



Fermi National Accelerator Laboratory

FERMILAB-TM-1973

Quasar Target Selection Fiber Efficiency

Heidi Newberg and Brian Yanny

*Fermi National Accelerator Laboratory
P.O. Box 500, Batavia, Illinois 60510*

May 1996

Disclaimer

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Quasar Target Selection Fiber Efficiency

Heidi Newberg Brian Yanny

May 14, 1996

Fermilab Technical Memo TM-1973

Abstract

We present estimates of the efficiency for finding QSOs as a function of limiting magnitude and galactic latitude. From these estimates, we have formulated a target selection strategy that should net 80,000 QSOs in the north galactic cap with an average of 70 fibers per plate, not including fibers reserved for high-redshift quasars. With this plan, we expect 54% of the targets to be QSOs. The North Galactic Cap is divided into two zones of high and low stellar density. We use about five times as many fibers for QSO candidates in the half of the survey with the lower stellar density as we use in the half with higher stellar density. The current plan assigns 15% of the fibers to FIRST radio sources; if these are not available, those fibers would be allocated to lower probability QSO sources, dropping the total number of QSOs by a small factor (5%). We will find about 17,000 additional quasars in the southern strips, and maybe a few more at very high redshift. Use was made of two data sets: the star and quasar simulated test data generated by Don Schneider, and the data from UJFN plate surveys by Koo (1986) and Kron (1980). This data was compared to results from the Palomar-Green Survey and a recent survey by Pat Osmer and collaborators.

1 Introduction

The SDSS quasar working group is charged with the task of determining a quasar target selection algorithm to be used throughout the duration of the 5 year survey.

The algorithm in the northern cap must be based primarily on the magnitudes, colors, and radial profile information (point-like or not) obtained by the photometric imaging part of the SDSS survey. Prior information about the distribution of magnitudes, colors and radial profiles of stars, galaxies and quasars from other astronomical sources may also be used. In addition, selection based on independent sources such as radio or X-ray flux catalogs can be considered.

Top consideration must be given to the scientific goals of the quasar part of the survey. These goals include: gaining a better understanding of the quasar luminosity function at all redshifts, investigating quasar-quasar and quasar-galaxy clustering, and studying quasar absorption lines and intrinsic properties of quasars.

The founding documents of the survey estimate that approximately 100,000 quasars will be detected in the survey, based on the average numbers of quasars known per square degree of sky to a given magnitude limit. Two crucial factors in determining whether or not 100,000 QSOs can be confirmed are: 1) the detection limit of the spectrograph for approximately 1 hour integrations, and 2) the fraction of quasar targets which turn out to actually be quasars (selection efficiency).

All large surveys must reconcile the competing goals of ‘statistical validity’ and ‘exploration’. A large survey must be defined well enough so that its results on detections and non-detections are statistically significant and quantifiable. At the same time, a survey should be designed with serendipity in mind so that new phenomenon, or phenomena which could not be seen in previous smaller or less sensitive surveys, are not excluded from being discovered in the larger survey.

In order to understand the statistical significance of the luminosity function and quasar clustering, we must know the the selection function for all areas of the sky that were searched for quasars. We would have an unacceptably low fiber efficiency if we tried to find a $> 90\%$ complete magnitude-limited sample of quasars at $g \sim 19.5$. The goal of counting the quasars at all redshifts (especially at high redshifts, $z > 4$) and the goal of detecting any remaining undiscovered bright QSOs ($g < 18$) imply that one may have to accept a lower selection efficiency than one could obtain by only looking for the most easily identified (and best studied) QSOs with $z \sim 2$.

Finally, the quasar section of the survey must not interfere excessively with the completeness of the main galaxy spectroscopic redshift survey. This means that the total number of fibers allocated to quasars should not exceed, on average, about 15% of the total fibers available to galaxies and that the distribution of numbers of fibers allocated to QSOs should not change excessively over the area of the survey in a way that affects galaxy detection completeness.

We do not calculate the efficiency of the specialized survey for quasars with $z > 5$. This search involves very few fibers per plate, and efficiency concerns are greatly outweighed by the possibility of even a handful of detections. We may learn something interesting even if these very rare objects do not turn out to be quasars. We require that the search cover a large fraction of the sky, but it is not necessary that this part of survey be uniform.

All of these issues go into the final quasar target selection algorithm. We investigate below the known information about quasar target selection based on color and other attributes. We extrapolate this data to the instrumentation of the SDSS to estimate more accurately how efficient various selection algorithms will be. We then propose a quasar selection algorithm for the SDSS to be refined during the test year.

2 Other QSO surveys

We list in Table 1 a summary of efficiencies obtained in other QSO surveys which have used the techniques of either ultra-violet color excess (UVx) or multi-color selection.

The best efficiency anyone in the literature has claimed to find is about 70%-80%, but

these efficiencies are either unconfirmed (Gaidos et al. 1993), or used a color space cut that was determined after a lower efficiency spectroscopic survey had already been completed (Warren et al. 1991b; Kron et al. 1995).

Table 1

Summary of QSO selection efficiency from other UVx and multi-color surveys

Method	Area	limit	candidates	QSOs	efficiency	ref
UVx	10700	$B < 16.1$	1715	93	5.4%	Green et al. 1986
UVx	20	$B < 21$	1400	420	30%	Boyle et al. 1990
ubvroi	46	$r < 20$	473	130	27%(58% ^a)	Warren et al. 1991
UVx	10	$B < 20$	200	97	50%	LaFranca et al. 1992
BVRI	0.33	$B < 22$	25	20?	80% ^b	Gaidos et al. 1993
UJFN	1.5	$J < 22$	188	130	70% ^c	Kron et al. 1995
ugriz	10000	$g < 20$	186000	100000	54% ^d	SDSS est

^aThe higher figure is from more conservative color space cuts.

^bConfirming spectra have not been obtained, 80% is a guess.

^cThis is based on after the fact separation based on d_5 and d_{locus} .

^dEstimate based on this work.

We conservatively estimate below that if we give consideration to all issues raised in the introduction, then overall fiber efficiency for confirming QSO candidates in the SDSS survey will be in the neighborhood of 54%.

We will investigate simulated test data and three of the data sets listed in Table 1 and attempt to justify the 54% efficiency expectation for the SDSS.

3 Simulated Test Data

The test data is in a first-draft state. It contains only ‘normal’ spheroid and disk stars and simulated quasars, with three times as many quasars per square degree as there are in the real sky. Two patches of test data were generated, one at $l=0$, $b=90$, and the other at $l=0$, $b=30$. We believe this simulated data will be most important for determining where the QSO locus is relative to the stellar locus for regions outside the ‘UV excess’ part of color space (that is, in regions with $z > 2$). There are not enough known quasars at high redshift to study their color-space distribution in detail using only existing data, so the simulations are important.

3.1 Description of Simulated Star Colors

The stars were generated by using the B-V colors and star densities from the Bahcall-Soneira (1980) model. The Gunn-Stryker (1983) atlas was used to transform B-V to the SDSS colors. The catalog was restricted to stars with $i < 20$.

3.2 Description of Simulated QSO Colors

The quasars are model quasars. The quasar spectrum is described by a set of parameters such as continuum slope, strength of ly-alpha emission relative to continuum and other emission lines, redshift, absolute luminosity, and the depth of the ly-alpha depression. Each parameter was drawn from the observed distribution of this parameter, and the effects of the ly-alpha forest were included.

3.3 Objects not yet modeled

Some types of objects which are expected to be confused with quasars on the basis of colors alone are not yet included in the simulations. They include:

- white dwarfs
- sdB, sdO blue sub-dwarfs
- odd metallicity stars
- binary stars
- compact blue galaxies
- non-Gaussian errors in photometry of ordinary stars.

3.4 Visualization of the survey data

We have examined both 2-D and 3-D projections of the multi-dimensional data. In general, we believe that the 3-D projection, under the correct rotation, gives more insight into how to best separate QSOs from stars than looking at specific 2-D projections along one color axis or another. If one performed a principal component analysis of the data (such as Dawn Lenz has done for simulated stellar colors), then one can determine a projection along some linear combination of color axes which serves to best separate the stars from the QSOs in 2 dimensions. Since it is not clear a-priori how to best do this principal component breakdown, the visual insight from our 3-D representation is useful. Are three dimensions sufficient? What about 4-D or higher dimensional representations? The data, in fact, have at least 5 and perhaps 6 easily identifiable dimensions. The dimensions are: magnitude, u-g, g-r, r-i, i-z and fwhm. We expect the i-z dimension to be redundant with the r-i dimension for all but the very highest redshift quasars. In this work, we will not consider extended objects with unusual colors as QSO candidates. We do not rule out the possibility of adding it later if tests indicate that there are not too many of these targets. The magnitude dimension is an additional dimension that needs to be considered if the stellar or quasar locus changes significantly with magnitude. Indications are that the stellar locus *does* change significantly with magnitude. Unfortunately, our available datasets are not yet large enough to quantify magnitude dependence of the stellar locus. Since it is difficult to visualize 4-D data, we propose to visualize the data in several 3-D cubes, each covering a different magnitude range, to explore the ‘magnitude’ dimension.

