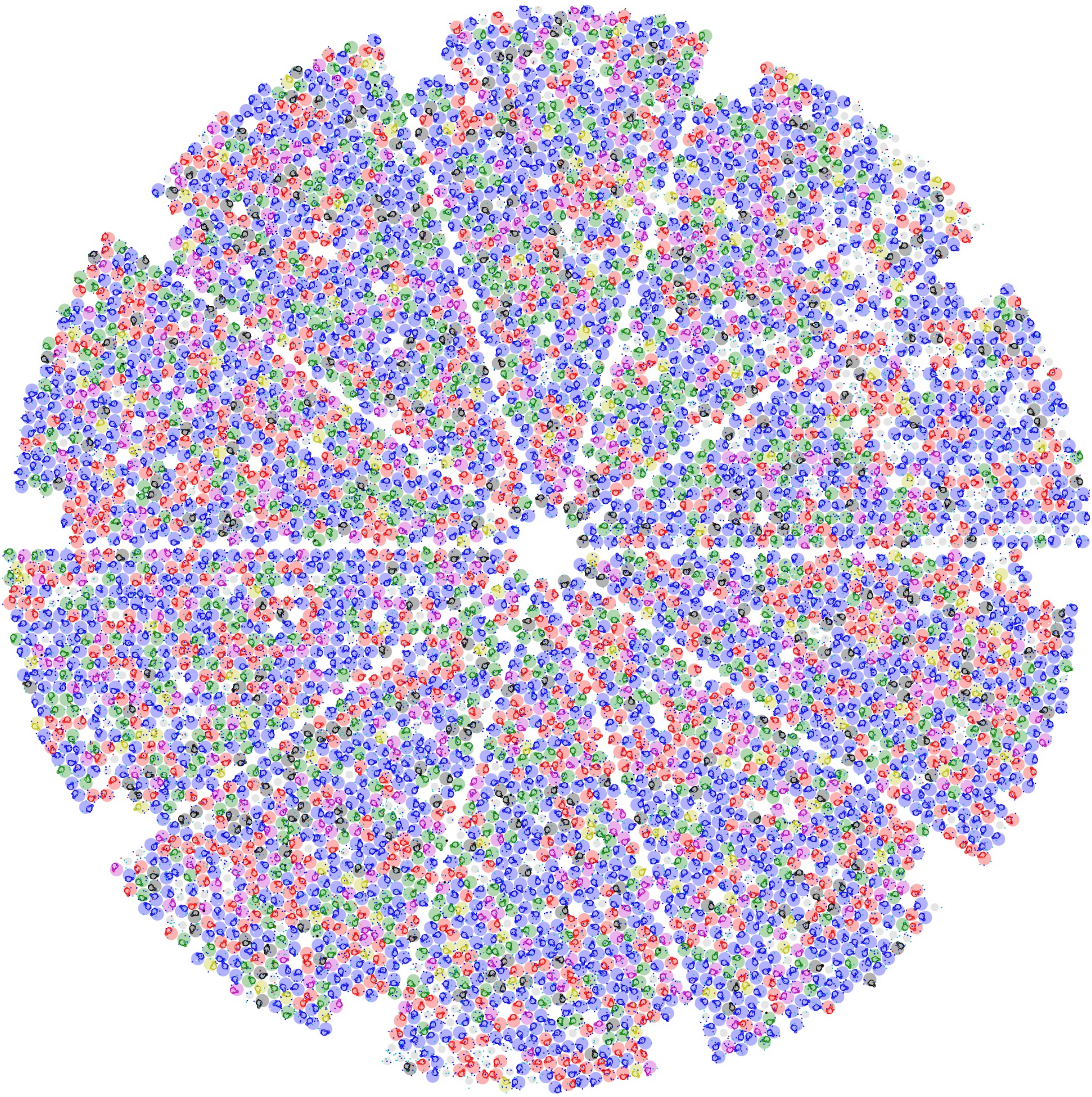
Louis Garrigue

Research internship from February to June 2015

Fiber assignment for DESI



Supervisor : Robert N. Cahn, Senior physicist at the Lawrence Berkeley National Laboratory - rncahn@lbl.gov

Mon. Not. R. Astron. Soc. **000**, 000–000 (0000) Printed 22 August 2015 (MN LaTeX style file v2.2)

Fiber assignment in DESI experiment

Louis Garrigue[1,2]⋆

[1] *Department of Cosmological Physics, LBNL, Berkeley*
[2] *Département de physique, École normale supérieure, Paris*

22 August 2015

ABSTRACT

Dark energy study began in 1998, when two independent teams, led by Adam Riess and Saul Perlmutter, found out, using quasars measurements, that the expansion rate of the universe was increasing. The then standard cosmological model, which components were baryonic and dark matter ruled by general relativity, had to be changed. Λ-CDM model now includes dark energy through the cosmological constant, a component explaining the fact that gravity is weaker than expected, and compatible with the accelerating rate of the universe. It is still today the standard model, but still suffers from a lot of imprecisions which make it questionable. Future breakthroughs in the three different branchs of cosmology - origin, evolution and fate of the universe - are now totally dependent on understanding the nature of dark energy, which constitutes 70% of the total amount of energy in the universe. This thus involves to observe millions of galaxies and quasars, their spectrum and redshift carrying a lot of worthy data, especially with Ly-α forest. Such a work needs huge surveys, like SDSS, BOSS, and now *DESI*, heir of previous ones. DESI will then significantly improve our knowledge on cosmological parameters. The power spectrum (Fourier transform of the correlation function of the matter density), imprinted by baryon acoustic oscillations (BAO) which went through the early universe, will also be computed much more precisely, and in high redshift regions. One of the key parts of DESI experiment is the assignment of its 5000 fibers on 10666 different tiles, observing celestial objects of seven different natures in a mock catalog of 70 million ones. Indeed, the efficiency of DESI will directly depend on multiple features of this assignment algorithm : number of unassigned fibers, number of times a QSO Ly-α will be observed, introduced bias, etc. Here is a brief presentation of the standard cosmological model and DESI experiment. We then describe our fiber assignment : design, C++ code organisation, principles, rules and results. We ran our jobs on the Edison NERSC Supercomputer - the code is freely and publicly available online on the DESI github repository.

Key words: gravitational lensing: desi experiment - cosmology - correlation function

1 INTRODUCTION

At the begining of the XX^{th} century, elaborating general relativity, Einstein introduced a cosmological constant (Einstein 1917) to counterbalance gravitation and make the universe static. He eventually gave up this idea when taking into account Edwin Hubble's discovery (Hubble 1929) that the universe is expanding, at a computed rate (though overestimated by an order of magnitude), observing nebulae. This was the begining of modern cosmology. Perlmutter (Perlmutter et al. 1997) and Riess (Riess et al. 1998), in the late 1990s, independently discovered that this expansion was accelerating, rather than slowing as had been long assumed. Combined with general relativity, this means that 70% of energy in the universe is ”dark”, which effect is pulling apart matter.

A lot of different theories try to explain what is this dark energy, for ”dark energy” is just a catch-up term. Though, none of them are currently satisfying. The first picture of the universe we have is taken by the Planck satellite (Tauber & ESA 2004), which scrutinizes the Cosmic Microwave Background, signal from the sky 379,000 after the big bang (recombination), that we still receive. Cosmologists try to find out how, from big bang, universe became the way the CMB tells us. They also try to understand how it then evolved from recombination to now under gravitation, and what will be its ultimate fate. All those investigations are deeply linked to dark energy, which makes it a key to understand our reality. Underlying of it, are one of the most relisient and impenetrable mysteries of science.

Current researches consist in creating new theoretical models, and on the experimental side to realize redshift studies of huge amounts of celestial objects. It leads to constrain parameters which appear cosmological equations. This is done by comparision of

⋆ e-mail : lgarrigue@lbl.gov ; louis.garrigue@ens.fr

simulations with observations, and leads to eliminate some models.

SDSS and BOSS where among those huge studies. Now DESI (formerly called BigBOSS (Schlegel et al. 2009)) is the succession. It is a ground-based stage IV survey, complementary to futur Euclid experiment (Amendola et al. 2013) (european spacecraft, planned to be launched in 2020), SDSS IV and eBOSS. DESI will be based at the 4-meter Mayall telescope at Kitt Peak, Arizona, to measure the spectra and redshifts of 50 million galaxies and quasars on a 14000 square degree patch of the sky. 5000 controled actuators mounted by optical fibers, placed on the focal plan of the telescope, will observe signals from those objects.

The work we achieved with Bob Cahn was the fiber assignment for DESI, which is the mapping between fibers positions and galaxies/quasars locations. This challenge consists in looking at as much objects as possible, though at the same time computationnally fast and trying not to bias the observation. Our computational jobs, very heavy, had been runned on Edison, the most powerful NERSC supercomputer with its 133,824 Cray XC30 cores, 357 TB of memory, and 7.56 PB of disk. We had to parallelize a part of the code, which would have been too slow to run even on Edison. This C++ 4,000 lines code I wrote under the supervision of Bob Cahn is divided into 5 modules (which provide low level to high level functions), a "main" and a parameters input file.

The Lawrence Berkeley National Laboratory, funded by the U.S department of Energy and linked with the University of California in Berkeley, is the main center of DESI organisation. It not only hosts research on Physical Sciences, but also on environmental sciences, computer sciences and biotechnologies.

We first briefly recall basic principles of the standard model of cosmology, we then describe more precisely DESI features. Next, we describe our fiber assignment and finally conclude presenting our results and futur work that will have to be done.

2 COSMOLOGY TODAY

2.1 Dark matter

Dark matter is invisible (doesn't react with electromagnetic radiations), and has been discovered when physicists noticed that there must be additional matter in clusters of matter. It accounts for 27% of total matter and implies that the standard model of particle physics must be incomplete. Through gravitational interaction with baryonic matter, they influence the large scale structure formation, but also small-scale clustering of galaxies lying in host dark matter halos.

