## ■ Scientific Justification

The majority of the baryons in the universe are found in the intergalactic and circumgalactic medium (IGM, CGM) (e.g. Lehner et al. 2007, Danforth & Shull 2008, Cen 2013 and references therein). Galaxies reside within large gaseous filaments, from which they draw new material for continuing star formation, and into which they eject enriched material. Observationally studying these complex interactions is challenging, yet paramount to properly understanding and modeling global galaxy evolution.

The properties of this 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 absorption equivalent width (e.g. Bowen et al. 2002), linewidth (e.g. Wakker & Savage 2009), column density (e.g. Rudie et al 2012), and metallicity (e.g. Kacprzak et al. 2014) decreasing with the distance to the nearest bright galaxy.

Nevertheless, many open questions remain concerning the details of how gas near galaxies (the CGM, generally considered to extend to $\sim 2R_{vir}$) interacts with and affects the evolution of the galaxies. Key questions we would like to answer include: a) how do the properties of CGM absorbers (e.g. equivalent width, location, velocity) compare to the properties of the galaxies they are associated with (e.g. size, inclination, morphology)? b) does the CGM follow the rotation direction and velocity of the associated galaxies, as predicted by simulations (e.g. Stewart et al. 2011)?

More generally, we would like to know whether galaxies shape their environments or vice-versa. Studying the properties of neutral H I via Ly$\alpha$ absorption is one of the most direct ways to probe the galaxy-IGM interface, which remains poorly understood. A large sample of absorbers with associated galaxies could provide clues to inflow and outflow behavior, halo gas dynamics and composition, and how gas and galaxies are oriented and interact in a broader, cosmological context. To answer these questions it will be necessary to probe gas both very near as well as physically far from galaxies. This is most easily done in the nearby universe, where the galaxy sample is highly complete to low luminosities. However, previous studies have suffered from several shortcomings. We propose a program using archival COS data to deal with the following 3 issues in particular:

1) **Galaxy Incompleteness**: As the completeness of known galaxies decreases sharply with redshift, many CGM studies suffer from limitations due to inhomogeneous and incomplete galaxy data. 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). This complicates the process of associating absorption with nearby galaxies, and may result in significant biases. Additionally, very few galaxies have published rotation curves, so comparing the velocity of the gas in the CGM to that within the disks of the associated galaxies is usually not possible. Distant galaxies also tend to have less detailed information (e.g. inclination, position angle, size), and are more prone to misclassification.

**Our solution - A new galaxy catalog for the redshift range $cz \leq 10,000$ km/s:**

In this range available galaxy data is complete to $\sim 0.2L^*$ at $10,000$ km/s (and better towards lower redshifts). We created this galaxy catalog by mining the NASA Extragalactic Database (NED), collecting redshifts, diameters, redshift-independent-distances, morphologies, inclinations, position angles, photometry, and more for each galaxy, and then normalizing these values beyond the basic work done by NED. This catalog is the most complete, comprehensive and up-to-date nearby galaxy catalog in existence (to be published in 2016), and is key to allowing us to draw direct comparisons between the properties of the absorbers in the CGM and the properties of galaxies on a larger scale then ever before.

2) **Reproducibility**: It is often unclear and sometimes arbitrary how absorption features are matched with a galaxy. Some authors (e.g. Kacprzak et al. 2011) pick the galaxy closest to the detected IGM absorber, while others (e.g. Wakker & Savage 2009) try to take galaxy size and other properties into account, and most tend to deal with ambiguities in an ad-hoc manner. This makes it difficult to compare results between different studies, and is preventing us from building a more global picture of the galaxy-IGM connection.