3.5 Results

The simulated data were plotted and examined by eye. Figure 1 shows a (u-g, g-r, r-i) color-color-color plot of the test data for $l=0$, $b=90$. Figure 2 shows the same for $l=0$, $b=30$. Keep in mind that the model has three times as many quasars per square degree as we can expect to see on the sky. One has the impression from these plots that the quasars are extremely well separated from the stars, though it is difficult to see the extent of it in the 2-D projection.

The distribution of stars and QSOs in color space can also be seen in a ‘color cube’ representation. Color space was sub-divided into cubes 0.5 magnitudes on a side. With larger data sets one can use smaller sub-cubes of 0.05 to 0.1 mags on a side. The color cubes were then filled by counting the number of quasars in each small cube and dividing by the total number of objects in each small cube (stars + quasars). Figures 3 and 4 show the result for the $l=0$, $b=90$ and $b=30$ simulated data respectively. The cyan cubes contain mostly quasars, the magenta cubes contain all stars, and other colors are in between. We also calculated the number of stars and quasars that were in boxes of each color. If we had selected as targets anything that fell in a cyan box, we would have found 91% of the quasars with 100 percent efficiency. At low galactic latitude, we would have found 91% of the quasars with 99% efficiency. These are, of course, strict upper limits on the efficiency for finding quasars.

One very important observation from this test data is that no quasars are expected to be located in the very red part of the color-cube, that part of the diagram is populated by a wide variety of M-dwarf stars and appears to be void of QSOs (at least those with $z < 5$). We propose then to exclude this region of color space from QSO target selection (high z targets excepted). This will significantly improve selection efficiency, with the possible drawback of missing the discovery of a class of point-like extra-galactic objects in this part of color space.

4 The Koo-Kron data

4.1 The data

We are grateful to have been given access to a partially published body of data obtained over the last 15 years by Koo, Kron, Munn, Majewski, Bershad, and Smetanka. (hereafter KK data; see Kron 1980, Koo 1986, Kron et al. 1995). It is based on about 1.5 square degrees of (KPNO 4-meter telescope, 1-hour exposure) UJFN plates (UJFN filters correspond approximately to SDSS ugri filters, respectively, see Appendix A) in four regions of the sky which sample a variety of galactic locations (l, b). The data are more or less complete and unsaturated in the $21.5 > J > 19.5$ range. Four-band photometry exists for approximately 2000 objects in this magnitude range, and over 200 3\AA resolution spectra covering (4000-7000 \AA) have been obtained for objects which fell outside the stellar locus. Types have been assigned to objects based on these spectra.

4.2 Results

The UJFN data were plotted and examined visually in 3-space. Figure 5 shows the KK data in color-color-color space. The plots do give a good impression what the real data will look like. The quasars are still clustered in the UVx region, but there are many more contaminating objects than exist in present simulations. We find that the contaminating objects are objects of unusual color, and are not likely to move back into the stellar locus with the more accurate colors expected from the Sloan survey.

Figure 6 shows the position of galaxies in color-color-color space. While we expect to be able to remove galaxies as QSO targets through use of a fwhm indicator, there will be some contamination, especially at the faintest magnitudes.

Figure 7 shows the color cube plot (cubes are 0.5 mag on a side) for this data. If we had selected as targets anything that fell in a cyan box, we would have found 23% of the quasars with 100% efficiency. There are two competing traps one can fall into doing this kind of analysis. If the boxes are too small, then each object will essentially have its own box; we would find 100% of the objects with 100% efficiency. If the boxes are too large, the populations would not be separated from each other; we would find few objects, or find many objects with tiny efficiency. Although this technique is interesting for visualization, we will abandon it for estimations of our quasar finding efficiencies.

Comparing with the simulated test data, we note that the main stellar loci are in excellent agreement. In areas of the color space where UV excess quasars are found, there is naturally more contamination of the QSOs by unusually colored stars and unknown objects in the KK data than in the simulations.

4.3 UVx Quasars

In the magnitude range of the Sloan survey, we expect to find about 17 QSOs per square degree. The majority of these quasars will lie in the ultraviolet excess (UVx) region of color space with blue $u - g$ and $g - r$ colors. We therefore wanted to know in more detail what objects were actually found in this region of the real data. We selected the 116 objects below the plane $0.72(U-F) + 0.62(J-F) + 0.30(F-N) - 0.28 = 0$. Jeff Munn then gave us the most up-to-date (although not yet final) information on what the spectra tell us about these objects. Some fraction of the KK data spectral identifications may be incorrect, however, a very conservative QSO identification criteria was chosen and nearly all objects tagged as QSOs really are QSOs. The spectra could be divided into the following six categories:

Table 2

UVx Objects identified in KK Survey

Type	number	Spectroscopic feature identified
Quasar	72	At least one solid broad line or two weaker broad lines
Galaxy	5	Primarily narrow emission line spectra
Star	6	Show emission/absorption lines at zero redshift. We included some stars with rather weak line if there were several of them.
White Dwarf	3	Showed broad absorption at H-alpha or H-beta
Not Quasar	12	Have featureless low signal-to-noise spectra that have enough signal-to-noise to show QSO emission lines, but do not have them.
Unknown	18	Were not observed, or showed no signal at all.
Total	116	

Figure 8 shows these UVx selected objects in color space, with different objects marked with points of different color. The orientation is chosen to show that a further color cut could be made to exclude a few more non-QSO objects.

4.4 Selection Efficiency vs. Galactic Latitude

Michael Strauss has suggested that for the purposes of target selection, the Northern Cap survey area be separated into two regions, a ‘low density region at the NGP’ $b > 45$ and a somewhat higher density region ‘approximately an annulus’ with $30 < b < 45$. The Galactic stellar density is significantly greater in the lower latitude region and scatter of stars from the stellar locus can significantly reduce selection efficiency for QSO confirmation. The division into two zones allows one to concentrate fibers for QSO selection in the NGP low stellar density region (with proportionally fewer fibers in the high stellar density region) and obtain more successful QSO confirmations than one would obtain by allocating fibers evenly over both regions. Heidi Newberg has suggested a modification to this technique that divides the survey area into the half with lower stellar density (survey center) and the half with higher stellar density (skirt). The star density in the survey center varies by a factor of two from minimum to maximum and takes into account the Galactic bulge which affects stellar density as a function of longitude. The stellar density in the skirt ranges from twice the Galactic Cap minimum to about 9 times the minimum. Figure 9 shows the stellar density as a function of (l,b) and locates the survey northern cap ellipse.

We can use the data from the KK survey to estimate the overall efficiency for finding QSOs in the galactic cap (for comparison with the Palomar-Green survey) and the separate efficiencies for the survey center and skirt. First, we adjust the data to account for the higher galactic latitude and lower average stellar density in their fields. Three of the four KK fields are in the less dense half of the cap, and one is in the more dense half of the cap. If we separately determine the efficiencies in these two regions and then average them, we expect to target QSOs 70% of the time in the magnitude range $21.5 > J > 19.5$. To determine this, we have ignored objects of unknown spectrum.

We expect the stellar contamination to be somewhat higher at low galactic latitude than at high galactic latitude. As expected, the three fields which lie in the survey center have about the same QSO density as the remaining field that lies in the skirt, but this low galactic latitude field has about 41% more UVx objects per square degree. We calculate from these numbers that the efficiency for finding UVx QSOs is about 82% in the center, and 58% in the skirt. It is interesting that the density of UVx stars increases at about the same rate as the stellar density from the Bahcall-Soneira model as one goes to lower galactic latitudes. Depending on what we call a star, there are between two and four times as many stars in the skirt field as in an average field in the survey center, and the stellar density in the lower field is three times as high as the fields at high galactic latitude.

5 Warren et al. Survey Data

Pat Osmer and collaborators have done a survey similar to the Koo-Kron survey (Warren et al. 1991a;1991b). The Warren et al. survey uses 6 filter bands (u b v r o i) and has magnitude errors of about 0.1 mag. The region of sky surveyed was 47 sq. degrees. Typical seeing was 1.5 arcsec. In discussions with Osmer about his data, several interesting points came up. First, separating QSO candidates from galaxies was not a big problem, and, in fact, NELG (narrow line emission galaxies) appeared to be well separated from QSOs in color space using u-b and b-v as discriminant colors. Second, there always appeared to be a significant number of ‘interloper’ stars which had colors very similar to that of QSOs in all bands for the region of color space where the bulk of the QSOs are located. The Warren et al survey is distinct from UVx surveys in that it attempts to find QSOs at higher redshift ($z > 2.2$) and specifically excludes the UVx region of color space from spectroscopic followup.

What we find in real surveys is that the contamination in the quasar locus (which makes it virtually impossible to target quasars with better than 70-80% efficiency) is caused by unusual objects that were not included in the simulations. They are white dwarves, other unusual stars, emission line galaxies, and objects with featureless spectra. Although some have argued that we would have very good efficiency for finding quasars because our colors will be more accurate than other surveys, it is probably not true.