2.2 Λ-CDM model

This is the currently standard model. It assumes that the universe is ruled by regular general relativity, dark matter and the Einsein's cosmological constant Λ. It respects the cosmological principle : our observational location in the Universe is not unusual or special ; on a large-enough scale (much greater than the typical size of galactic clusters), the Universe looks the same in all directions (isotropy) and from every location (homogeneity). Then comes the well-known FLRW metric :

$$g_{\mu\nu}dx^{\mu}dx^{\nu} = -c^2dt^2 + a(t)^2(\frac{dr^2}{1-kr^2} + r^2 d\Omega) \qquad (1)$$

There is a growing consensus among cosmologists that the total density of matter is equal to the critical density, so that the universe is spatially flat (curvature $k = 0$). Approximately 24% of this is in the form of a low pressure matter, most of which is thought to be non-baryonic dark matter, while the remaining 71% is thought to be in the form of a negative pressure dark energy, like the cosmological constant. If this is true, then dark energy is the major driving force behind the fate of the universe and it will expand forever exponentially. Here is the dynamical equation used is Einstein's field equation :

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + g_{\mu\nu}\Lambda = \frac{8\pi G}{c^4}T_{\mu\nu} \qquad (2)$$

We apply it to the FLRW metric, on a perfect fluid of energy density ρ and pressure p to get Friedmann equations :

$$\left(\frac{\dot{a}(t)}{a(t)}\right)^2 + \frac{c^2k}{a^2} = \frac{8\pi G}{3}\rho + \frac{c^2\Lambda}{3} \qquad (3)$$

$$\frac{\ddot{a}(t)}{a(t)} = -\frac{4\pi G}{3}(\rho + 3\frac{p}{c^2}) + \frac{c^2\Lambda}{3} \qquad (4)$$

Since it's a perfect fluid, we also use $p = w\rho c^2$ the state equation of the universe. One of the most interesting thing to compute from them is the scale factor $a(t)$. Using those Friedmann equations, one easily gets :

$$\frac{\dot{\rho}}{\rho} = -3(1+w)\frac{\dot{a}}{a} \qquad (5)$$

and then, if w is a constant : $\rho \propto a^{-3(1+w)}$.

For baryonic/leptonic matter, $w = 0$. For radiation, $w = 1/3$. The cosmological redshift z is defined by $1 + z \equiv 1/a$, the Hubble parameter by $H(z) \equiv \frac{\dot{a}}{a}$ and H_0 is the Hubble parameter today.

We define the energy density parameter for matter and radiation $\Omega \equiv \rho_i/\rho_{critic}$, where $\rho_{critic} \equiv \frac{3H_0^2}{8\pi G}$ is the critical density of the universe, beyond which it would recollapse due to gravitation. We also define $\Omega_k \equiv -c^2k/H_0^2$ and $\Omega_\Lambda \equiv c^2\Lambda/3H_0^2$, energy densities associated with curvature and cosmological constant (Weinberg et al. 2013).

The first Friedmann equation 3 can then be rewritten :

$$\frac{H(z)^2}{H_0^2} = \Omega_m(1+z)^3 + \Omega_r(1+z)^4 + \Omega_k^3 + \Omega_\Lambda \qquad (6)$$

This is the relation between dark energy (through its equation of state $w(z)$) and the expansion (through $H(z)$).

2.3 Dark energy

Saul Perlmutter's investigations showed that $\ddot{a}(t)$ is actually positive, which means that the expanding rate of the universe is still inscreasing and this was the beginning of dark energy hypothesis.

Dark energy is in fact a catch-all term for the origin of cosmic acceleration, and could be either a modification of general relativity, or a new kind of energy. Several other theoretical solutions (Mortonson, Weinberg & White 2013) are possible.

The density of dark energy, $7 \cdot 10^{-27} kg \cdot m^{-3}$ (with mass-energy equivalence), is much lower than the density of ordinary matter within galaxies. It comes to dominate the mass-energy of the universe because it is uniform in space.

Here are several possibilities for dark energy :

- Einstein's cosmological constant (a constant energy density filling space, and which can be formulated to be equivalent to vacuum energy, with repulsive gravitational effects)

• a dynamical scalar field called quintessence, sometimes considered to be a fifth fundamental force, and whose energy density can vary in time and space. Scalar fields that do change in space can be difficult to distinguish from a cosmological constant because the change may be extremely slow. In equation 6, the cosmological constant is equivalent to a fluid with state parameter $w = 1$ and a repulsive gravitational effect, and we call it dark energy as we don't know anything else about it. Now, lets imagine that dark energy as a fluid that fills the whole Universe but with a parameter w that can evolve in time, i.e. depending on the scale parameter a. We find then that equation is modified by $\Omega_\Lambda \leftarrow \Omega_\Lambda exp(3\int_0^z \frac{1+w(z\prime)}{1+z\prime}dz\prime)$

• an error in GR. Theorists for example try to replace the Ricci scalar $\mathcal{R}$ with a function $\mathcal{R} + f(\mathcal{R})$

• an illusion even within general relativity, owing to an incorrect treatment of averaging in an inhomogeneous universe (Wiltshire 2008)

2.4 Galaxy-matter bias

The galaxy bias is defined as : $b = \frac{\delta_g}{\delta}$ (Kaiser 1984) in an attempt to reconcile the different clustering scale lengths of galaxies and rich clusters, which could not both be unbiased tracers of mass. From a Gaussian distribution of matter, the peaks which first collapse to form galaxies will be more clustered than the underlying mass distribution.

The galaxy "bias" - the relationship between the spatial distribution of galaxies and the underlying dark matter density field - is a result of the varied physics of galaxy formation which can cause the spatial distribution of baryons to differ from that of dark matter. On large scales the bias tends to be a constant value, which cosmologists take in their equations.

2.5 Correlation function

2.5.1 Definition

The correlation function and its Fourier transform, the power spectrum, are one of the main objects of study in today's cosmology investigations on dark energy. Main parameters of cold dark matter and dark energy models have an influence on them. This function then not only carries the information of BAO. Methods of MCMC are used to strongly constrain parameters, comparing theory and measurement.

The two-point correlation function ξ (TPCF, called the correlation function), can be defined from the fluctuations of the galaxy density by :

$$\xi(\mathbf{r}) = \langle\delta(\mathbf{x})\delta(\mathbf{x+r})\rangle \tag{7}$$

where the average is on the spacialvariable x

Galaxy redshift surveys take information of angular position and redshift of several kinds of objects, to build a 3D map of a section of the sky. *A priori*, those studies build the correlation function of baryonic matter only. Fortunately, distributions of regular and total matter are linked by the Kaiser model (Kaiser 1987) on the concerned scales :

$$\delta_g(\mathbf{k}) = (b + f\mu^2)\delta_m + \epsilon \tag{8}$$

b(z) is the linear galaxy-matter bias, it is given by galaxy formation physics. μ is the cosine of the angle between the line of sight and $\mathbf{k}$. $f(z) = \frac{d\ln D}{d\ln a}$ where D(z) is the growth rate ($\delta_m = D(z)\delta_m^{initial}$). f stands for the redshift-space distorsion : the influence of galaxies peculiar velocities (radial) on their redshift measurement. ϵ is a white noise. Taking the fourier transform, one can get :

$$P_g(\mathbf{k}) = (b + f\mu^2)^2 P_m(k) + \bar{n}^{-1} \tag{9}$$

$\bar{n}$ is the mean galaxy density.

One of the key points is the fact that there has not to be any bias added by the experiment techniques. And if there is any, it has to be known and counterbalanced. Bianchi et al. (2015) explains how to compensate the bias on correlation function.

2.5.2 Data and BAO

The clustering results in an increase of the correlation function at short distances, which usually exhibits a power law behavior :

$$\xi(r) \sim (\frac{r}{r_0})^{-\gamma} \tag{10}$$

with $\gamma \simeq 1.7$ and the power spectrum is then :

$$P(k) \sim k^{n_s} \tag{11}$$

where n_s is the spectral index

But the correlation carries much more information than only that. Baryon acoustic oscillations (BAO) are regular, periodic fluctuations in the density of the visible baryonic matter of the universe. BAO matter clustering provides a "standard ruler" for length scale. The length of this standard ruler (∼490 million light years in today's universe) can be measured by looking at the large scale structure of matter using surveys like DESI. Those oscillations come from recombination, when the universe was only 379,000 years old (z $\simeq$ 1000). At this time photons decoupled from baryonic matter, and then slight overdensities of matter began to grow under gravitation. In this early plasma acoustic waves were propagating, which were due to two competitive forces : radiation pressure pulling plasma apart on one hand, and gravitation leading to gather plasma in the other. Then (almost) only gravitation led the evolution of the universe at large scales, at the first order.

Those BAO thus can be seen on the correlation function and on the power spectrum. BAO are very important because they lead to constrain cosmological parameters. One can see their trace on Figure 1, taken from Benitez et al. (2009). The dashed line is the linear correlation function scaled with the linear halo bias ($b = 3$), while the black solid line corresponds to the nonlinear prediction. Their difference shows the degradation coming solely from nonlinear clustering. In addition, the triangle (red), square (blue) and cross (green) symbols show the measured correlation function after a Gaussian error degradation in the line-of-sight position of the halos is introduced.

3 DESI

The Dark Energy Spectroscopic Instrument (DESI) (Levi et al. 2013) is a Stage-IV experiment which will make the next major advance in dark energy, starting in 2018 and supported by the US Department of Energy.

3.1 Material features

The DESI instrument, that will be installed on the Mayall telescope in Arizona, consists of a new wide-field (3.2 deg. linear field of view) corrector plus a multi-object spectrometer with up to 5000

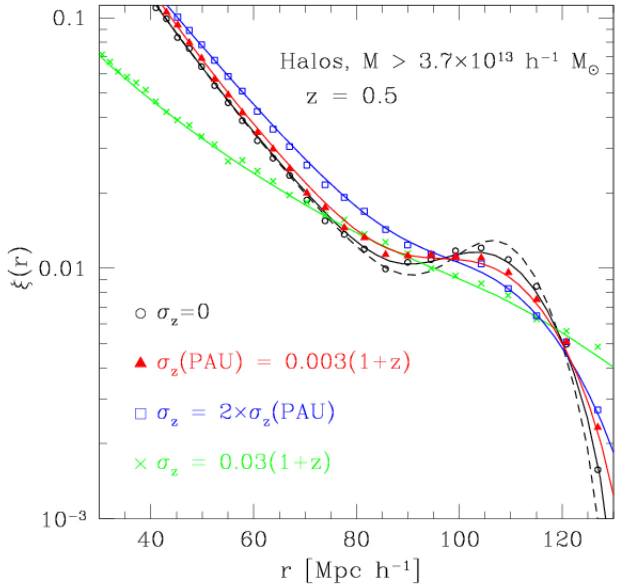


Figure 1. BAO peak imprinted in the correlation function

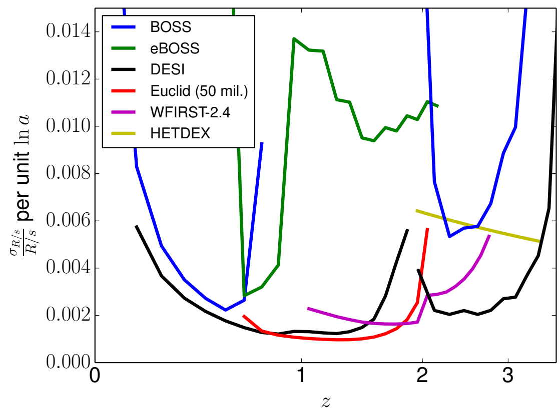


Figure 2. Comparison of DESI with other studies : fractional error on the dilation factor as a function of redshift, this is in some sense a distance error density

robotically positioned optical fibers, with a resolving power of R = 5000. The fibers feed 10 three-arm spectrographs (500 fibers for each spectrographs, also called petals) producing spectra that cover a wavelength range from 360-980 nm and have resolution of 2000-5500 depending on the wavelength. It will watch at 14,000 deg^2 in 10666 tiles from 2018 to 2022, with roughly 7 different tiles each night. Fiber positioners will have a reach of ∼6 mm from their central position.

