■ Scientific Justification

The majority of the baryons in the universe are found in the intergalactic medium (IGM) (e.g. Werk et al. (2014), MORE). The properties of this important reservoir of matter are most directly measured via absorption in the spectra of UV-bright background QSOs. Studies using this method have found numerous IGM-galaxy proximity effects, such as increasing absorption linewidth (e.g. Wakker & Savage 2009), column density (e.g. Rudie et al 2012), and metallicity (e.g. Kacprzak et al YEAR?) anti-correlating with the distance to the nearest bright galaxy.

In order to move forward from these results however, we need to confront several shortcomings that are nearly universal among studies of the IGM-galaxy connection.
1) Sample size: MORE, CITATIONS
2) Reproducibility: By definition, a decision must be made matching a galaxy to an absorption feature, but it is often unclear or arbitrary how this matching is done. Some (e.g. CITATION) simply picked the closest neighboring galaxy to the detected IGM absorber, while others (e.g. CITATION) try to take galaxy size and other properties into account as well. No method has yet emerged as a favorite.
3) Incompleteness: the completeness of known galaxies decreases sharply with redshift, which can further complicate choosing the galaxy to associated absorption with. For example, Mathes et al. (2014) and Werk et al. (2014) are only complete to $\sim L^*$ (at $0.12 < z < 0.67$ and $z \sim 0.2$, respectively).

In attempt to combat these issues, we have begun a large scale survey of nearby COS sightline-galaxy associations, restricting ourselves to $cz \leq 10,000$ km/s, where we have assembled a galaxy catalog complete to $L^* \sim 0.1$. To address the reproducibility issue, we have developed a likelihood parameter to automatically choose which galaxy to associate an absorption line with.

The UV initiatives undertaken by the Cosmic Origins Spectrograph on HST have produced a wealth of high resolution and high signal-to-noise spectra that are ideal for a wide variety of IGM studies. However, the identification and measurement of spectral features is a major undertaking, and limits the viability of large-scale studies using archival HST data. To combat this we have developed a pipeline that automates the IDing of targets, which will allow us to produce a sample of hundreds of spectra and thousands of absorption lines. As the final step of our proposed program, we will create a legacy data archive of identifications and measurements of COS sightlines, complete with likelihood-based galaxy associations as well.

Archival data

At the time of writing 275 COS QSO targets have been observed by HST and are publicly available on MAST (at $z > 0.03$ and S/N$\geq$10, the minimum to be useful for our purposes). X are nearby, in the redshift range $0 \leq z \leq 0.1$, X are mid-redshift, $0.1 \leq z \leq 1$, and X are high-redshift, at $z \geq 1$. Of those, we have already reduced and identified 100.

Line Identification

Description of the identification code and method.

Nearby Galaxy Dataset We have made extensive use of the NASA Extragalactic Database to construct a dataset of galaxies in the $0 \leq cz 10,000$ km/s redshift range.

Galaxy Matching: Likelihood Method

Studies of the circumgalactic medium, or the IGM-galaxy connection, generally rely on matching absorption measured in the spectra of background QSOs to a nearby foreground galaxy. The methods used for matching a galaxy to an IGM absorber are universally ad-hoc, however, and range from picking the nearest galaxy in physical impact parameter, to some combination of impact parameter, size, and judgement. In French et al. (2016, in prep) we are introducing a reproducible, likelihood-based method to streamline these types of decisions. We define the likelihood as follows:

$$\mathcal{L} = e^{-(\rho/R_{vir})^2} e^(- (\Delta v/200)^2, \tag{1}$$

where ρ is the physical impact parameter between the sightline and a galaxy, R_{vir} is the galaxy's virial radius, and $\Delta v = v_{galaxy} - v_{absorber}$, the difference in velocity between the galaxy and the absorption line. In order for a galaxy to be deemed "associated" with an absorption line, we require $\mathcal{L}$ for any potentially associated galaxy be a factor of 5 larger than $\mathcal{L}$ for all other galaxies, and $\mathcal{L} \geq 0.001$. This hard limit translates to an absorber located at $\sim 2R_{vir}$ and ~ 350 km/s in physical and velocity separation, respectively. This edge agrees nicely with observational results of **EXAMPLES AND REFERENCES**.