6 Palomar-Green Survey

To meet the scientific goals of the quasar working group, we must detect ‘essentially all bright ($B < 18$) QSOs’. To determine the numbers of bright QSOs and numbers of bright objects with colors similar to QSOs we refer to the Palomar-Green (1986) survey for reference.

Since our quasar sample will be somewhat brighter than that of the KK data, we will interpolate between the KK data, which reaches $B \sim 22$, and the Palomar-Green Catalog of Ultraviolet Excess Stellar Objects, which has a limiting magnitude of $B = 16.2$. In the Palomar Green survey, 1715 objects were found in 10700 square degrees of the North Galactic Cap. The percentage of each type of object is given in Table 3.

Table 3

Percentage of Each Type of Object in the Palomar-Green Survey

Type	Percentage
Quasar	5.4
Seyfert	1.3
BL Lac	0.2
Galaxy	2.0
sdB Star	39.9
sdO Star	13.4
Other Star	6.1
White Dwarf (DA)	21.0
Other White Dwarf	5.1
Planetary Nebula	0.5
Var Star	5.0
Unknown	0.3

Table 4 shows these numbers of objects per square degree extrapolated from the counts at 16 magnitude and $d(\log N)/dB$ from Green et al.

6.1 Scale height of populations of objects

One notices immediately from Table 4 that the numbers of objects for different populations of objects (QSOs vs. field stars vs. WDs) grow at different rates with magnitude.

We analyze this as follows: Let's assume that the luminosity function for a given type of star does not change as a function of how far it is away. Then, the slope $d(\log N)/dB$ should be 0.6, independent of what the luminosity function is. We expect that stellar distributions will decrease as we go away from the center of the galaxy, so as we look at fainter stars in the Galactic Cap, the slope should fall off. The Palomar-Green survey is consistent with a uniform distribution of white dwarfs; if white dwarfs are typically $B = 16$, this shows that the scale height of white dwarfs is not sampled at about 10pc (which is expected). Since the slope for the sdB and sdO stars is shallower at $B = 16$, we expect that we are sampling its scale height. It is difficult to determine what that scale height is, since it depends on the absolute magnitude and luminosity function of these stars. The steep slope for quasars (greater than Euclidean) shows that the QSO population is changing with survey depth, for instance, it is a population at a (approximately fixed) very large distance and the observed slope samples the intrinsic slope of the luminosity function.

Although we know that the slope for each type of object should decrease at some rate, we do not have enough information to predict what this rate will be. Therefore, we extrapolate the numbers linearly. The rate has been investigated by others for QSOs and appears to flatten somewhat at about $B = 19.5$. It is unclear whether this will cause us to overestimate or underestimate our efficiency for finding quasars, since it depends on which types of objects fall off faster. It is reassuring that the extrapolation is reasonably consistent with the KK

results (see Table 5). If anything, we have overestimated the efficiency for finding quasars in comparison. The KK results (and the presumed scale height of white dwarfs) suggest that we have overestimated the fraction of white dwarfs and underestimated the fraction of galactic stars at 20th magnitude.

Table 4

Number Densities (per square degree) of UVx Objects

Type	$\frac{d(\log N)}{dB}$	N(B<16)	N(B<18)	N(B<18.5)	N(B<19)	N(B<19.5)	N(B<20)
sdB	0.32	0.049	0.21	0.31	0.45	0.65	0.93
sdO	0.35	0.014	0.07	0.10	0.16	0.24	0.35
WD	0.57	0.027	0.37	0.72	1.38	2.67	5.14
QSO	0.90	0.0054	0.34	0.96	2.71	7.63	21.50
Total		0.144	0.99	2.09	4.70	11.19	27.92

Notice how the number of QSOs rises faster than the number of stars or WDs as one goes to fainter mags

7 QSO finding efficiency estimates

In Table 5 we show the fraction of the UVx targets that will be QSOs. The first and last data columns show results from the surveys, and the center three columns show our extrapolations, assuming that 10% of all objects are galaxies or stars that are not included in the extrapolation. By further breaking down the figures for low and high galactic latitude regions, we estimate the QSO fiber efficiencies for UVx candidates in Table 6. These were calculated using the relative efficiencies between low and high galactic latitude in the Koo-Kron data.

There is less information available on how well we can separate non-UVx quasars from the stellar locus. To estimate this, we counted the number of outliers we could see in the Koo-Kron data, and noted how many of them were QSOs. There were about seven quasars, and about twenty-some things that were similarly separated from the stellar locus. So, we estimate this efficiency to be about 30% at high galactic latitude to $B < 20$. For all other magnitude limits and latitudes, we scale the efficiency with the UVx efficiency. These numbers assume that we don't look at outliers near the red end, and that otherwise we do not look specifically in the quasar locus.

We can get a better grasp of the problem from the Warren et al. (1991b) search, which has a limiting magnitude of $m_{or} = 20$. They surveyed two fields. One field is at low stellar density (SGP), like our survey center, and the other is in our survey skirt (F401). We consider only the high redshift quasars ($z > 2.2$) that are found as outliers from the stellar locus; we assume that the low redshift QSOs would be in our UVx region. They give results for the cut they actually used and the one for which they have spectroscopy for every object. For the latter color-space cut, they found 67 quasars with 28% efficiency in the SGP, and 33 quasars

with 16% efficiency in F401. These numbers are very similar to our estimates from the KK data. For the higher (complete) threshold, they found 46 quasars with 53% efficiency in SGP, and 25 QSOs with 30% efficiency in F401. It is possible that we could increase our efficiency for finding quasars in color space by a factor of two over the estimates we will use in the next section; it depends how close to the stellar locus we wish to cut color space in selecting QSO targets.

Table 5

Estimated efficiency for finding UVx objects by type (%)

Type	PG							KK
	B<16.2	B<16	B<18	B<18.5	B<19	B<19.5	B<20	B<21.5
sdB	40.	46.	19.	13.	9.	5.	3.	-
sdO	13.	13.	6.	4.	3.	2.	1.	-
WD(DA)	21.	26.	34.	34.	26.	21.	17.	>=4.
Quasar	5.	5.	31.	41.	52.	61.	69.	70.
Other	21.	10.	10.	10.	10.	10.	10.	>=4.

Table 6

Estimated UVx QSO fiber efficiency by Galactic Latitude Zone (%)

Galactic Latitude	B<16	B<18	B<18.5	B<19	B<19.5	B<20
high	7.	36.	48.	61.	71.	81.
low	4.	26.	34.	43.	51.	57.

8 Proposal for QSO target selection

8.1 The Proposal

Based on the limited data in our possession, we propose a plan that will allow us to accomplish all of our science goals. Here we set broad goals for the quasar target selection algorithm which will allow us to judge the success of an algorithm against reasonable expectations. We will discuss the more specific algorithm in the next paper. Since all of our efficiencies were estimated using limiting magnitudes in B, we assume that we will also use a B limiting magnitude. In practice, we may decide to set the magnitude limit some other way; we note that for objects with QSO colors, $B \sim g + 0.3$. Our proposal assumes that we can allocate more fibers in the central region, that it is necessary to find all quasars brighter than B=18th magnitude, that it is not required to find all 100,000 quasars in the North Galactic Cap, and that we will use on average 70 fibers per plate. We will discuss afterwards the consequences of changing some of these assumptions.

We will do all of our calculations per square degree. We assume that we should average 70 fibers per plate, where each plate subtends 7.1 square degrees on the sky. The way Jon

Loveday and the tiling algorithm currently calculate this, we have about 80 fibers per plate with 87% efficiency for assigning fibers to quasars, and then the 13% left over is assigned to non-tiled targets. According to the tiling simulations, the plates will overlap by 48% on average. This will give us $70 * 1.48 / 7.1 = 14.6$ fibers per square degree. Our proposal attempts to do as much science as possible with about that average density of fibers.

We propose to use a limiting magnitude of about $B=20$ in the central half (5000 square degrees) of the survey, defined by the locus of points at which the Bahcall-Soneira model gives us twice the minimum stellar density. The other 5000 square degrees will be searched for the bright ($B < 18$) quasars that are of great use for absorption line studies.

We divide the quasars up into four groups: the UVx quasars, the quasars that are not UVx but are reasonably far from the stellar locus, the radio sources, and the quasars that cannot be detected with better than 30% efficiency due to the high density of stars:

- The UVx quasars will give us the highest fiber efficiency, and are numerous enough that we believe we can complete a uniform survey of them across the sky.
- We hope we will be able to do a uniform survey of color space, but we are not certain that this will be possible; it may be difficult to define an area of color space that will have a high density of quasars (and guarantee that these targets will not consume too many fibers) before the entire survey is complete. We are concerned that we must set a threshold for several-sigma events, and that small differences in data quality or position in the sky could swamp the QSO fibers. Mapping the appropriate parts of color space as a function of galactic latitude and longitude could be a big job.
- We are planning to observe radio sources that are detected as point sources brighter than 20th magnitude in the Sloan survey. This is of course subject to solving logistical problems, and assuming our estimate of how many fibers this would require is not too far off.
- We consider the quasars very near or in the stellar locus to be unfindable, except for some radio sources. We hope that radio sources will help to locate QSOs at problematic redshifts around $z = 3.2$ where the QSO locus intersects the stellar locus in all colors.