A "pacman" version is envisaged in case the study runs out of money. This version has less spectrometers and the plate then looks like a pacman.

3.2 General aspects

It will study baryon acoustic oscillations (BAO) and the growth of structure through redshift-space distortions. The goals are :

(i) probe the effects of dark energy on the expansion history using BAO
(ii) measure the gravitational growth history through redshift-space distortions
(iii) measure the sum of neutrino masses
(iv) study the signatures of primordial inflation

The resulting 3-D galaxy maps at z $<$ 2 and Lyman-alpha forest at z $>$ 2 will make 1%-level measurements of the distance scale in 35 redshift bins. This will provide unprecedented constraints on cosmological models.

This survey is the successor of the Stage-III BigBOSS survey and complements imaging surveys such as the Stage-III Dark Energy Survey (DES, currently operating) and the Stage-IV Large Synoptic Survey Telescope (LSST, planned start in the next decade). Figure 2, taken from Font-Ribera et al. (2014), and which is usually displayed in DESI presentations, shows the projected BAO distance errors : a distance error which shows in what extent DESI will concurrence other studies on BAO measurement.

There is a comparison with BOSS (USA, ground-based, 2009-2014), eBOSS (USA, ground-based, 2014-2020), Euclid (Europe, space-based, 2020), WFIRST (Spergel et al. 2015) (NASA, space-based, after 2020) and HETDEX (USA, ground-based). One can particularly note the efficacity of DESI on high redshifts.

3.3 Tiles

14,000 deg^2 will be observed, divided into 10666 tiles. In average, a galaxy is covered 5.24 times (by different fibers), with only 2.6% of the footprint having a coverage of 3 or less. Figure du tiling. Figure des fibers. There are different tiling strategies, depending on what

3.4 Targets

À qui donc parles-tu, flocon lointain qui passes ?
À peine entendons-nous ta voix dans les espaces.
Nous ne te distinguons que comme un nimbe obscur
Au coin le plus perdu du plus nocturne azur.
Laisse-nous luire en paix, nous, blancheurs des ténèbres,
Mondes spectres éclos dans les chaos funèbres,
N'ayant ni pôle austral ni pôle boréal :
Nous, les réalités vivant dans l'idéal,
Les univers, d'où sort l'immense essaim des rêves,
Dispersés dans l'éther, cet océan sans grèves
Dont le flot à son bord n'est jamais revenu ;
Nous les créations, îles de l'inconnu !

Victor Hugo

It will seek for targets tracing the evolution of dark energy from redshift 0.2 out to 3.5 using 3 kinds of celestial objects : 4 million luminous red galaxies (LRG), 18 million emission line galaxies (ELG) and 3 million quasars. QSO Ly-α are the most important targets. Then comes LRG, then ELG. For each petal (of 500 fibers), 10 are dedicated to Standard Stars and 40 to Sky Fibers for calibration.

- QSO (quasi stellar object) targets are divided into three categories. There are those at $z < 2.1$ (QSO-I), which are used as tracers only, and as those at $z > 2.1$ which are used for the Ly-α forest. Finally, there are false ones, objects that are *a priori* believed QSO but are actually not (which is known once observed). Quasars are extremely luminous, there spectra is very large and doesn't correspond to a star's one. They are likely to be central regions of massive galaxies surrounding a supermassive black hole, and their energy release must be caused by mass falling onto the accretion disc of the black hole. They are excellent tracer population at redshifts $1 < z < 2$. QSO Ly-α with $z > 2.1$ are especially important because it is the only opportunity for DESI to get information on dark energy at so high redshifts, and get information on the matter distribution between us and the quasar. The Lyman-α forest represents the holes in the spectrum that correspond to an absorption ray of the hydrogen (Weinberg et al. 2003) (at different redshifts), and by this mean we can compute the matter distribution of hydrogen (consequently of all baryonic matter) along the corresponding line of sight.

- LRG (luminous red galaxy) are the most luminous and red galaxies in the universe at $z < 1$ (Eisenstein et al. 2001). They are highly biased and easy to target from spectroscopic data.
- ELG (emission line galaxy) are galaxies where stars are being created. They have strong emission lines and characteristic OII and OIII doublets at a rest frame 327 nm and 500 nm, which enables to precisely get their redshift.
- Sky Fibers and Standard Stars are calibration targets, necessary to have a reference during measurements. There have to be 10 Standard Stars and 40 Sky Fibers on each petal.

3.5 Fiber assignment

Fiber assignment is the correspondance between fibers of each tile to objects. It tells which fiber is assigned to which galaxy, or unassigned (in regions where there are not enough galaxies). It depends upon fiber locations on the plate, and on the tilling pattern of the sky. One has to make prioritizations because there are several possibilities to target selection for each fiber. We discovered a broad range of different possible strategies for assignment.

4 OUR FIBER ASSIGNMENT

The original catalog of 50M galaxies has only the position and the type of the galaxy, which is either LRG, ELG or QSO. Those information were given by a previous survey, only based on colors, the redshift wasn't then computed. A python code generates the file of galaxies, simulating with appropriate probabilities each type, taking into account the fact that there are correlations between positions of galaxies (accordingly with BAO and gravition).

4.1 Introduction

Martin White started to develop a C++ code for fiber assignment, modified by Lado Samushia, then by Bob Cahn and finally by me, intern for 5 months. With Bob, we've been exploring a lot of different ideas, good, bad, sometimes useless and thus erased, to finally end after five months with this algorithm, which seems the closest to the theoretical optimal solution. We included the improvement, redistribution and update steps, to increase the number and the quality of the assignment, and at the same time to have a realistic process, in a way the code can be used. It provides a library of functions which can be easily adapted, with a "main" producing the assignment. It is well "pipelined" and someone can use it easily. The code can be found on DESI git repository with a short description of how to run it.

4.1.1 Input files

They are found in the NERSC repository (see features.txt for addresses). The produced executable (to compile with `make all` in src repository) is `assign`. The calling sequence will look like : `./assign features.txt`. An example of how running it on NERSC is provided in the `run` script, then executed by `qsub run`. Input files are :

- Parameters : all parameters used in the fiber assignment are written in this file, including addresses of other input files
- Target database : information, before the study, on all possible targets : (NERSC rep) /projects/projectdirs/desi/mocks/preliminary/objects_ss_sf0.rdzipn, created by `make_catalog`. The samples are taken from Martin's mocks and stored on NERSC at /project/projectdirs/desi/mocks/preliminary/. From those files we create a single file containing the appropriate mix of ELG, LRG, QSO, SS (Standard Stars) and SF (Sky Fibers) using the python script in git fiberassign/bin/make_catalog_starsandsky.py. In the same place is the python code to produce a mixture of galaxies without any correlations, but with the correct $\frac{dn}{dz}$
- Observations database : database constructed from the ongoing DESI observations after the data has been processed by the Spectroscopic pipeline. This database has not been designed yet
- Survey tiles : file containing the positions of all the tiles to be observed in /project/projectdirs/desi/software/edison/desimodel/0.3.1/data/footprint/desi-tiles.par
- Fiber positions : locations of the positioners in the focal plane in /project/projectdirs/desi/software/edison/desimodel/0.3.1/data/focalplane/fiberpos.txt

To change the location of other input files, one can simply change it in the features file.

4.1.2 Output files

They are produced in a directory defined in features.txt, in the format `tile54.fits` for example for the 54th tile. There are therefore 10666 such binary files. They consist in 5000 lines (fibers) with the following columns :

- fiber: [0-4999]
- positioner: [0-4999]
- number of available objects
- ID of available objects
- objtype : ELG, LRG, QSO, SKY, STDSTAR, GAL, OTHER (-1 if the fiber isn't assigned)
- targetid : unique target identifier to get back to target selection info
- desi target : 64 bit mask of targeting info (not yet)
- ra : degrees [0-360]
- dec : degrees [-90 - +90]
- xfocal : mm from center in positioner coordinate system
- yfocal : mm from center in positioner coordinate system

4.2 Source files

Source files are called file.h and file.cpp and are, in the inscreasing dependency order :

- misc : a home-made library of structures (and functions on them) needed to manipulate concerned data, but independent of them. There are pair (of int), List (of int), Table, Cube, and timing, printing, string conversion, error report items

Kind	Id	Priority	Nobs	Density ($obj \cdot deg^{-2}$)
QSO Ly-α	0	1	5	50
QSO Tracer	1	1	1	120
LRG	2	3	2	300
ELG	3	4	1	2400
Fake QSO	4	1	1	90
Fake LRG	5	3	1	50
Standard Star	6	5	1	140
Sky Fiber	7	6	1	1400

Table 1. Characteristics of galaxy samples as set in make_catalog_rnc.py

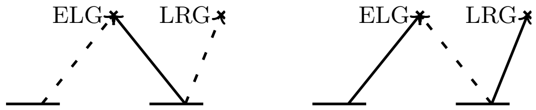


Figure 3. Improve

Figure 4. Redistribute

- collision : used to build polygons/circles representing positioners and compute collision checking between them
- features : carries all input parameters
- structs : structures of the manipulated data
- global : main high-level functions and algorithms used in the program to collect input information, assign fibers, print statistics and output the assignment. Important ones are described further
- main : neat and quickly understandable code that sums up all the steps of the assignment

4.3 Parameters

Here are some features on input galaxies simulated catalog :

A number of parameters are defined in features.txt :