**Our solution - Likelihood Method:** In French & Wakker (2016, in prep) we are introducing a reproducible method to streamline these types of decisions. We define the likelihood as:

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

where $\rho$ 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 our pilot study (French & Wakker 2016, in prep, see below) we require $\mathcal{L}$ to be a factor of 5 larger than $\mathcal{L}$ for all other galaxies, and $\mathcal{L} \geq 0.001$, for a galaxy to be marked "associated" with an absorption line. This hard limit translates to an absorber located at $\sim 2R_{vir}$ and $\sim 350$ km/s in physical and velocity separation, respectively. This edge agrees with previous practice, but is rigorously defined (see Figure 1 for an example).

Furthermore, we can use this method as a tunable parameter, and break down our results into bins of likelihood. This allows us to study how, e.g., the equivalent width and $\Delta v$ of an absorber population evolve with stricter or more relaxed galaxy-association criteria.

3) **Sample size**: Probing the CGM in absorption requires a serendipitously located background source, usually only one of which can be found for a particular galaxy. It is thus necessary to study the statistics of a large dataset of single galaxy-absorber matches. However, current CGM surveys max out at fewer than 100 galaxy-absorber pairs (e.g. 93 in Wakker & Savage 2009, 44 in Tumlinson et al. 2013, 89 in Steidel et al 2010, 71 in Stocke et al 2013).

**Our solution - 1) Archival data:** At the time of writing 550 QSO targets have been observed with COS. 300 spectra have S/N≥10; of these 120 have $0.03 \leq z \leq 0.2$, such that intrinsic Ly$\alpha$ is in the G130M wavelength range, 90 have $0.2 \leq z \leq 0.45$, putting Ly$\alpha$ in G160M, and 90 have $z > 0.45$. Nearly all of these pass within 500 physical kpc of at least one galaxy in the $cz \leq 10,000$ km/s redshift range. In our pilot study of 35 sightlines, we measured 176 Ly$\alpha$ absorption lines, 42 of which we paired with a nearby galaxy using our likelihood method. Hence, we predict 300 total spectra should produce over 1500 absorption lines and 360 absorber-galaxy pairings.

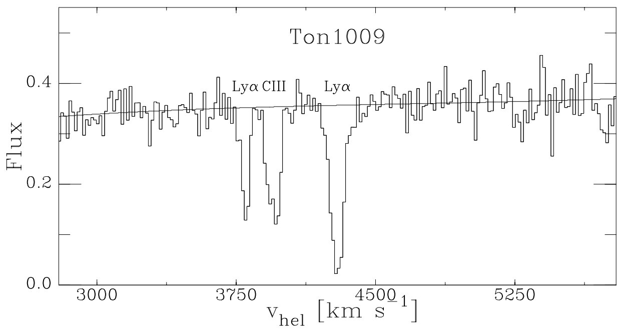
**Our solution - 2) Line Identification**: The identification and measurement of spectral features is a major undertaking, and limits the viability of large-scale studies using archival HST data. To combat this we have developed a pipeline that helps automate the IDing of targets, which will allow us 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 probable galaxy associations.

4) **Pilot Study:** In French & Wakker (2016, in prep) we will present results from an initial study of 35 COS sightlines. This sub-sample yields 176 Ly$\alpha$ systems at $cz \leq 10,000$ km/s, 32% of which can be unambiguously paired with a single galaxy using our likelihood selection algorithm. In agreement with previous studies (e.g. Rudie et al 2012, Wakker & Savage 2009 and references therein), we find a steep increase in Ly$\alpha$ equivalent width ($W$) with decreasing $\rho$, *but only when normalizing for galaxy size.* Absorbers are also preferentially detected around highly inclined galaxies, with 73% associated with a galaxy of inclination $> 50°$.