The grey book states that at magnitude $B=20$, there are 17 QSOs per square degree. From the KK data we estimate that about 11 of these are UVx, 2 are outliers, and 4 are undetectable. It is more difficult to estimate the radio sources. The FIRST search web page allows us to estimate the number of sources that fit our criteria. They see 110 sources per square degree. About 15% of them are brighter than 20th magnitude. Of those, about 65% are point-like. Multiplying this out, we expect about 10 sources per square degree that fit our criteria. The claim is that most of these are quasars. If 80% really are quasars, and if the radio-loud quasars are distributed evenly in color space, then we should already have targeted 6 of the 8 radio quasars per square degree. Therefore, we require an additional 4 fibers per square degree, and we expect two of them to be quasars. Table 7 shows the calculation of the number of fibers per square degree. We need about 24.3 fibers per square degree in the central section.

Table 7

Fiber use estimates for Finding Quasars

	UVx	Outliers	Radio Sources	Total
Survey Center				
QSOs/sq.deg.	11	2	2	15
efficiency	0.81	0.30	0.50	0.62
fibers/sq.deg.	13.6	6.7	4.0	24.3
Skirt region				
QSOs/sq.deg.	0.7	0.2	0.1	1.0
efficiency	0.26	0.10	0.20	0.20
fibers/sq.deg.	2.6	2.0	0.5	5.1

In the lower 50% of the survey, we expect about 1.0 quasar per square degree to be brighter than 18th magnitude. We used the same ratios between classes of quasars that we used for the central survey. Table 7 shows that we need about 5.1 fibers per square degree in the skirt.

Within the errors of the calculation, this plan uses up all of the fibers we were nominally allocated. It projects that we will find 80,000 quasars, with an overall efficiency of 54%. The majority of these (55,000) are UVx quasars in the central survey, and most of these will be 19-20 magnitude. Only 5000 quasars will be found in the skirt. Although this is quite a bit lower than original projections, it is the same as or better than people typically get from a color-selected sample. The projections show that we will find about 80% of the quasars above our magnitude limits.

8.2 Discussion of unbalanced allocation of fibers

One criticism of this plan is that it requires about 5 times as many fibers to be allocated to quasars in the central part of the survey as it allocates to the skirt. We would like to make sure that this does not jeopardize the survey time-to-completion or the measurements of the large scale distribution of matter. Jon Loveday has done some target selection simulations which indicate that there will not be much difference for the galaxies with this plan. He assumes that the QSOs are randomly distributed on the sky, and that the survey will be tiled in strips that are about ten degrees wide and 100 degrees long. He uses galaxies from the APM catalog at a magnitude limit which makes the density on the sky about the same as the Sloan Survey expects to see. Although at first it seems unbelievable that the QSO working group could demand on average 116 fibers per plate in the center and 23 fibers per plate in the skirt and not cause havoc, we can give a plausible argument of how this could be true. Over the 10,000 square degrees of the survey, Jon is tiling 2100 plates, which cover 14,800 square degrees. So, there is about a 48% overlap of the plates on average. By putting $24.3 - 14.6 = 9.7$ extra fibers per square degree in the center, we pull an extra 91 plates into the center, which leaves 91 fewer plates in the skirt. The overlap in the center is now 61%, and the overlap in the skirt is now 36%. On the average plate, there are fifteen percent more

fibers than there are galaxy and quasar targets. So, we would have to have a large positive fluctuation in galaxy counts for the quasars to start bumping fibers off of galaxies. We expect to get a larger fraction of the close pairs in the center than we will get in the skirt, but the overall efficiency will be about the same. Jon will continue to do more simulations, and the large scale structure working group will have to tell us whether this is a significant problem.

8.3 Can we get 100,000 QSOs?

Since the POO claims that we will get 100,000 quasars, it is interesting to see what we would have to do to fulfill this.

Our current best estimates for the North give 80,000 QSOs confirmed (see Table 7). The easiest way bring this number up to 100,000 is to pull them out of the southern survey. If we image three stripes in the south covering about 1000 square degrees, we look for quasars to 20th magnitude in these stripes, and we assume that we do not have radio sources available, then we expect to be able to find 13,000 quasars. We are not as concerned with fiber efficiency in the southern survey, since there will be ample time for spectroscopy. If one of the stripes is imaged multiple times and we take two hour exposures in this region rather than one hour exposures, then the grey book states that we should be able to find QSOs to $r = 19.5$ ($g = 20.5$). Since there are twice as many QSOs to $r = 19.5$ as there are to $B = 20$, we expect to find an extra 4000 quasars in this stripe. Adding this up, we get $80,000 + 17,000 = 97,000$ quasars altogether.

It is instructive to see what we would have to do to push the projected number of quasars in the North Galactic Cap close to 100,000. There are two ways of attacking this - either we request more fibers or we try to increase the efficiency.

Although we might be able to push the northern spectroscopic survey deeper than 20th magnitude, it is a bit risky. We will start to lose our ability to separate stars from galaxies, we may not be able to make accurate magnitude measurements in all five filters, and it will push the spectrograph near its limit. To find 100,000 quasars by requesting more fibers, we should push the limiting magnitude down to $B = 19$ in the skirt. At this magnitude, there are about 5 quasars per square degree. If we find 80% of them this would increase the number of quasars found in the skirt from 5,000 to 20,000 giving us a projected 95,000 quasars. The cost of this is that instead of needing 5.1 fibers per square degree in the skirt, we need about 17.9 fibers per square degree. The average number of fibers per plate allocated to quasars would have to be increased by about 45%. One advantage of this scheme is that the number of fibers required is much more evenly distributed across the sky. However, this would probably be prohibitively costly in terms of fibers or total number of plates needed to complete the survey.

Another way to look at the trade off between the technical goal of high selection efficiency and the science goal of exploring the luminosity function of QSOs at all redshifts is the following: Within the magnitude limits we have chosen, we have picked the quasars with the highest finding efficiencies. If we instead wanted to find quasars with higher efficiencies, we would have to move fibers from the brighter quasars to fainter quasars in such a way that we get our average efficiency up from 54% to about 71%. We definitely want to keep the central

UVx quasars ($z < 2.2$) which give us the highest finding efficiency. For scientific reasons, we really want to keep the outliers in the central regions of color-color-color space, since they represent the bulk of objects with $z > 2.2$. If we keep these, we have used up 20.3 fibers per square degree, and we have found 65,000 quasars. So, we need to find 7 quasars per square degree with 8.9 fibers per square degree. This requires 79% efficiency. There is no way to do this in the skirt, unless we use only radio sources, maybe. We are rapidly forced to the conclusion that in order to get an efficiency around 70%, we cannot sample all of color space. We would have to go after only the UVx quasars that are in regions of highest quasar density, hurting the scientific objectives of the survey.

8.4 Outline of Target Selection Algorithm

We divide the target selection into survey center target selection, and target selection in the skirt. In the central section, we would like to:

1. Begin with a list of all point sources that are bright enough to detect as a quasar with the spectrograph. To get the numbers in this paper, the list would contain all stellar objects with $B < 20$, but we would probably tune this when we have more information about what stars and quasars look like in the survey colors.
2. Choose as targets UVx and candidates that are distinct from the stellar locus based on their position in color space. We will probably also divide the QSO target color space into regions of different priority, so that if some targets have to be culled, we will be able to throw out the least important ones. We are tentatively planning to do this in the four-dimensional space: u-g, g-r, r-i, some magnitude.
3. We would also like to target the very red objects that could be the $z > 5$ quasars. These will probably be chosen from r-i, i-z, and z, as well as lack of u or g flux. These have not been discussed elsewhere, because at this time little is known about how we will find these and what the background will be. We understand that these get about three fibers per plate, but it is not yet determined whether these are tiled with quasars or added in at the end with serendipity. Their fibers may also be reserved before the main tiling to make sure these rare but important objects are targeted.
4. We choose all radio sources that are in our set of point sources that are bright enough to detect with the spectrograph. Again, this limit is somewhere near $B = 20$.

The quasar algorithm will set a target flag that will say whether it was targeted as UVx, as an outlier, as a high redshift candidate, or as a radio source.

In the skirt, we will use essentially the same algorithm, except the limiting magnitude will be $B < 18$ for both the color-selected and radio-selected samples. Also, we are allowed to make the color selection criteria depend on galactic latitude and longitude in the skirt. The high redshift quasars will be chosen the same way, and to the same magnitude limit that they are chosen in the central survey. If the stellar locus does not move substantially, and reddening does not cause problems, the bright sample should be uniform over the whole sky.

After selecting targets, we will run a quality assurance task to make sure the algorithm has selected sensible candidates. We would like to see, for instance, whether the targets are more or less evenly distributed across the sky. We would also like to use our own systems to reduce the target list in the event that it will drastically overrun our fiber budget. This monitoring task will hopefully get smaller and smaller as the test year progresses.