- Input and output directories
- Output = true, whether one wants to release the output
- Randomize = false, whether randomizes order of plates in making plans
- Pacman = false, whether selects only spectrometers 0, 1, 2, 7, 8, 9 of the pacman
- Npass = 5, number of passes
- MaxSS = 10 ; MaxSF = 40, number of fibers assigned to SS and SF on a petal
- PlateRadius = 1.65 m, radius of the plate
- InterPlate = 0, minimal number of plates between two observations of the same galaxy
- Analysis = 7, tile distance for getting the information from previously observed tiles
- InfDens = false, simulate infinite density of SS and SF
- TotalArea = 15789.0 deg^2, total area of the sky considered (useful to do statistics)
- invFibArea = 700, inverse area, in deg^2, accessible to a fiber (which is also the fiber density for a deg^2)
- moduloGal = 1, if 2 for instance, reads only one celestial object over two in galFile. Reads all the objects iff 1
- moduloFiber = 1, same when reading fiber positions
- PatrolRad = 6.0 mm, maximum distance, on plate coordinates, that allows a fiber to observed a galaxy
- Collision = true, whether we want to allow collisions, to compute the collision rate for instance, and very practical when one wants to make quick tests, to run the fiber assignment faster
- Exact = true, whether one wants exact collision checking (exact geometry of components) or just circles
- AvCollide = 3.2 mm, in case of no exact collision checking, limit distance between two galaxies for their positioners to collide (so that we have the same collision rate than with exact geometry)
- Collide = 1.98 mm, minimum distance allowed on plate projection of two assigned galaxies on the same fiber (optimizes collision checking)
- NoCollide = 7 mm, maximum distance between two galaxies for the collision of their corresponding galaxies (optimizes collision checking)
- NeighborRad = 14.0 mm, maximum distance to consider that two fibers are neighbors
- PlotObsTime, etc, whether we want to plot some information into output files
- Verif = false, whether we verificate that the assignment is sane (no collision, sane mapping, etc...)

4.4 Classes and structures

Classes and structures are built to be independent of each other, flexible, quickly understandable, logical, and with no redundant information as much as possible.

4.5 Functions

In algorithms, j stands for a plate, k for a fiber, (j,k) for a tile-fiber (combination of a tile and a fiber), p for a petal, g for a galaxy (object). They are integers in the code. Arguments of functions are not all written.

4.5.1 Improvement functions ideas

Here we present the three ideas of improvement of this first assignment, which increase the number of used fibers and the quality of the assignment. It represents the body, the main purpose of our work. They need to be somewhere ”orthogonal”, so that they don’t loose efficiency when we call one of them after an other. Those three ideas are coded into the three following functions :

Improve : the idea, illustrated on Figure 3 is to take an unassigned fiber and to look at the galaxies it can reach and that are already assigned. We then look at the second fiber, which is observing this galaxy and try to reassign it to an other galaxy, and assign the first fiber to the first object. It is powerful because we use it when making a plan, and so the two fibers can (and are almost always) from different plates, which can be arbitrarily separated.

In **redistribute**, illustrated on Figure 4, we look at an unassigned fiber, and try to make the observed object observed by and other fiber, which can come from a different plate. It ”anneals” the assignment. In this way, it doesn’t improve anything, but we then

Structure name	Meaning	Description
element	Element of a geometric figure	Either a set of juxtaposed segments, or a circle
polygon	Polygon of segments + circles	Set of elements
Feat	Features/Parameters	Carries all needed parameters defined in features.txt and some other computed in the input reading process
PP	Plate Parameters	Carries locations of positioners on the plate, spectrometer linking, and neighboring fibers information
onplate	Plate coordinates	Used for coordinates in the focal plane, in mm
plate	A plate	Locations in the sky of the tile in terms of a unitary vector derived from RA (right ascension) and DEC (declination). Carries also the identifier of the tile, its pass, and the available galaxies it is able to reach. Plates is vector of plate, and carries all information on tiles
galaxy	A galaxy	Information on a celestial object (which can be something other than a galaxy) : kind which corresponds to Table 1, a position in the sky (in two differents ways), and the available tile-fibers that can observe it. Gals is vector of galaxy and carries all information on objects
Assignment	An entire assignment	Carries mapping of tile-fibers to galaxies on a table, and other redundant information for optimization

Table 2. Classes and structures

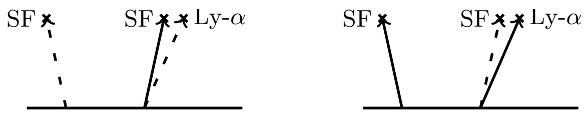


Figure 5. Improve from kind

apply the improve function again, which is more efficient. The idea is the improve, then redistribute, then improve again, redistribute etc. If there is no redistribution between a new improve, the improve is almost useless. Furthermore, among possible fibers, we assign it to the least used petal.

Improve from kind, the third idea, on Figure 5, is takes an unassigned fiber that is able to reach a SF (or SS), assigns it, and then reassigns a fiber assigned to a SF. As the number of SF per petal has to remain the same as previously (40) we search the second fiber among the same petal of the first fiber. It's not used anymore since it becomes useless when we first use new_assign.

4.5.2 *Some inportant functions in structs.cpp*

`plate_dist` turns radians into mm on the focal plane, i.e. it is the plate scale as a function of the angle

`projection` combines a galaxy and a particular plate to give the coordinates the galaxy will have in the focal plane

`find_collision` returns the fiber number of (one of) the fiber that conflicts with tilefiber (j,k), or -1 if there is no collision

4.6 Collecting

`collect_galaxies_for_all` is multithreaded, and for each fiber of each tile, collects reachable galaxies. It uses kdTree and htmTree libraries written by Martin White. It's absolutely necessary in order to do computations in a reasonable time with supercomputers

`collect_available_tilefibers` computes, using the previous work, available tile-fibers for each galaxy (inverse map)

4.6.1 *Useful sub-functions for global functions*

`ok_assign_g_to_jk(g,j,k)` checks whether one can assign g to the tile-fiber (j,k), according to assignment rules described further

`find_best(j,k)` finds the best reachable galaxy for this fiber, according to assignment rules (not observed in two tiles separated by less than Interplate, choose the one with the higher priority, then if several ones compete, the least observed)

`assign_fiber(j,k)` tries to assign this fiber using find_best

`improve_fiber(j0,n,j,k)` uses the improve idea described on Figure 4. If the fiber k is unused, tries first to simply assign it, and if impossible, tries to reassign some used one (jp,kp) where $j0 \leqslant jp \leqslant j0+n$. Before : (jp,kp) - g ; (j,k) & gp free. After : (j,k) - g & (jp,kp) - gp. The power of this function lies in the fact that j and jp can correspond to different passes. Among all galaxies of same priority and same number of observation, it takes the one belonging to the most unused petal, to distribute ELGs

Algorithm 1 Improve_fiber(j0,n,j,k)

if k is not assigned **then**
    try to assign running assign_fiber(j,k)
    if k couldn't be assigned this way **then**
        initialize a set of variables jp, kp, g, gp
        for each galaxy g available to (j,k) **do**
            if it's possible to assign g with k **then**
                for each chosen tile-fibers (jp,kp) which chose g (where $j0 \leqslant jp \leqslant j0+n$) **do**
                    $gp \leftarrow find\_best(jp,kp)$
                    memorize jp, kp, g, gp if it's a better set (gp is worthier according to priorities/observation numbers) than previous one
        if $gp \neq -1$ **then**
            Unassign $(jp,kp) \longleftrightarrow g$
            Assign $(j,k) \longleftrightarrow g$
            Assign $(jp,kp) \longleftrightarrow gp$

4.6.2 Assigning making a plan

In the simulated catalog, we know all information on objects (if a QSO is a real one for example). During the real study, one won't have access to this information prior to the observation plus analysis time. Thus, in the code, we have to simulate that we have this information only when we have at least once observed the object, and when the analysis of received data tells which exact type it is (fake or not for ELG for example). Algorithms do a plan with only types of objects (QSO, LRG, ELG) available in the catalog. A plan consists in an assignment on a set of consecutive tiles (from 0 to 2000 for instance). The assignment has better results when there is only one plan.

The argument "next" (integer) in following functions means that we treat all next "next" plates in the plan we make. If it is -1, it will deal with all left plates. For example, if $next\_plate = 10$ and one launches a function with $next = 100$, this function is going to do his job on plates from 10 to 110.

`simple_assign` makes a first simple assignment plan : for each fiber, assign it to the best available galaxy

Algorithm 2 Simple_assign(j0,n)

```
for each plate j from j0 to j0+n do
    for each fiber k in a random order do
        try to assign_fiber(j,k)
```

`new_assign` makes a first assignment plan trying to first assign QSO, then LRG then ELG (neither SS nor SF). This way, a QSO can't be lost because of a collision with an object of an other kind

Algorithm 3 New_assign_fibers(j0,n)

```
for each plate j from j0 to j0+n do
    for each petal p of this plate, in a random order do
        for each fiber k of this petal, in a random order do
            Try assign_fiber(j,k) only allowing QSO
        for each fiber k of this petal, in a random order do
            Try assign_fiber(j,k) only allowing LRG
        for each fiber k of this petal, in a random order do
            Try assign_fiber(j,k) only allowing ELG
```

`improve` improves the plan applying improve_fiber to all of the fibers involved in the plan

`improve_from_kind (kind)` for every concerned petal, tries to assign unassigned fibers to objects of kind "kind" (useful when it's SS or SF), to release an other fiber (formerly assigned to a SS or SF) that would be reassigned to a regular galaxy (kind is SS or SF when we call it). There is also the function `improve_fiber_from_kind` that does the same but on only a fiber. We don't use it anymore, since we use new_assign, which already includes potential improvement brought by this function.