After IDing all spectral features above the 2σ???? level and matching nearby features with our galaxy dataset, we will have produced the largest and most complete CGM survey to date. This dataset is a natural extension of our own research goals, but will also serve the science goals of many other, future IGM-related projects.

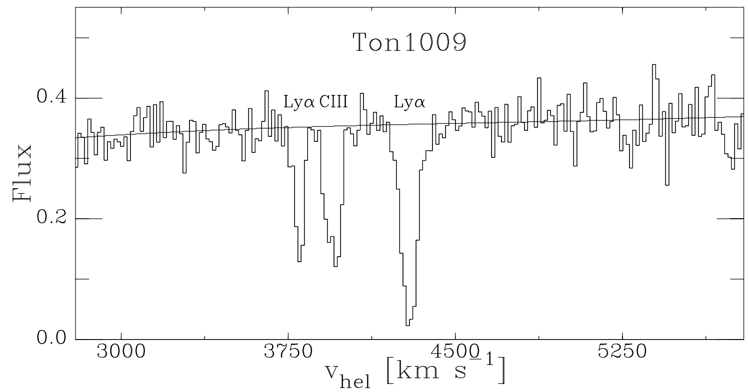
Analysis Plan

Line Identifications and Measurements

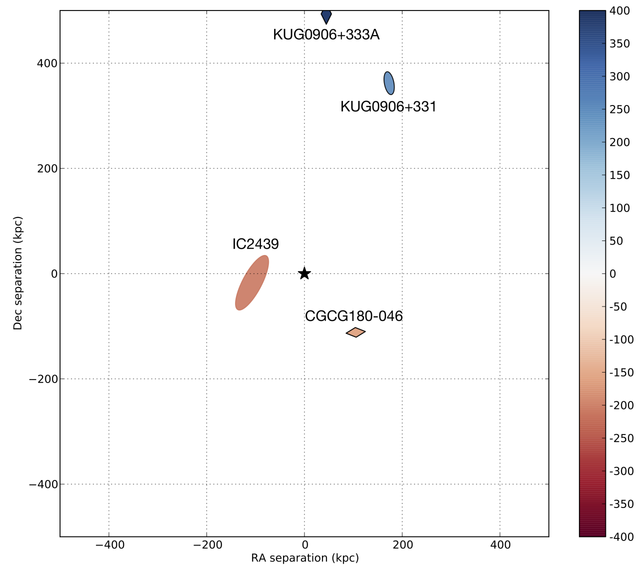
The first step will be aligning and combining multiple exposures to produce a single, clean spectrum. We will then apply the line identification code and method described above. This will result in a dataset of all absorption lines with ID's, velocities, equivalent widths, linewidths, and column density estimates.

Galaxy Matching

Second, we will correlate our galaxy dataset with the newly produced absorber dataset. This will produce matched absorber-galaxy systems, complete with association likelihood estimates.



(a)



(b)

Figure 1: a) An example Lyα line found in a sightline towards target TON1009 at 4295 km/s. b) A map of *all* galaxies within a 500 kpc impact parameter target TON1009 sightline and with velocity (cz) within 400 km/s of absorption detected at 4295 km/s (central black star). The galaxy IC2439 ($v = 4494$ km/s, inclination $= 71°$) can be unambiguously paired with the Lyα absorption feature at $v = 4295$ km/s because it is the largest and closest galaxy in both physical and velocity space to the absorption feature.

Publication

The final data set will be made available publicly in a machine readable format.

Management Plan

Line Identifications and Measurements

Spectra prep and line IDing will be led by Wakker, with additional input by French.

Galaxy Matching

Galaxy matching and final dataset preparation will be led by French.