Also during the test year, we need a monitoring process that runs after the spectroscopic pipeline. We need to be able to monitor the quasar targets to make sure that we are getting high efficiency and that we are getting a reasonable sample of quasar redshifts. This process will probably continue throughout the whole survey in some form, since we would like to watch for things that are scientifically interesting, like the highest redshift quasar.

9 Projects To Be Completed Before Test Year

Latest time estimates suggest the test year will begin sometime between May and October of 1996. Between now and then there is time to work on several tasks so that we are ready to evaluate the test year data quickly. These tasks are:

9.1 Code Algorithms

Write code to actually select the QSOs from the list of all objects with colors. The algorithms we are trying are:

1. UVx multi-dimensional plane cut: Using the 3-D tools described above, one plots the stellar objects in (u-g, g-r, r-i) space. If one has any information from spectra or other means on which objects actually are QSOs, those targets are plotted in a different color. One then, by eye, manipulates a plane in 3-D space which separates the spheroid end of the stellar locus from the bulk of the UVx QSOs. The coordinates of this plane are recorded and serve as a cut to select UVx QSOs. This process may be repeated for several magnitude bins ($r < 17$, $17 < r < 18$, $18 < r < 19$, etc), as the stellar locus may change somewhat with magnitude. This algorithm is already coded.
2. Fifth nearest neighbor: A stellar locus is filled with colors of a known number of stars (say 100,000). Then each point in color space that one wishes to test is examined. Successive growing hyper-spheres are grown outward from the test point. When another star is enclosed in the growing sphere, a counter is incremented. When the count reaches 5 (an arbitrary threshold), the radius of the sphere at that point is recorded as the fifth nearest neighbor distance. If the radius is above some threshold, then the point in color space is a QSO candidate, since it is sufficiently far from a dense region of the stellar locus. If we use a fixed or a set of fixed stellar loci for the survey (fixed during the test year), then calculating this distance is a one-time procedure. If the stellar loci change significantly with position (l,b), then this procedure may need to be applied in 'real time' to stellar sources taken from local (l,b) space. Since this algorithm is a bit tricky to code, an alternative is:

3. Convolve nearby color space, count neighbors: A color cube is divided into discrete sub-cubes, say 0.04 mags on a side. A stellar locus is filled with counts of stars having colors within each sub-cube. Then each point (sub-cube) in color space that one wishes to test is examined. The sub-cubes surrounding the cube in question are then weighted by a Gaussian function which diminishes as one moves further from the central sub-cube to be tested. Object counts found in the surrounding cubes are added to a sum weighted by the Gaussian weight. The width of the Gaussian is determined by errors in the magnitudes of the measured objects or in the test object. If the sum is less than a threshold, than the objects in that sub-cube are candidates.
4. Distance to ‘normalized locus’: There are several ways to determine a metric distance to a stellar locus. We may either ‘straighten’ the stellar locus by using a principal component analysis technique such as Dawn Lenz has done for simulated stars, or we may define a piecewise linear stellar locus with errors as described in Gaidos et al. (1993), or we may use a combination of these techniques. Distances from such a locus, appropriately weighted by errors, are then easy to obtain. Although this requires more modeling, it may require fewer square degrees for a reliable solution than setting the threshold on a bin-by-bin basis in a color cube.
5. Rank Stellar/Rank QSO: For objects outside of the UVx region (which is the plane defined region, see item 1), QSO targets will be ‘ranked’ with a number that is a linear combination of their distance from the stellar locus and their distance to the QSO locus. Their distance to the stellar locus may be defined by item 2, 3 or 4 above. The distance to the QSO locus may be done with simple plane cuts or it may be done with a more sophisticated ‘perpendicular distance’ based on a technique similar to that for stars in item 4 above. The weighting will be such that ‘close to QSO locus = higher weight’ and ‘further from stellar locus = higher weight’. The relative weighting is to be determined by trial and error. If we find too many targets per square degree, we can eliminate those with the lowest priority.
6. FIRST radio source: A limited number of fibers will be placed on radio source detections from the FIRST (radio) catalogs. A list of such sources is expected to be available prior to the beginning of the survey. These sources have positional errors of $< 1''$ and so can be matched to our detections. If we see a positional match to one of our objects within $< 1''$ and it has $B < 20$ or so, then we count this as a potential source. We may exclude some regions of red-dwarf color space.
7. Hi-z QSO: The high-z QSO selection algorithm will look for objects with very large $r - i$ or $g - i$ colors and essentially no u or g flux. The r flux may be below the $r < 19$ detection limit as long as there is sufficient flux in the i and z bands to get the red portion of the spectrum in a one hour spectroscopic observation.

9.2 Simulations

The simulations can be enhanced by adding realistic white dwarf and low-metallicity sdB, sdO stars to the color tables. Galaxies with appropriate colors, including blue-compact galaxies could also be added to the simulations. Don Schneider has code to do most of these, given time. Dawn Lenz has provided Don with spectral and photometric data on low-metallicity stars.

9.3 Map QSO locus with real data from telescopes

Since most published QSO photometry is of low accuracy and only in 2 or 3 bands (*UBV*), and since it is important to know where the QSOs are expected to be in color space, a program to get ugriz colors of known QSOs should be continued. The MT can reach objects of $r = 18$ with some effort; data for about 25 bright QSOs has been obtained by Gordon Richards. This data still needs to be reduced and calibrated against standard stars. To get data for QSOs with $r = 20$ would require excessively long integrations on the MT. It may be possible to use the 3.5m telescope with the DSC and a filter wheel this winter. Another possibility is the USNO 1m (which is already equipped with filter wheel and ugriz filters). Photometry for about 100-150 QSOs in five bands at all redshifts is the goal.

9.4 Map Stellar locus from DSC data

A set of data was obtained in May-June 1995 of a square degree or so of sky with accurate magnitudes possible down to $r = 20.5$. This data provides a good check against the KK data, and will provide about 1000 data points which can be used to generate a sample stellar locus with real survey colors. About 5 known QSOs are expected to be detected and the above selection algorithms can be tested (in a limited sense) to see if we would have targeted these few objects.

9.5 Process KK data further

The KK data can be put onto a ugriz color system using the color transforms supplied by Jeff Munn and reproduced in Appendix A (see Figure 6). We can compare in detail the KK stellar locus with that of the simulations and that of the DSC data. Steve Majewski may be able to provide zero points for the KK data which will then help set a limit on how much the stellar locus moves as a function of (l,b) around the Galaxy.

9.6 Make Test Year Plan

With insight from the above data and QSO selection algorithms, we need to write down a specific test year plan which specifies how much imaging camera data (2 nights – 1 week) and how much spectroscopic data (3 partial plates to 10 full plates) are needed to tune the QSO selection algorithm for the actual survey. Also, we need to specify which parameters are tunable and have software in place to answer the following:

- Does stellar locus change significantly with (l,b)?
- Where should the boundary between the skirt and the cap be defined?
- Into how many magnitude bins should the loci be divided?
- Where should the UVx region be in color space?
- What should the relative weightings be between ‘distance from stellar locus’ and ‘distance to known QSO locus’?
- Should objects with non-stellar FWHM (and fainter than the galaxy limit) be considered as targets if they have non-stellar or QSO-like colors?

10 Appendix A: UJFN \rightarrow ugriz color transformations

Jeff Munn has calculated filter transforms: Below are listed transformations from photographic UJFN to SDSS ugriz. They are based on the transformations given in Majewski (1992) and Fukugita et al. (1995)

Note the particularly easy and useful approximation (which is good to about a 10th of a magnitude over most colors): $r = F + 0.1$.

$$\begin{aligned}
 g &= F + 0.72(J - F) - 0.11 \quad -0.46 \leq J - F < -0.05 \\
 &= F + 0.83(J - F) - 0.12 \quad -0.05 \leq J - F \leq 1.77
 \end{aligned}$$

$$\begin{aligned}
 r &= F - 0.19(J - F) + 0.13 \quad -0.46 \leq J - F < -0.05 \\
 &= F - 0.07(J - F) + 0.12 \quad -0.05 \leq J - F \leq 1.77
 \end{aligned}$$

$$\begin{aligned}
 u - g &= 1.39(U - J) + 1.12 \quad -1.48 \leq U - J < 0.01 \\
 &= 1.22(U - J) + 1.00 \quad 0.01 \leq U - J \leq 1.41
 \end{aligned}$$

$$\begin{aligned}
 r - i &= 0.72(F - N) - 0.25 \quad -0.26 \leq F - N < 1.57 \\
 &= 1.03(F - N) - 0.73 \quad 1.57 \leq F - N \leq 2.76
 \end{aligned}$$

$$\begin{aligned}
 r - z &= 1.19(F - N) - 0.42 \quad -0.26 \leq F - N < 2.31 \\
 &= 1.97(F - N) - 2.21 \quad 2.31 \leq F - N \leq 2.76
 \end{aligned}$$

11 Acknowledgments

One of the advantages of working on a big project like the Sloan survey is that we have access to many people with knowledge or data that apply to the problem we are trying to solve. We would like to thank Jeff Munn for helping us to interpret the Koo-Kron data and look through the spectra of quasar candidates. Rich Kron provided useful discussions of the data. Dawn Lenz provided preliminary principal component analysis of simulated stars. Steve Majewski provided us with pre-publication work on white dwarfs in the sample. Pat Osmer agreed to come to Fermilab to give a talk on his color-selected quasar sample. His insights were useful. Jon Loveday ran numerous target selection simulations to determine whether it would cause selection biases to tile the quasars differently in the center of the survey than in the skirt. He and Steve Kent also answered numerous questions regarding target selection.