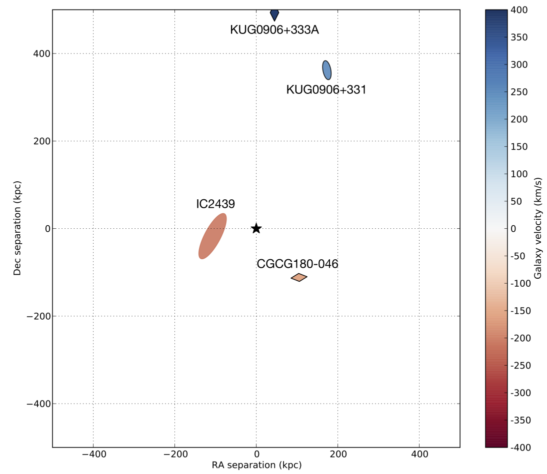
We also discovered a dichotomy in the equivalent width of absorption detected red-ward vs blue-ward of the systemic velocity of the associated galaxies (see Figure 3), with $\overline{W}$(redshifted) = 153±14 mÅ, and $\overline{W}$(blueshifted) = 321±13 mÅ . This difference is significant at the 99% level from results of both the Anderson-Darling and Kolmogorov-Smirnov statistical distribution tests. Following the suggestions of recent simulations (e.g. van de Voort et al. 2012), this could be evidence of cold, filamentary inflows, which are expected to be dense, and hot, lower density outflows. However, the robust statistics and increased parameter space of our proposed survey are necessary to fully explain these intriguing new results.

**Summary:** Galaxies have a complex relationship with their environment, and are known to both eject and accrete material from the surrounding IGM. Although many studies have probed this CGM gas, none yet have overcome the issues of incompleteness, small sample size, and reproducibility. We propose a large scale survey to overcome these common shortcomings, and provide robust statistics on how the properties of Ly$\alpha$ absorption depend on the properties of nearby galaxies. We will select 300 of the 550 available COS sightlines to ID and cross-correlate with our nearby galaxy catalog, and produce the largest yet galaxy-absorber catalog. This will also provide a legacy dataset of ID'd and measured absorption lines for a majority of COS sightlines, which will continue to be useful to the community long after the end of this project.

**References:** Bowen, D.V., et al. 2002, ApJ, 580, 169; Cen, R. 2013, ApJ, 770, 139;




(a)




(b)

Figure 1: a) An example Ly$\alpha$ 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 Ly$\alpha$ absorption feature at $v = 4295$ km/s following our selection criteria ($\mathcal{L}_{IC2439} = 0.45$, many times larger than all other nearby galaxies). Note that the galaxy sizes have been scaled up by a factor of 10 in this figure.

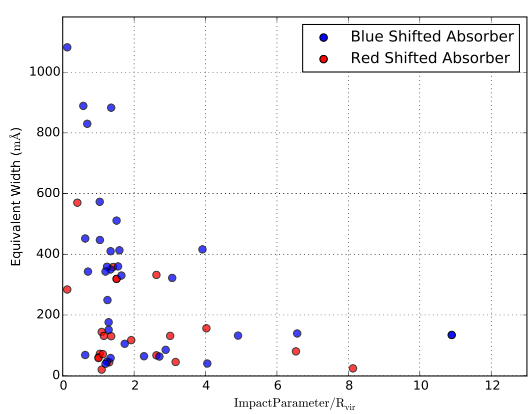



Figure 2: French & Wakker (2016, in prep) plot showing equivalent width of Ly$\alpha$ absorbers as a function of $\rho/R_{vir}$, the impact parameter to the associated galaxy normalized by the galaxy virial radius.

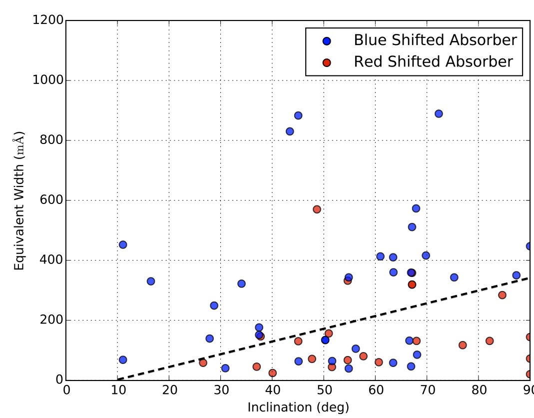



Figure 3: French & Wakker (2016, in prep) plot showing equivalent width of Ly$\alpha$ absorbers as a function of the inclination of the associated galaxy. The black dashed line highlights the dichotomy between red and blue shifted absorbers.

