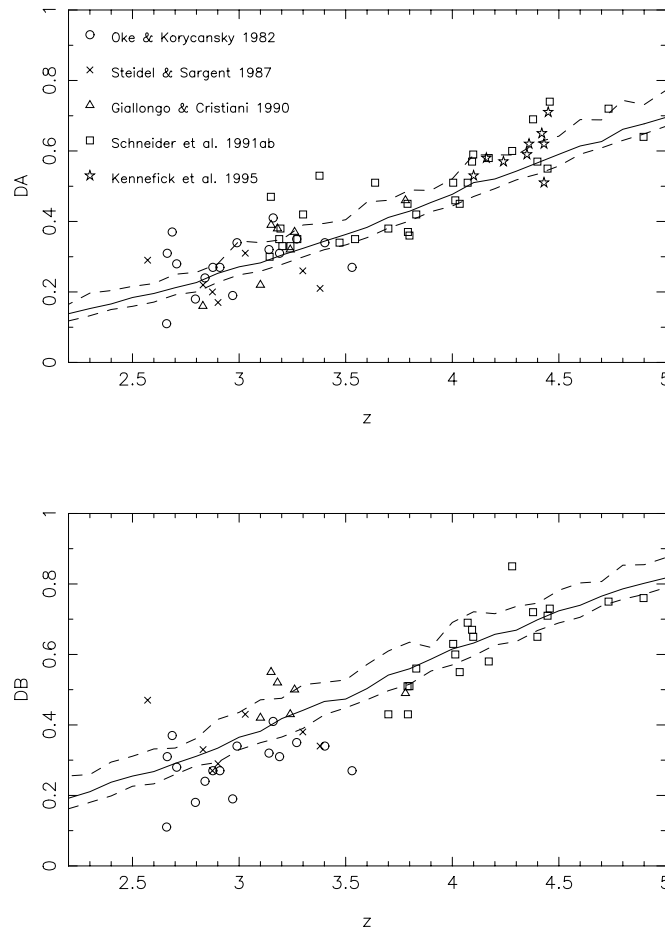


FIG. 22.—Lyman depression D_A and D_B as a function of redshift. Solid and dashed lines are the median and 68% scatter from the simulation.