12 References

References

- [1] Bahcall, J. N. & Soneira, R. 1980, ApJS 44,73
- [2] Boyle, B. J., Fong, R., Shanks, T. & Peterson, B. A. 1990, MNRAS 243, 1.
- [3] La Franca, F., Cristiani, S., & Barbieri, C. 1992, AJ 103, 1062
- [4] Fukugita, Ichikawa, Gunn, Doi, Shimasaku, and Schneider, 1995, preprint: "The Sloan Digital Sky Survey Photometric System".
- [5] Gaidos, E. J., Magnier, E. A. & Schechter, P. L. 1993, PASP 105, 1294
- [6] Gunn, J. E. & Stryker, M. 1983, ApJS 52, 121
- [7] Green, R. F., Schmidt, M. & Liebert, J. 1986, ApJS 61, 305.
- [8] Koo, D. C. 1986, ApJ 311, 651.
- [9] Kron, R. G., Bershadsky, M. A., Munn, J. A., Smetanka, J. J., Majewski, S., & Koo, D. C. 1995, preprint
- [10] Majewski, S. 1992, ApJS, 78, 87
- [11] Kron, R. G., 1980, ApJS 43, 305.
- [12] Warren, S. J., Hewett, P. C., Osmer, P. S., & Irwin, M.J. 1991a, ApJS 76, 1.
- [13] Warren, S. J., Hewett, P. C., Osmer, P. S., & Irwin, M.J. 1991b, ApJS 76, 23.

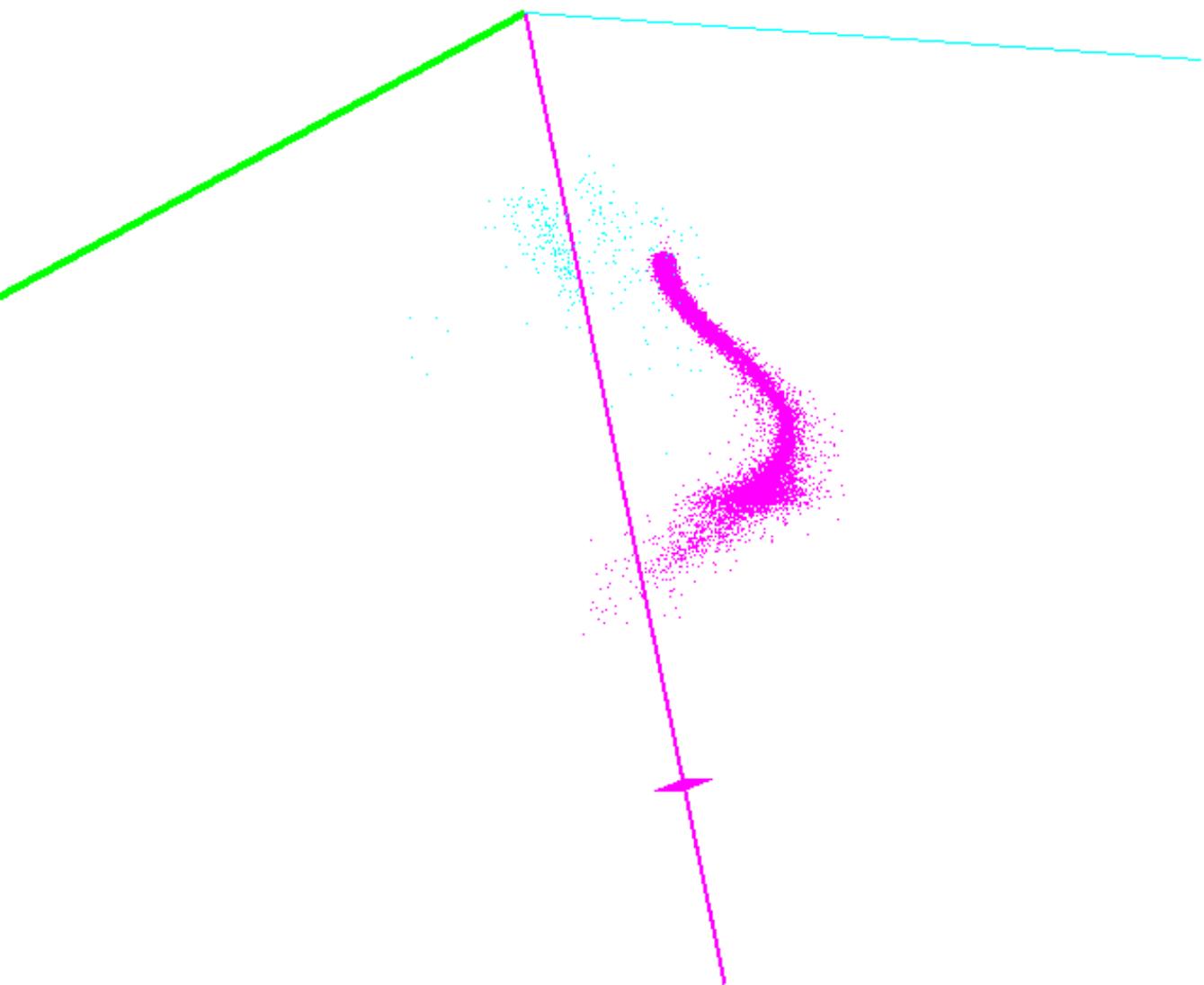
13 Figure Captions

1. **Simulated data b=90 degrees.** This shows a 2-D projection of simulated stars (magenta) and QSOs (cyan) in the 3-D color space $u-g$, $g-r$, $r-i$. The projection of a square on the magenta axis shows that this axis is tilted out of the plane of the page, while the lack of a projected square on the other two axes shows that these axes are tilted into the page. This projection shows the highest redshift QSOs as the clump farthest to the left in the image. There are as many stars represented as we expect to find in 3 sq. deg. of the Sloan survey, and as many QSOs as we expect to find in nine sq. deg. The quasars are over-represented by a factor of three. Notice the clean separation of QSOs from stars. WDs, sub-dwarf stars and other stars with QSO-like colors are not included in the simulations.
2. **Simulated data b=30 degrees.** This 2-D projection was chosen to best separate the stars from the QSOs. Note that even though the star density is over 10 times as high as at the galactic pole, the stars and QSOs are still well separated.
3. **Simulated data, b=90 degrees, color cube.** Color space was divided into cubes of 0.5 mags on a side. The number of QSOs in each sub-cube was divided by the total number of objects in that sub-cube to determine the ratio of stars to QSOs. For each sub-cube containing any objects, we plotted a smaller cube whose color denotes the percentage of quasars in that cube: 0% (all stars) - magenta, 1%-40% - orange, 40%-60% - grey 60%-80% - blue, 80%-100% (nearly all QSOs) - cyan. The orientation is the same as in Fig. 1. This technique allows us to look for the regions of highest target efficiency.
4. **Simulated data, b=30 degrees, color cube.** This figure is the same as Fig. 3, except it is at lower galactic latitude. The orientation is the same as Fig. 2. Again, we see the good separation of stars and QSOs.
5. **KK data, sum of four fields.** This figure shows the sum of the objects from four fields of various galactic latitudes and longitudes that cover an area of 1.5 square degrees. The limiting magnitude is $J=21.5$, which is somewhat fainter than the spectroscopic limit of the Sloan survey. As before, we have represented the QSOs by cyan points. All other types of objects are indicated with magenta points. The QSOs have been confirmed by spectroscopy. Not all the stars have been confirmed spectroscopically. As the simulations show, most of the QSOs reside in the UVx region of color-color space. The real data shows that other types of objects of blue color exist near the quasar locus, and that there are also outliers from the stellar locus in other regions of color space.
6. **KK data, including galaxies.** This shows three of the four KK sub-fields from Figure 5 (objects with $19.5 < J < 21.5$) converted to $ugri$ colors, and include extended objects (galaxies) as grey dots. Note the position of the galaxies relative to the stars and QSOs and the contamination of the QSO color-space by galaxies. It is possible that we could

target extended objects in some of the color space where we are looking for QSOs, but certainly not in all of it.