Algorithm 4 Improve_from_kind (kind,j0,n)

```
for each plate j from j0 to j0+n do
    for each petal p of this plate, in a random order do
        for each unassigned fiber k do
            initialize a set of variables kp, g, gp
            for each available galaxies g of k do
                for each fiber kp of p assigned to a galaxy of kind do
                    if no conflict & g is worthier than previous one then
                        memorize kp, g, gp
            if g ≠ −1 then
                Reassign kp to g
                Assign k to a galaxy gp of kind "kind"
```

`redistribute_tf` "anneals" the distribution (so that "improve" is more efficient after) and at the same time redistributes in a way that the number of unused fibers per petal tends to be the same for every petal. It is very important to prepare the step of assigning SS and SF, in order to replace the least number of ELG and LRG for having the needed number of SS/SF

Algorithm 5 redistribute_tf(j0,n)

```
for each plate j from j0 to j0+n do
    for each assigned fiber k do
        get g, the galaxy to which (j,k) is assigned
        initialize a set of variables (jpb,kpb) to memorize values of (jp,kp)
        for each available tile-fibers of g do
            if (jp,kp) is better (j0 ⩽ jplj0 + n, unassigned, it's ok to assign, and the petal of (jp,kp) is more unused than the one of (jpb,kpb)) then
                memorize (jp,kp) in (jpb,kpb)
        if jpb ≠ −1 then
            Reassign g to (jpb,kpb)
```

`update_plan_from_one_obs(j)` updates the plan formerly made so that if for example on the plate $j - Analysis$ plates (to simulate there is a delay between observation and result on analysis) we observed a QSO which the analysis reveals fake but is still planed to be observed again further in the plan, we will unassign the corresponding fiber and try to reassign it with improve_fiber. In the input catalog we only know the basic type (QSO, LRG, ELG, SS, SF) of the galaxies. The precise type is only infered by a simulation coded in python, but is generated with randomness. So in the algorithms, we can know the nature of the object only when it has been observed once at least and analysed (this delay of "Analysis" number of plates is inserted between those two events). In a plan, we try to optimize assignment, but of course, we can't know the precise type (fake, ...) so it is possible to foresee observing for example 4 times a fake QSO. We won't do it, so after having planned, when we begin to apply it and do the real observation, if we have just previously observed a fake QSO, we will remove further observations in this plan. We then try to assign the released tile-fibers.

Algorithm 6 Update_plan_from_one_observation(j0,end_plan)

get the list of galaxies observed by this plate that are discovered fake or target
for each of those galaxies g **do**
  get the list of tile-fibers that are supposed to observed this galaxy further in the plan $j0 \leqslant jp \leqslant end\_plan$
  for each of those tile-fibers (jp,kp) **do**
    Unassign $(jp, kp) \longleftrightarrow g$
    run improve_fiber(j0+1,end-j0,jp,kp) (tries to reassign (jp,kp)

`assign_sf_ss(j)` just before the observation of a tile, we assign SF and SS, replacing ELG then LRG if necessary

Algorithm 7 assign_sf_ss

for each petal p of this plate, in a random order **do**
  try to assign unused fibers to 40 SF and 10 SS
  if not enough SS and/or SF **then**
    try to replace some ELG for SS
    try to replace some ELG for SF
  if not enough SS and SF **then**
    try to replace some LRG for SS
    try to replace some LRG for SF

`assign_left(j)` just before the observation of a tile, for all unassigned fibers, tries to assign them to a galaxy in reach that is supposed to be observed later, and remove the later observation. The goal is to observe and thus analyse objects as soon as possible, not waiting further in the plan.

4.6.3 Displaying results

Histograms are written in Tikz format.

`results_on_inputs` displays some statistics on input files (histogram of number of objects in range of a fiber, histogram of number of tile-fibers which can observe a given object, histogram of redshifts by type)

`display_results` provides the tex-formated results to make Table 3 and writes histograms one can see further in this document. Data are written into .dat files, which can be compiled by the Latex module Tikz to create figures like histograms present here

`write_FAtile_ascii(j)` writes output of the assignment in ASCII format

`fa_write(j)` writes output of the assignment in FITS format

`pyplotTile(j)` writes a file tileX.py from which (executing `python tileX.py`), one can produce the pdf graphic colored plot of the tile. In it, are also ploted objects requiring more observations. An already twice observed LRG won't appear for example

`verif(j)` verification which checks several features of the assignment to know if it is sane (no collision, regular mapping, etc...)

4.7 Rules of assignment

- a tile-fiber is only assigned once
- no collision between fiber positioners
- two observations of a QSO or an LRG can't be separated by less than `InterPlate` plates

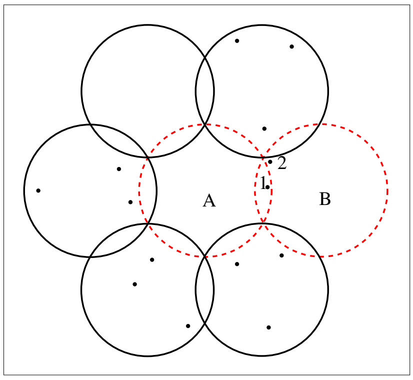


Figure 6. Choice by density : A is chosen to observe 1

- in last pass, only ELG, SS and SF are considered
- basically, we choose best galaxy among the ones which are available and observed less than maxgoal(kind) times. Among them, we take those with highest priority, and among them we then choose the least observed one
- between an already observed Ly-α (observed less than 5 times) and an unknown QSO, we choose the Ly-α
- exactly 10 and 40 fibers are assigned to SS and SF in each petal
- SS has priority over SF
- SS and SF don't have limit on the number of observations
- our policy is : while improving the assignment, never unassign an object without reassigning it right then

4.8 Choice by density

The assignment could include the idea well described in (Morales et al. 2012) which consists in, when you have the choice of several fibers to observe an object, you choose the one which has the least density of (weighted) remaining available objects in reach. This way, "busy" fibers are more available. Of course, it will introduce a bias in the survey, and since it doesn't improve results (it's likely because redistribution/improvement functions already make this optimization) we can avoid it. An illustration taken from (Morales et al. 2012) is in Figure 6.

4.9 Collision problem

The exact geometry as built in our program can be seen on Figure 7, which consists in two circles and several segments, optimized to reduce the number of computations when checking for collisions. The black dot point is the center of the positioner, and other dots are objects in reach.

4.10 Algorithm

`main` proceeds by :

- loading input files : features of galaxies to make F (features), reading in the galaxies catalog to make G (galaxies), the positions of the fibers on the plate to make pp (plate parameters), the plate centers and positioner locations to make P (plates)
- collect available galaxies for each fiber for each plate, and compute the inverse map (available fibers for each galaxies, useful for improvement functions)

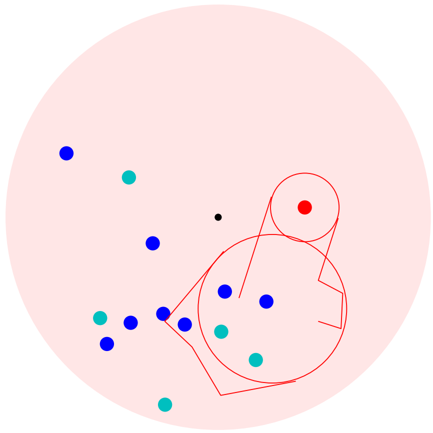


Figure 7. Geometry of a fiber positioner, with fiber holder and central body oriented

- before starting the study, make a plan of assignment
- begin the study, and update the plan as we get information from objects

4.10.1 *Redistribution-Improvement step*

Assigning only QSO, LRG and ELG, not SS and SF when making the plan leads to have a lot more degrees of freedom for the improvement functions, which is then much more efficient. Redistributing unused fibers by petal occupancy permits to replace the least number of ELG when assigning SS and SF just before an observation. We try by this mean, as long as possible, to have 50 free fibers per petal before starting observations.

When applying the plan, sometimes redistributing-improving the plan is very efficient, it updates the further plan (compared to the initial done plan, just before starting the study), taking into account that some new fibers have been released and that we have new information on precise kind of objects. We also do improvement/redistribution process during the 5-years observation time.

Here is presented, on Figure 8, histograms of unused fibers per petal, during redistribution/improvement process when making the first plan. The goal is to have 50 unused fibers per petal before starting observations. On Figure 9 are the histograms but with more redistributions-improvements when making the plan. Figure 10 is the same with a linear y axis.

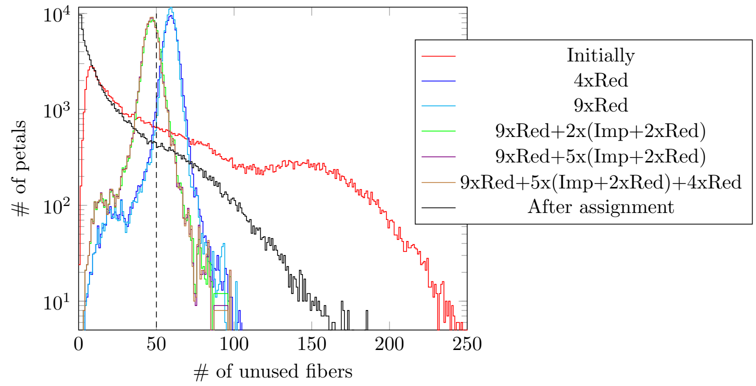


Figure 9. Unused fibers, at different steps, with more improvements

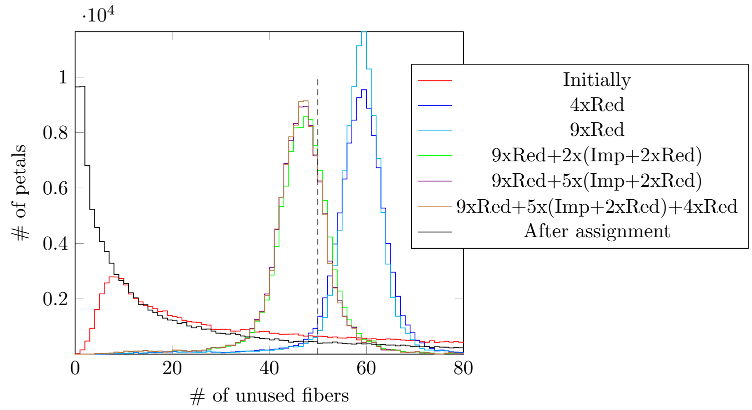


Figure 10. Same, but in linear scale

4.11 Pseudo-code of the algorithm

Algorithm 8 Reference assignment in main program

Phase I - Make a plan for all plates

```
Run, globally, on the list plates from 0 to last :
"New" assign fibers
Redistribute (several times)
Improve + 2x Redistribute (several times)
Redistribute (several times)
```

Phase II - Observation time

```
Beginning of the study
for each plate j of the plan, in order do
    Assign_sf_ss
    Assign_left
    Possibly "pyplot" the planned observed tile
    Real observation is here
    Update information collected on the previously observed Analysis'th past tile with update_plan_from_one_obs
    If we are in the plate 100, or 300, or 800, or... 2xRedistribute + Improve + Redistribute (on all the plan of remaining tiles)
```