REFERENCES

- Allard, F., & Hauschildt, P. H. 1995, *ApJ*, 445, 433
Bahcall, J. N. 1986, *ARA&A*, 24, 575
Bahcall, J. N., & Soneira, R. M. 1981, *ApJS*, 47, 357
———. 1984, *ApJS*, 55, 67 (BS)
Bergeron, P., Saffer, R. A., & Liebert, J. 1992, *ApJ*, 394, 228
Bergeron, P., Wesemael, F., & Beauchamp, A. 1995, *PASP*, 107, 1047
Bertelli, G., Bressan, A., Chiosi, C., Fagotta, F., & Nasi, E. 1994, *A&AS*, 106, 275
Boyle, B. J. 1989, *MNRAS*, 240, 533
Boyle, B. J., Shanks, T., & Peterson, B. A. 1988, *MNRAS*, 239, 935
Burstein, D., & Heiles, C. 1982, *AJ*, 87, 1165
Carraro, G., Ng, Y. K., & Portinari, L. 1998, *MNRAS*, 296, 1045
Cohen, M. 1993, *AJ*, 105, 1860
———. 1994, *AJ*, 107, 582
———. 1995, *ApJ*, 444, 874
Crampton, D. C., Cowley, A. P., & Hartwick, F. D. A. 1987, *ApJ*, 314, 129
Fan, X., & Chen, J. 1993, *A&A*, 277, L5
———. 1994, *Chinese Astron. Astrophys.*, 18, 393
Fan, X., et al. 1999, in preparation
Fluks, M. A., Plez, B., Thé, P. S., de Winter, D., Westerlund, B. F., & Steenman, H. C. 1994, *A&AS*, 105, 311
Francis, P. J. 1996, *Proc. Astron. Soc. Australia*, 13, 212
Francis, P. J., Hewett, P. C., Foltz, C. B., Chaffee, F. H., Weymann, R. J., & Morris, S. L. 1991, *ApJ*, 373, 465
Freeman, K. 1979, in *Photometry, Kinematics, and Dynamics of Galaxies*, ed. D. S. Evans (Austin: Dept. of Astronomy, Univ. of Texas at Austin), 85
Fukugita, M., Ichikawa, T., Gunn, J. E., Doi, M., Shimasaku, K., & Schneider, D. P. 1996, *AJ*, 111, 1748 (F96)
Giallongo, E., & Cristiani, S. 1990, *MNRAS*, 247, 696
Giallongo, E., & Trevese, D. 1990, *ApJ*, 353, 24
Gilmore, G. 1984, *MNRAS*, 207, 223
Greenstein, J. L., & Liebert, J. W. 1990, *ApJ*, 360, 662
Gunn, J. E., & Knapp, G. R. 1993, in *ASP Conf. Ser. 43, Sky Surveys: Protostars to Protogalaxies*, ed. B. T. Soifer (San Francisco: ASP), 267
Gunn, J. E., & Weinberg, D. H. 1995, in *Wide Field Spectroscopy and the Distant Universe*, ed. S. J. Maddox & A. Aragon-Salamanca (Singapore: World Scientific), 3
Gunn, J. E., et al. 1998, *AJ*, 116, 3040
Hall, P. B., Osmer, P. S., Green, R. F., Porter, A. C., & Warren, S. J. 1996, *ApJ*, 462, 614
Haywood, M., Robin, A. C., & Crézé, M. 1997a, *A&A*, 320, 428
———. 1997b, *A&A*, 320, 440
Haywood, M. 1994, *A&A*, 29, 129
Hartwick, F. D. A., & Schade, D. 1990, *ARA&A*, 28, 437
Hewett, P. C., et al. 1991, *ApJ*, 101, 1121
Horowitz, I. 1994, Ph.D. thesis, Caltech
Irwin, M. J., McMahon, R. G., & Hazard, C. 1991, in *ASP Conf. Ser. 21, The Space Distribution of Quasars*, ed. D. Crampton (San Francisco: ASP), 117
Kennefick, J. D., Djorgovski, S. G., & de Carvalho, R. R. 1995, *AJ*, 110, 2553
Koo, D. C., Bershad, M. A., Wirth, G. D., Stanford, S. A., & Majewski, S. R. 1994, *ApJ*, 427, L9
Koo, D. C., Kron, R. G., & Cudworth, K. M. 1986, *PASP*, 98, 285
Kriszunas, K., Margon, B., & Szkody, P. 1998, *PASP*, 110, 1342
Kroupa, P., Tout, C., & Gilmore, G. 1993, *MNRAS*, 262, 545
Lanzetta, K. M. 1991, *ApJ*, 375, 1
Larid, J. B., Rupen, M. P., Carney, B. W., & Latham, D. W. 1988, *AJ*, 96, 1908
Leitherer, C., et al. 1996, *PASP*, 108, 996
Lejeune, T., Buser, R., & Cuisinier, F. 1997a, *A&AS*, 125, 229
———. 1997b, *A&AS*, 130, 65
Lenz, D., Newberg, H. J., Rosner, R., Richards, G. T., & Stoughton, C. 1998, *ApJS*, 119, 121
Liebert, J., Dahn, C. C., & Monet, D. G. 1988, *ApJ*, 332, 891
Majewski, S. R. 1993, *ARA&A*, 31, 575
McQuade, K., Calzetti, D., & Kinney, A. L. 1995, *ApJS*, 97, 331
Møller, P., & Jakobsen, P. 1990, *A&A*, 228, 299
Newberg, H. J., Richards, G. T., Richmond, M., & Fan, X. 1999, *ApJS*, submitted
Newberg, H. J., & Yanny, B. 1997, *ApJS*, 113, 89
Ng, Y. K. 1994, Ph.D. thesis, Leiden Univ.
Ng, Y. K., Bertelli, G., Chiosi, C., & Bressan, A. 1997, *A&A*, 324, 65
Oke, J. B., & Gunn, J. E. 1983, *ApJ*, 266, 713
Oke, J. B., & Korycansky, D. G. 1982, *ApJ*, 255, 11
Osmer, P. S., Kennefick, J. D., Hall, P. B., & Green, R. F. 1998, *ApJS*, 119, 1890
Oswalt, T., Smith, J. A., Wood, M. A., & Hintzen, P. 1996, *Nature*, 382, 692
Pei, Y. C. 1995, *ApJ*, 438, 623
Prandoni, L., et al. 1998, *A&A* submitted (astro-ph/9807153)
Rana, N. C. 1991, *ARA&A*, 29, 129
Rana, N. C., & Basu, S. 1992, *A&A*, 265, 499
Reid, N., & Majewski, S. R. 1993, *ApJ*, 407, 74
Richards, G. T., Yanny, B., Annis, J., Newberg, H. J., McKay, T., York, D. G., & Fan, X. 1997, *PASP*, 107, 39
Robin, A. C., & Crézé, M. 1986, *A&A*, 157, 71
Salzer, J. J. 1989, *ApJ*, 347, 152
Salzer, J. J., MacAlpine, G. M., & Boroson, T. A. 1989, *AJ*, 70, 479
Sargent, W. L. W., Steidel, C. C., & Boksenberg, A. 1989, *ApJS*, 63, 703
Scalo, J. 1998, in *ASP Conf. Proc. 142, The Stellar Initial Mass Function: Proc. of the 38th Herstmonceux Conf.*, ed. G. Gilmore & D. Howell (San Francisco: ASP), 201
Schmidt, M., & Green, R. F. 1983, *ApJ*, 269, 352
Schmidt, M., Schneider, D. P., & Gunn, J. E. 1995, *AJ*, 110, 68
Schneider, D. P., Gunn, J. E., & Hoessel, J. G. 1983, *ApJ*, 264, 337
Schneider, D. P., Schmidt, M., & Gunn, J. E. 1991a, *AJ*, 101, 2004
———. 1991b, *AJ*, 102, 83
———. 1994, *AJ*, 107, 1245 (SSG)
Storrie-Lombardi, L. J., & Irwin, M. J. 1994, *ApJ*, 427, L13
Storrie-Lombardi, L. J., McMahon, R. G., Irwin, M. J., & Hazard, C. 1996, *ApJ*, 468, 121
Steidel, C. C., & Sargent, W. L. W. 1987, *ApJ*, 313, 171
VandenBerg, D. A. 1985, *ApJS*, 58, 711
Wainscoat, R. J., Cohen, M., Volk, K., Walker, H. J., & Schwartz, D. E. 1992, *ApJS*, 83, 111
Warren, S. J., Hewett, P., & Osmer, P. S. 1994, *ApJ*, 421, 412
Weidemann, V. 1995, in *White Dwarfs: Proc. of the 9th European Workshop on White Dwarfs*, ed. D. Koester & K. Werner (Berlin: Springer), 1
Wilkes, B. J. 1986, *MNRAS*, 218, 331
Wood, M. A. 1995, in *White Dwarfs: Proc. of the 9th European Workshop on White Dwarfs*, ed. D. Koester & K. Werner (Berlin: Springer), 41