7. **KK data, color cube.** This figure shows the color cube representation (see Fig. 2) of the KK data. The percentage of QSOs in each sub-cube is given by: 0% (all stars) - magenta, 1%-40% - orange, 40%-60% - grey 60%-80% - blue, 80%-100% (nearly all QSOs) - cyan. We can see from this that there is no way to rotate the projection to completely separate the quasars from the non-quasars.
8. **UVx selected points from KK data.** This figure shows a sub-section of Fig. 5, selecting only those objects which meet a UVx color plane cut. Objects are color-coded by type (determined from follow-up spectroscopy): QSOs - cyan, stars - purple, Narrow Line Emission Galaxies - green, not QSOs (faint low S/N spectrum stars) - magenta, white dwarf - grey, unknown (unobserved or low S/N spectrum) - orange. We have rotated the color space to separate the quasars from the non-quasars as much as possible.
9. **Definition of the survey center and skirt.** Using the Bahcall-Soneira model, we have plotted the stellar density at $V=20$ as a function of galactic latitude and longitude. We have included the limits of the Sloan Digital Sky Survey as the red ellipse. The galactic center is in the center of the drawing. The galactic latitude at any point on the plot is proportional to its distance from the center. The edge of the circle is 20° galactic latitude. Star counts for lower galactic latitudes were not calculated. The minimum stellar density is 1202 stars/sq. deg. (Notice that it is not at the galactic pole!) The central area has star density between 1.0-1.5 times the minimum density. Each successive area out from the center has 1101 more stars per sq. deg. than the previous region. The color map we used saturates at about six times the minimum stellar density. The blue areas show the proposed position of the survey center, in which we will search for quasars to 20th magnitude. The skirt is the yellow-orange region interior to the red ellipse.

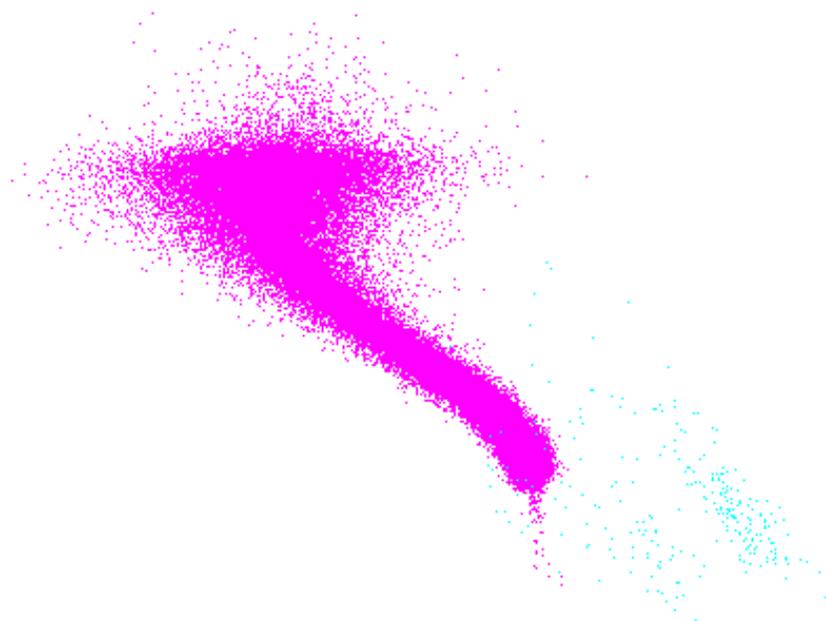
Simulation $b=90$ $l=0$: QSOs (cyan) and Stars (magenta)



plane rx: 0
plane ry: 0
plane z: 0
rotation around x: -16.8
rotation around y: 137.4

u-g: -1 5
g-r: -1 4
r-i: -1 4

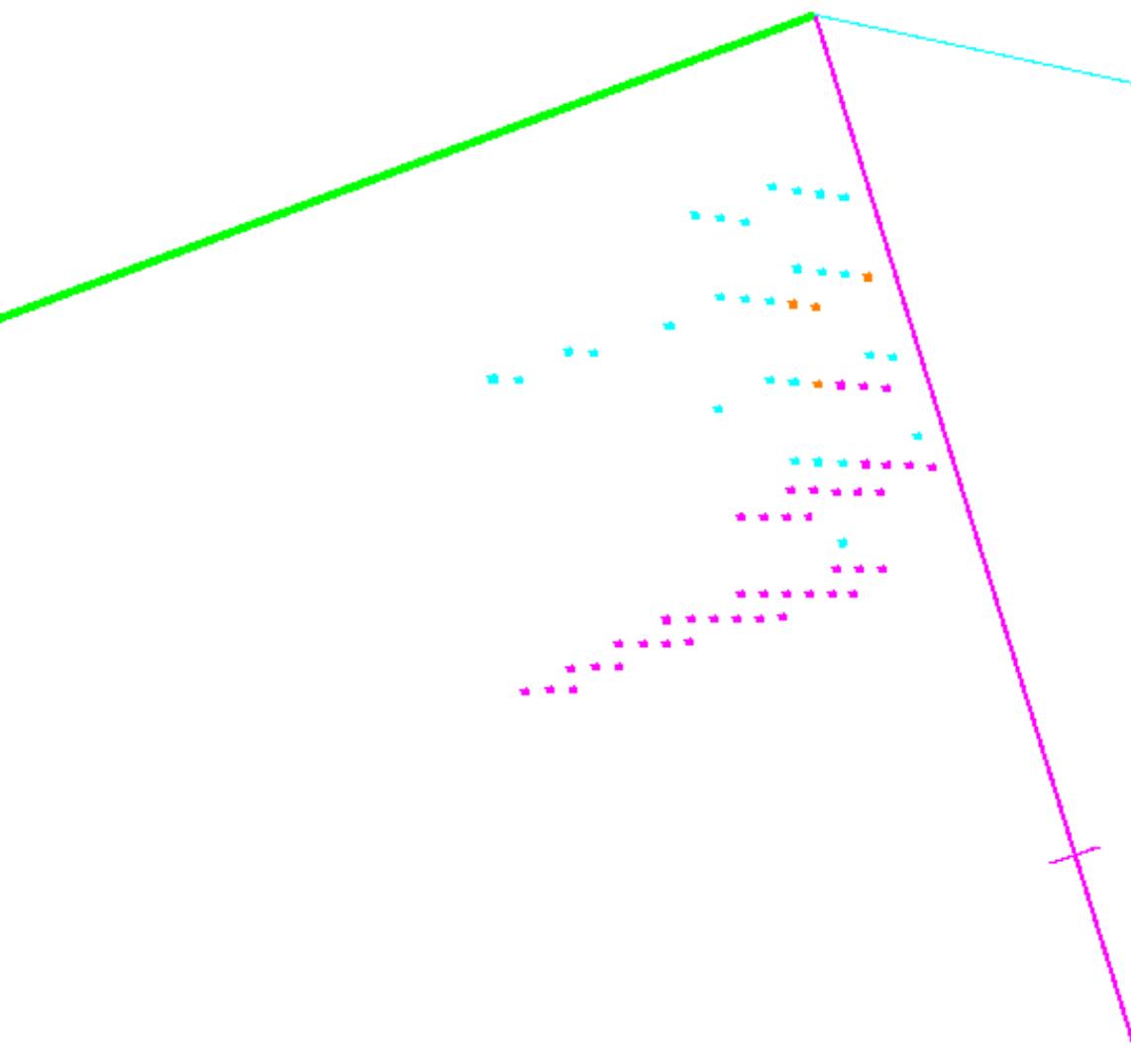
Simulation b=30 l=0: QSOs (cyan) and Stars (magenta)



plane rx: 0
plane ry: 0
plane z: 0
rotation around x: -192
rotation around y: 10,2

u-g: -1 5
g-r: -1 4
r-i: -1 4

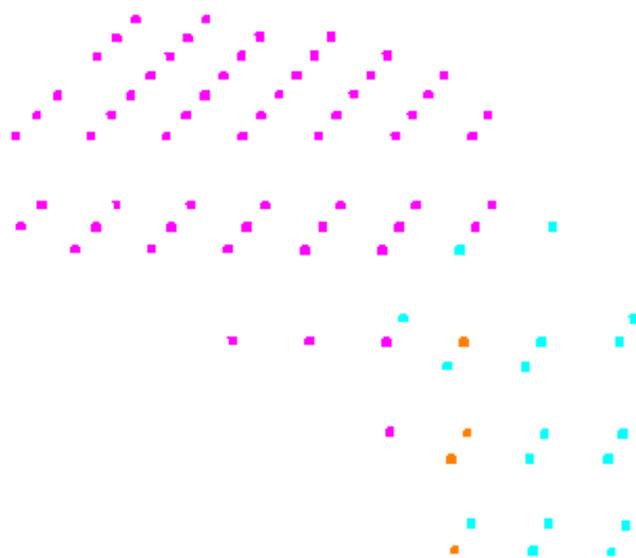
Sim 90 - Q50 frac: magnta:0% or:1-40% red:40-60% bl:60-80% cyan:80-100%