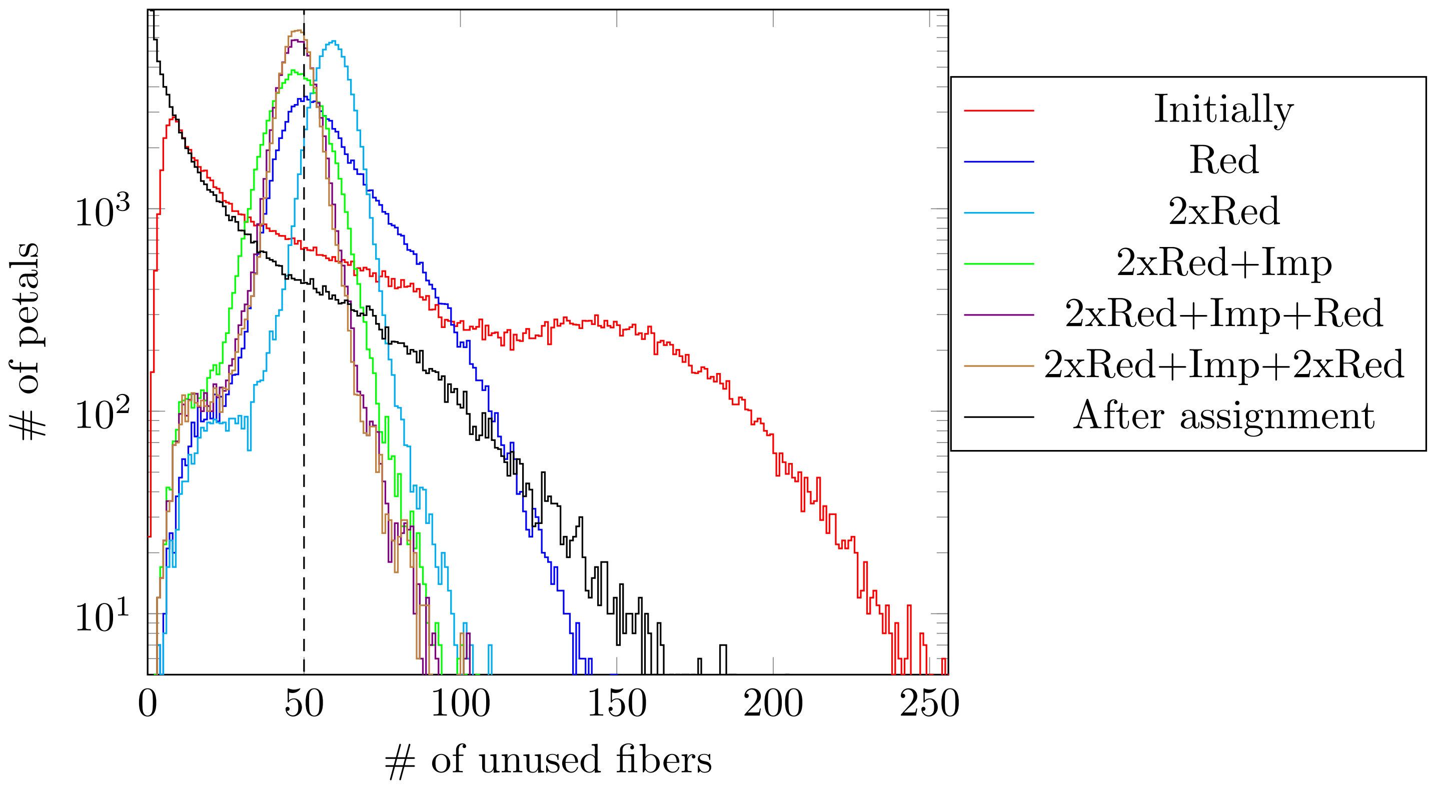


Figure 8. Unused fibers, at different steps

A standard output display of an execution looks like that (for assignment plan and applying):

```
# Read 71,998,144 galaxies from
/desi/.../objects_ss_sf0.rdzipn
# Read 10,666 plate centers from
/desi/.../desi-tiles.par and 5000 fibers from
/desi/.../fiberpos.txt
# Start building HTM tree at 13.8 s
# ... took : 25.5 s
# Begin collecting available galaxies
# ... took : 31.6 s
# Begin computing available tilefibers
# ... took : 1 mn 12.5 s
# Start assignment at : 2 mn 27 s
# Begin new assignment :
50,518,743 assignments on all left next plates
# ... took : 18 mn 38.3 s
# Begin improve :
565,801 more assignments (1.120 % improvement)
# ... took : 36.5 s
# Begin redistribute TF :
1,760,465 redistributions of couples of TF
# ... took : 23.3 s
# Begin improve :
206,948 more assignments (0.405 % improvement)
# ... took : 30.2 s
# Begin redistribute TF :
1,575,337 redistributions of couples of TF
# ... took : 23.1 s
# Begin improve :
92,508 more assignments (0.180 % improvement)
# ... took : 26 s
# Begin real time assignment at 23 mn 24.5 s
- Plate 0 : 550 not as - 3852 unas & 2772
replaced
- Plate 1 : 98 not as - 3784 unas & 2773 replaced
- Plate 2 : 104 not as - 3985 unas & 2844
replaced
- Plate 3 : 121 not as - 3602 unas & 2563
replaced
- Plate 4 : 105 not as - 3773 unas & 2685
replaced
- Plate 5 : 85 not as - 3797 unas & 2853 replaced
- Plate 6 : 145 not as - 3537 unas & 2514
replaced
- Plate 7 : 92 not as - 3964 unas & 2856 replaced
- Plate 8 : 123 not as - 3714 unas & 2578
replaced
- Plate 9 : 143 not as - 3543 unas & 2378
replaced
- Plate 10 : 148 not as - 3528 unas & 2472
replaced
- Plate 11 : 86 not as - 3507 unas & 2885
replaced
# Begin redistribute TF :
1,160,465 redistributions of couples of TF
# ... took : 23.3 s
# Begin improve :
106,948 more assignments (0.205 % improvement)
# ... took : 30.2 s
```

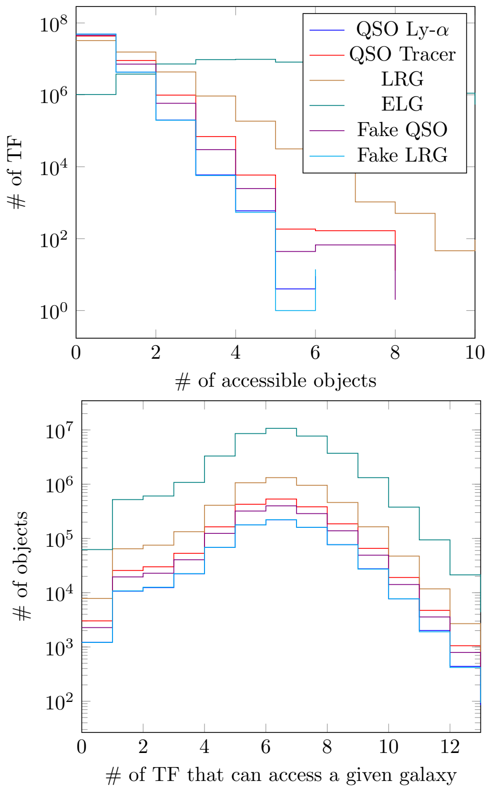


Figure 11. Available galaxies for a fiber on a tile and available tile-fibers for a galaxy (by kind)

```
- Plate 12 : 48 not as - 3128 unas & 2475
replaced
- Plate 13 : 96 not as - 3407 unas & 2485
replaced
...
```

After an observed tile, we print the number of unassigned fibers on it, the number of further objects unassigned because they were analysed as fake, and the number of the corresponding tile-fibers that have been successfully reassigned.

5 RESULTS

5.1 Results on the input catalog

On Figure 11 are histograms of number of accessible objects for a tile-fiber, and the inverse. On Figure 12 is the histogram of number of plates which can access an object, for the regular and pacman versions, it is important for understanding final results. Figure 13 gives the repartition of redshifts.

5.2 Results of the assignment

We run the program with all information (prior knowledge of information on fake, target, etc) to compare with realistic simulation, this result is interesting to see the effect of this prior unrealistic knowledge.

We have taken a reference strategy, which was a trade-off between computation time and quality of results. We use improvement functions in a way such that they are still efficient given

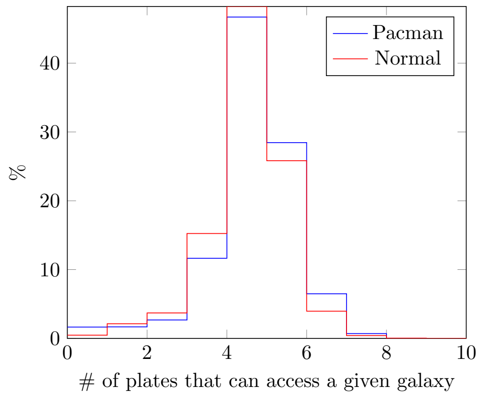


Figure 12. Available plates for a galaxy (without 5th pass)

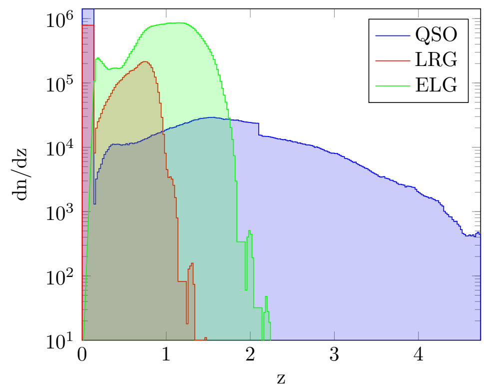


Figure 13. Distribution of objects as a function of redshift

the time they take (because the more improvement execution we launch, the less they are efficient).

We tried to run it with an ”ideal” infinite density of SS and SF, which only led to an 0.5% increasing of the number of observed ELG. Increasing the number of SS and SF in the catalog could lead to this improvement, but Bob already took all SS and SF currently known by the cosmologists community.

5.3 Sum-up of assigned galaxies

The weighted score of a certain kind is defined by :

$$Score(kind) = 100 \cdot \frac{\sum_{g \in kind} obs(g)}{\sum_{g \in kind} goal(g)}$$

where $obs(g)$ is the number of times g is observed.

A sum-up table of general results on assignment is provided on Table 3.

Furthermore, there are 51,044,452 assignments in total (95.7143 % of all fibers).

5.4 Free fibers

Here are the histogram of petals as a function of free fibers and the number of free fibers for each plate, in inscreasing order on Figure 14. One can see on the second one the 5 different passes.