u-g:-1 5
g-r:-1 4
r-i:-1 4

rotation around x: -18
rotation around y: 110.4

Sim 30 - Q50 frac: magnta:0% or:1-40% red:40-60% bl:60-80% cyan:80-100%



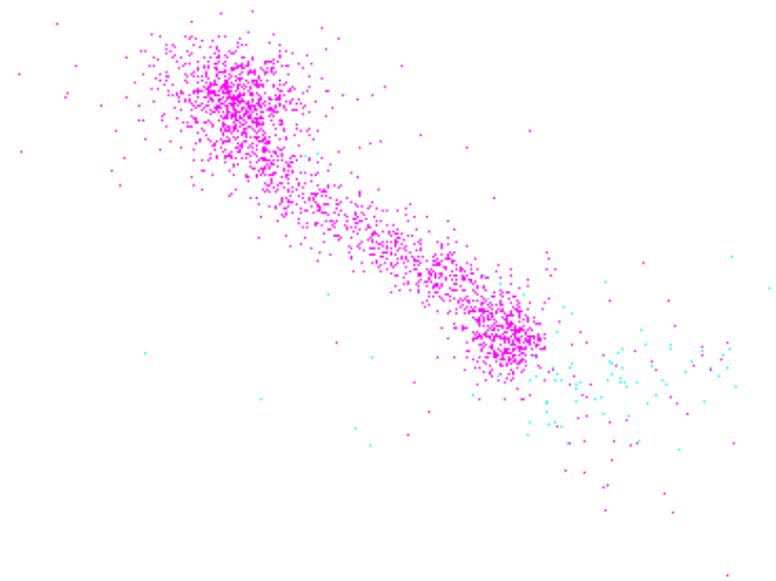
u-g:-1.5

g-r:-1.4

r-i:-1.4

rotation around x: -193.2
rotation around y: 10.8

KK data: QSOs (cyan) and Stars (magenta)



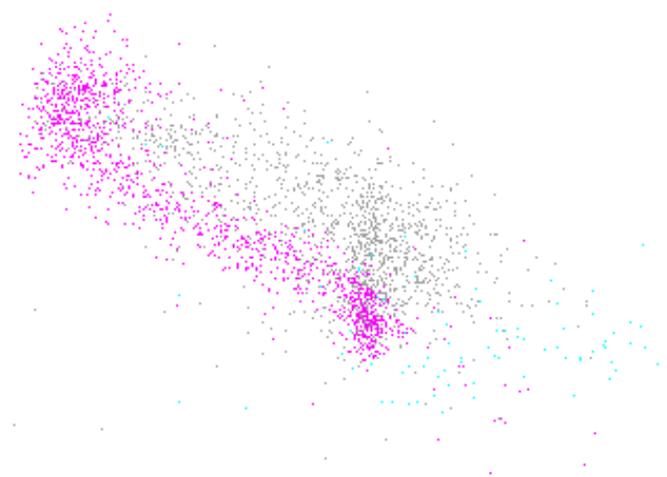
plane rx: 0
plane ry: 0
plane z: 0
rotation around x: -193.2
rotation around y: 10.2

U-J: -2 3

J-F: -2 3

F-N: -2 3

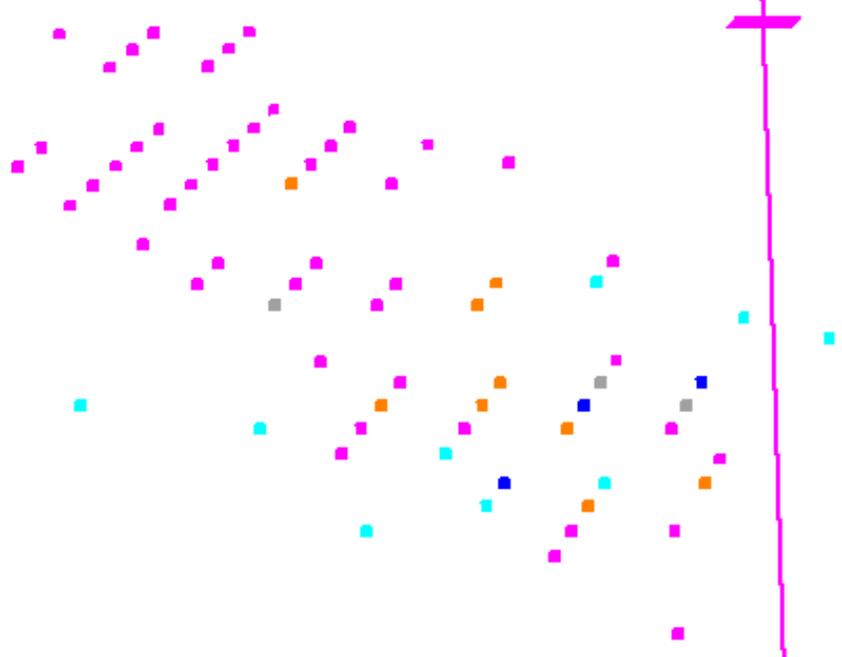
KK data: Stars:magenta Q50s:cyan Galaxies:gray



u-g:-2 3
g-r:-2 3
r-i:-2 3

plane rx: 0
plane ry: 0
plane z: 0
rotation around x: -194.4
rotation around y: 10.8

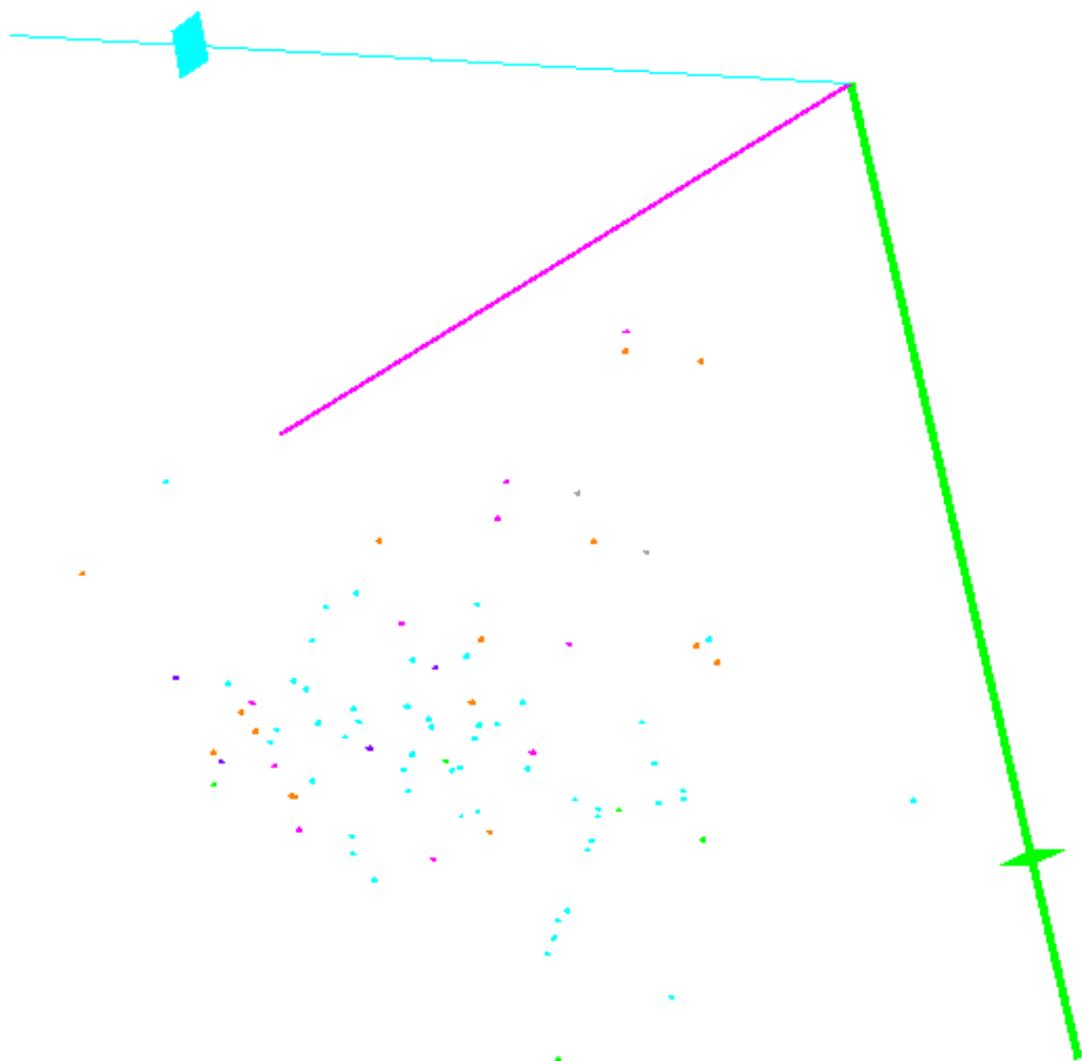
KK data - Q50 frac: magnta:0% or:1-40% grey:40-60% bl:60-80% cyan:80-100%



U-J:-2 3
J-F:-2 3
F-N:-2 3

rotation around x: -193.2
rotation around y: 11.4

UVx: QSO:cyan Star:purple Gal:green notQSO:magenta WD:grey Unknown:orange



plane rx: 0
plane ry: 0
plane z: 0
rotation around x: -70.8
rotation around y: -37.2

U-J: -1 0
J-F: 0 1
F-N: -1 1