On Figure 15, one can see the proportion of observed objects

	Times observed										
	0	1	2	3	4	5	Total	Fiber used	Once observed	Observed %	Weighted %
QSO Ly-α	0	1	5	12	19	10	49	180	49	99.151	72.206
QSO Tracer	1	118	0	0	0	0	119	118	118	99.141	99.141
LRG	13	42	243	0	0	0	298	528	285	95.505	88.434
ELG	480	1,930	0	0	0	0	2,411	1,930	1,930	80.054	80.054
Fake QSO	0	89	0	0	0	0	90	89	89	99.139	99.139
Fake LRG	2	47	0	0	0	0	50	47	47	95.792	95.792

Table 3. Densities (objects/deg^2) as a function of # of observations (with total), % observed, weighted %

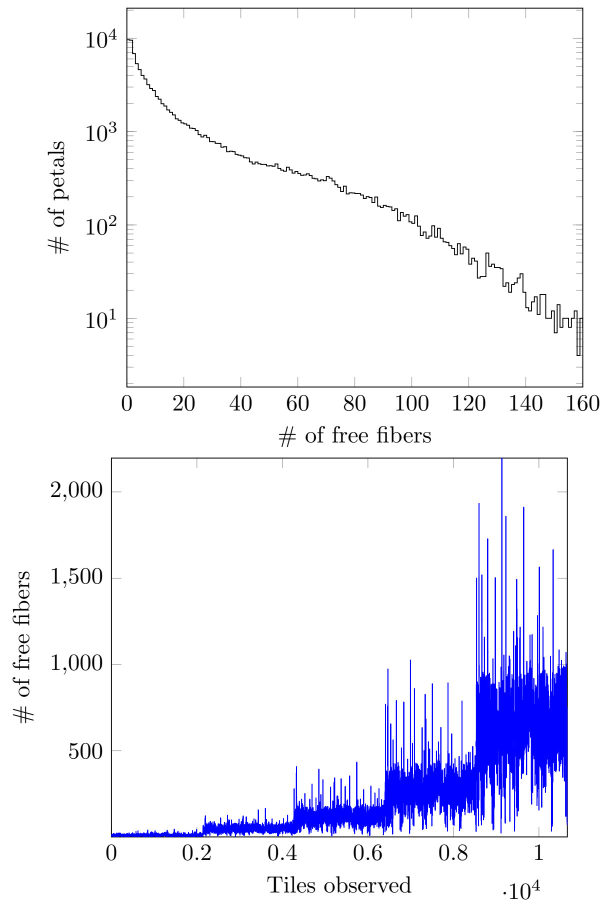


Figure 14. # of petals with that many free fibers and free fibers as a function of time (plates)

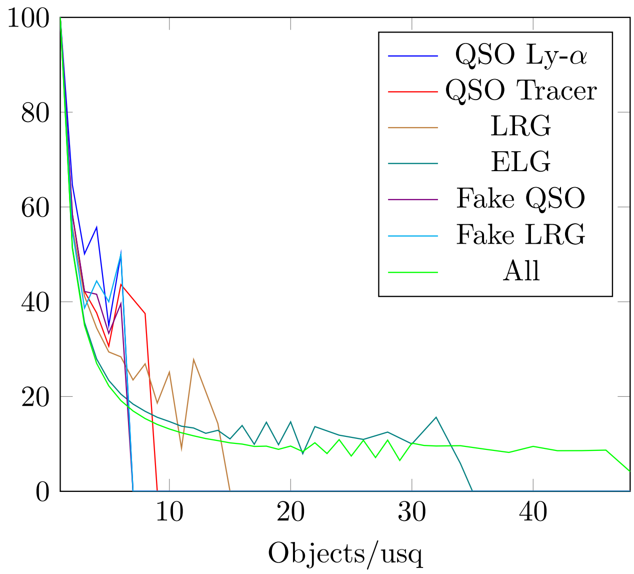


Figure 15. % of observed galaxies as a function of objects density

as a function of their density. ”usq” stands for unity of square degrees, it’s the area of the sky reachable by a single fiber. As expected, when density inscreases, proportion of observed objects decreases.

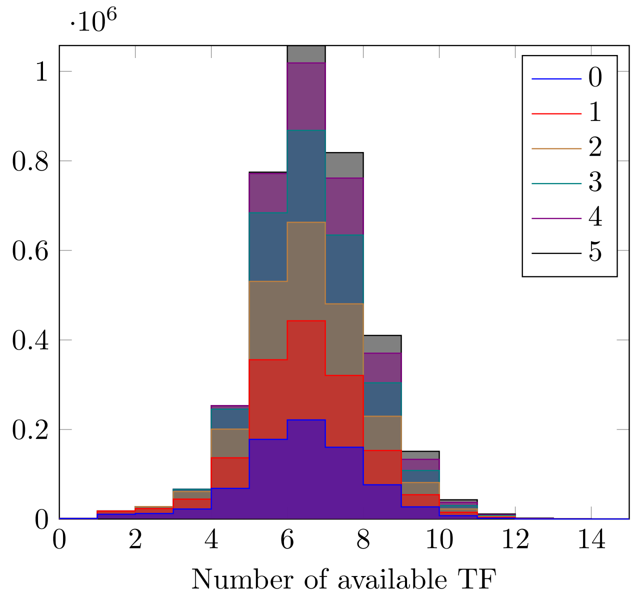


Figure 16. # of QSO Ly-α (with their number of observations) as a function of available tile-fibers

5.5 Results on Ly-α particularly

Figure 16 gives the histogram of proportion of assigned Ly-α and the number of observations as a function of available tile-fibers.

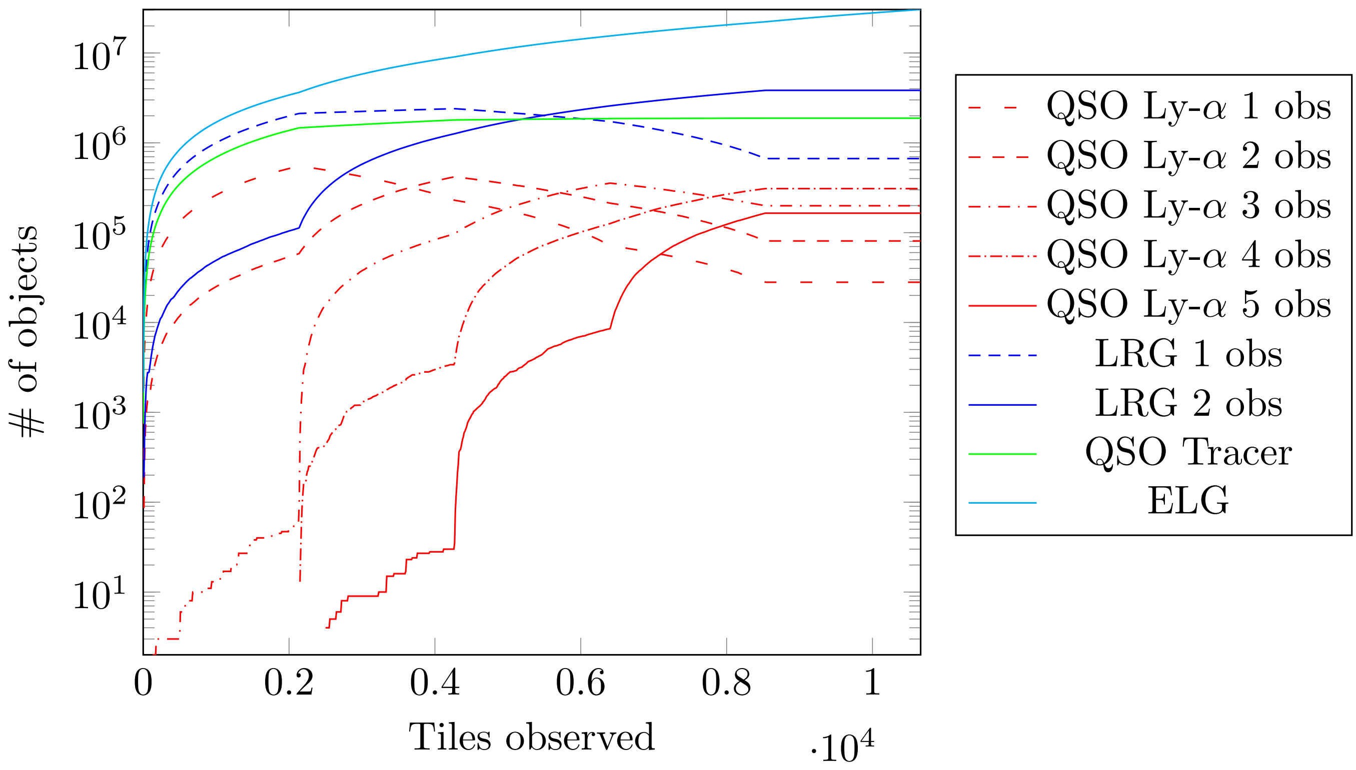


Figure 18. Observed galaxy kind as a function of time (plates seen)

5.6 Study on evolution over time

5.7 Tiling strategy

The tiling strategy is important and should separate overlapping tiles for more than "Analysis". This way, the assignment can be updated in order to avoid observing twice the same object. The results are then improved in a non negligible rate : compared to a simple assignment, we increased the proportion of observed objects (+8% for ELG, +3% for LRG, +5% for weighted QSO Ly-α). Fiber assignment results may help to choose the best strategy.

5.8 To better understand bias from positioners

In the features file the "Exact" boolean indicates whether we choose the exact geometry of the fiber positioners. If not, collisions are computed with only circles at the position of the fiber holder, of radius $AvCollide$, which we calibrate such that we get the same collision rate than with the exact geometry ($\sim$ 10.8% of fibers collide). If the exact geometry is on, the modules collision.h and collision.cpp are used to check. They then build sets of circles or segments, and check collisions between those belonging to the first fiber positioner to the ones belonging to the second one. Computationnally cheap methods were absolutely necessary and used for segment-segment and circle-segment checking, because checking functions are called a lot of time, and optimization of those methods lead to considerably decrease total computation time. That is why we use circle and segments, not only segments : creating an arc with segments needs a lot of them, but only one circle.

Figure 19 is the histogram of distances between two galaxies, in a collision case, and its integral. It will be very important to understand the introduced bias when time of data analysis will come, in 9 years.

(Makarem et al. 2014) is an article about the problem of reaching final position for the positioner without colliding a neighboring one.

5.9 Tile ploting

The function pyplotTile builds a tile.py file, which can be executed manually with `python tile.py` to create the pdf plot. The cover page of this report contains the plot of the 1000th tile. Only objects that would need at least one remaining observation are projected on the plate and ploted. The function has to be called just before real observation time, otherwise, if it is called at the end for example, information on number of observations of an object would be the one that we have at the end, and not the ongoing one.

A grey circle is ploted when there is an unassigned fiber at some place, not to mix up with fiducials (totally white), places where there is physically no fiber. Watching at a dozen of tile plots, one can convince oneself that the assignment is very close to the optimum.

5.10 Mathematical optimum

Have we reached the optimum ? We have to master, understand, results we get (of Table 3) from mathematical computations.

Imagine a portion S of the sky of the size the reacheable area by a single fiber. Let i be the indice for different kinds (Ly-α , fake QSO, ... except SS or SF). Let λ_i be the densities of the kinds of objects in S. Let N be the indice for the portions of sky with exactly N objects (when we take an area of size S). Let f_N be the fraction

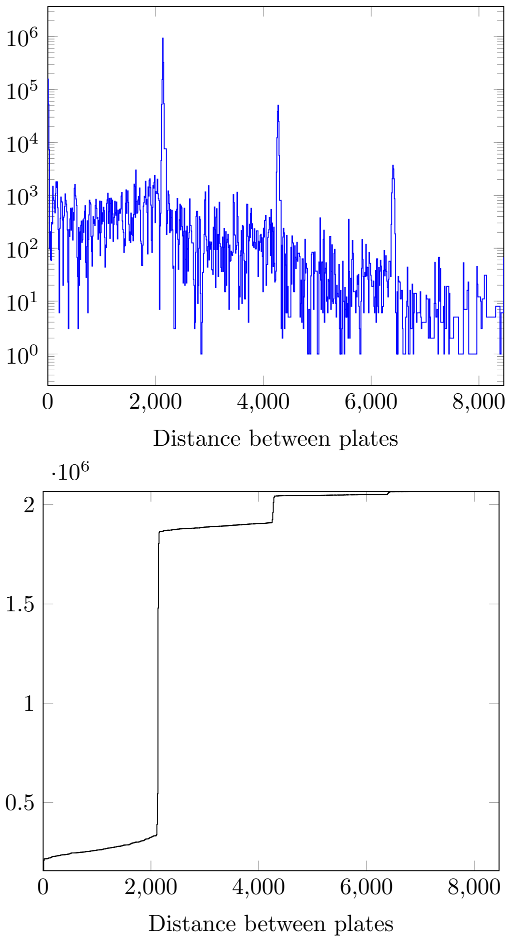


Figure 17. Histogram of distances (in number of plates) between two consecutive observed QSO Ly-α and integral

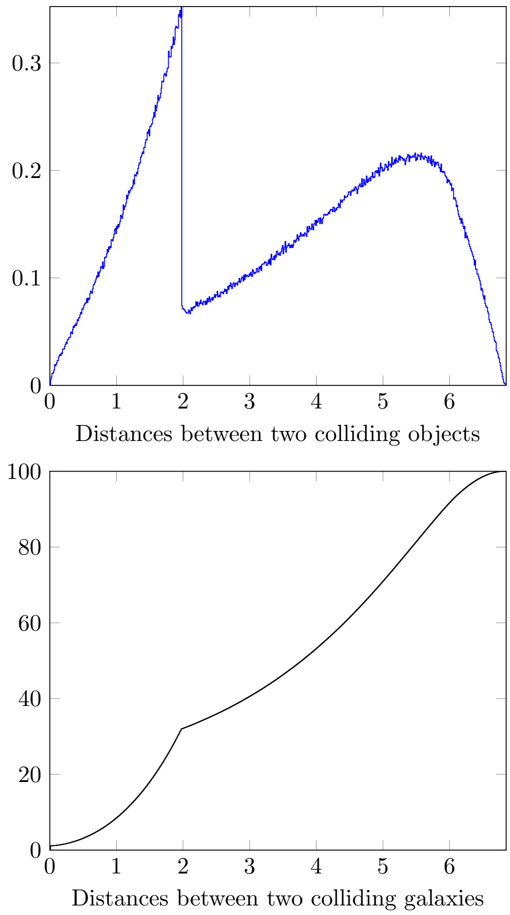


Figure 19. Histogram of distances between galaxies in collision case, and integral

of the sky which density corresponds to N. Let $poisson(x, n) = \frac{x^n e^{-x}}{n!}$ be the Poisson distribution. Then, for a given ELG in S named e, the probability p to observe it is :

$$\begin{aligned}
p &= \sum_{1 \leqslant N \leqslant 10} f_N \sum_{0 \leqslant A \leqslant N-1 \text{ and } 1 \leqslant B \leqslant \infty} p(\text{A objects having priority over an ELG}) \cdot p_1 \\
&= \sum_{1 \leqslant N \leqslant 10} f_N \sum_{0 \leqslant A \leqslant N-1 \text{ and } 1 \leqslant B \leqslant \infty} p(A) \cdot p(\text{B ELG}) \cdot p(\text{e is chosen among the N-A firsts}) \\
&= \sum_{1 \leqslant N \leqslant 10} f_N \sum_{0 \leqslant A \leqslant N-1} \sum_{1 \leqslant B \leqslant \infty} p(A) \cdot \frac{\lambda_{ELG}^B e^{-\lambda_{ELG}}}{B!} \cdot \frac{N-A}{B} \\
&= I \cdot \sum_{1 \leqslant N \leqslant 10} f_N \sum_{0 \leqslant A \leqslant N-1} p(A)(N-A)
\end{aligned} \tag{12}$$

where $I = e^{-\lambda_{ELG}} \int_0^{\lambda_{ELG}} \frac{e^{\lambda}-1}{\lambda} d\lambda$

and $p_1 = p(\text{B ELG and e is chosen among the N-A firsts})$

Also we define the set L of objects with priority against ELG, which are Ly-α , fake QSO, QSO target, LRG and fake LRG. Then, for $i \in L$:

$$p(A) = \sum_{n_i st \sum_i goal(i) n_i = A} \prod_i poisson(\lambda_i, n_i)$$

One can use the same method to compute theoretical percentage of observed QSO and LRG. Unfortunately, the computed numbers weren't credible, it must be caused by the fact that this calculus doesn't take into account that there can be overlaps between fiber observational areas.

5.11 Future of the fiber assignment

The code will have to be updated to take into account further caracteristics, which are being defined as DESI requirements are precised. For example, the weather during observations is not predictable and assignment will then have to be adapted everyday : QSO will be observed during not cloudy nights. Bob Cahn is going to take care of it. I've been keeping contact with Bob during this summer to help solving a few coding problems and will keep being in touch with people wanting precisions on my work.

One of the consequences of the fiber assignment that has to be investigated is to what extent it introduces an experimental bias. The fact that a lot of rules are used in the assignment is likely to introduce a bigger bias. It will be an unexplored problem for data analysts, because in BOSS and further studies, since fiber assignments were much simpler and fibers didn't have positioners, introduced bias were very different.

Lado Samushia is beginning to explore this problem, trying to understand the bias comparing correlation functions given with this fiber assignment and unbiased one.

ACKNOWLEDGEMENTS

I'm extremely grateful towards Bob Cahn, who is not only an excellent mentor and professor, but also a very optimistic and motivating researcher, making our daily work very enjoyable. I thank him having entrusted me jointly realizing the fiber assignment, which is a key part of the DESI achievement. Our collaboration had been very exciting and put me to a high degree of responsibility for an intern. Thanks to him, I had the broader researcher's work overview I could have hoped.

I thank Stephen Bailey, Joe Silber, Florian Beutler, Andreu Font, Julien Guy, generous with their time to help me resolving coding problems and with who I had very interesting discussions about cosmology and science. I thank Ted Kisner and Rollin Thomas for having helped me to use NERSC supercomputers. Thanks a lot to Anytra Henderson, Anthony Spadafora and Ben Ortega for their help and kindness on administrative issues. I thank Adel Bilal, Pierre Binétruy and George Smooth who got me in touch with Bob Cahn. I also thank the supercomputer Edison for its breathtaking computational power, which made me amazed every one day even after five months.

The collaboration with other researchers of the LBNL was tight and friendly. The cosmology group combines a high scientific level, cutting edge resources, a wonderful environment (on Berkeley's hill) and a nice social atmosphere. I attended most of the talks given at the lab, which gave me a state-of-the-art and almost exhaustive outlook of today's research on cosmology. This internship had been a professional and social extremely enriching experience. I would advise any following normalien to do its internship there if he has the opportunity.

REFERENCES

Amendola L. et al., 2013, Living Reviews in Relativity, 16, 6
Benitez N., Gaztanaga E., Miquel R., Castander F., Moles M., et al., 2009, Astrophys.J., 691, 241
Bianchi D., Gil-Marín H., Ruggeri R., Percival W. J., 2015, ArXiv e-prints
Einstein A., 1917, Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften Berlin
Eisenstein D. J., et al., 2001, Astron.J., 122, 2267
Font-Ribera A., McDonald P., Mostek N., Reid B. A., Seo H.-J., Slosar A., 2014, 5, 23
Hubble E., 1929, Proceedings of the National Academy of Sciences of the United States of America
Kaiser N., 1984, APJL, 284, L9
Kaiser N., 1987, Mon.Not.Roy.Astron.Soc., 227, 1
Levi M., et al., 2013, .
Makarem L., Kneib J.-P., Gillet D., Bleuler H., Bouri M., et al., 2014, Astron.Astrophys., 566, A84
Morales I., Montero-Dorta A. D., Azzaro M., Prada F., Sánchez J., Becerril S., 2012, MNRAS, 419, 1187
Mortonson M. J., Weinberg D. H., White M., 2013, .
Perlmutter S. et al., 1997, in Bulletin of the American Astronomical Society, Vol. 29, American Astronomical Society Meeting Abstracts, p. 1351
Riess A. G. et al., 1998, The Astronomical Journal, 116, 1009
Schlegel D. J. et al., 2009, ArXiv e-prints
Spergel D. et al., 2015, ArXiv e-prints
Tauber J., ESA, 2004, Advances in Space Research
Weinberg D. H., Davé R., Katz N., Kollmeier J. A., 2003, in American Institute of Physics Conference Series, Vol. 666, The Emergence of Cosmic Structure, Holt S. H., Reynolds C. S., eds., pp. 157–169
Weinberg D. H., Mortonson M. J., Eisenstein D. J., Hirata C., Riess A. G., Rozo E., 2013, ., 530, 87
Wiltshire D. L., 2008, in Dark Matter in Astroparticle and Particle Physics, Klapdor-Kleingrothaus H. V., Lewis G. F., eds., pp. 565–596