Danforth, C.W., & Shull, J.M. 2008, ApJ, 679, 194D; Kacprzak, G.G., et al. 2011, MNRAS, 416, 3118; Kacprzak, G.G., et al. 2014, ApJ, 792L, 12K; Lehner, N., et al. 2007, ApJ, 658, 680L; Mathes, N.L., et al. 2014, ApJ, 792, 128; Rudie, G.C., et al. 2012, ApJ, 750, 67; Steidel, C.C., et al. 2010, ApJ, 717, 289; Stewart, K.R., et al. 2011, ApJ, 738, 39; Stocke, J. T., et al. 2013, ApJ, 763, 148; Tumlinson, J., et al. 2013, ApJ, 777, 59T; Wakker B.P. & Savage B.D. 2009, ApJs, 182, 378; Werk, J. K., et al. 2014, ApJ, 792, 8;

Savage & Sembach 1991 Carswell et al. 2002 Kim et al. 2007

## Analysis Plan

**1. Line Identifications and Measurements:**

We will begin by aligning and combining multiple exposures to produce a single, clean spectrum. We then apply the line identification code, the mechanics of which are relatively simple and robust. First, we fit gaussian a profile to all spectral features above $2\sigma$ significance to determine their centroid wavelengths. Second, this list is fed into our algorithm, which produces IDs for each feature. We inspect each ID and make adjustments as needed. In this way each spectrum is both fit by machine and checked by eye, which minimizes miss-fits and machine artifacts. This will result in a dataset of all absorption lines with ID's, velocities, equivalent widths, and linewidth estimates. The time needed to ID a sightline can be as little as half an hour for an easy, low-redshift target, to several days for one at $z > 1$; for the first 100 we have already completed, the median is $\sim 2$ hours.

Once ID'd, we carefully measure all spectral features in the $cz \leq 10,000$ km/s redshift window. After fitting a low order (1st or 2nd) polynomial to line-free regions around each absorption line, we calculate the column density and estimate a linewidth using the apparent optical depth method (see Savage & Sembach 1991), as well as by making a Voigt profile fit using the VPFIT package (see Carswell et al. 2002, Kim et al. 2007).

**2. Galaxy-Absorber Matching:**

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

**New Nearby Galaxy Catalog:** Our nearby galaxy catalog contains over 108,000 galaxies, and is mostly complete to $\sim 0.2L^*$ at $10,000$ km/s (progressively better towards lower redshifts). The data table contains the following entries for each galaxy: coordinates, redshift, heliocentric and Virgocentric flow corrected velocities, redshift-independent-distance, inclination, position angle, morphology, extinction, photometry, group association, and alternative galaxy names. NED makes an effort to homogenize these data, but we go further preferentially using 2MASS $K_s$-band values for inclination, position angle, and diameter measurements. When these are not available, we scale measurements in other bands (e.g. SDSS $g$ and $r$) to the expected $K_s$-band values based on correlations between photometry bands. 2MASS values were chosen as it is an all-sky survey, and measurements are available for the majority of galaxies. Additionally, we calculate best-estimate B-band magnitudes and $L^*$ values from the multitude of disparate photometry values included in NED.

**3. Interpretation and Publication:**

With both the absorption line and nearby galaxy datasets in hand, we will investigate how Ly$\alpha$ equivalent width varies as a function of associated galaxy inclination, impact parameter, morphology, virial radius, azimuth angle, and system velocity with respect to absorption velocity. Particular attention will be paid to how these effects vary with proximity to the galaxies (i.e. within bins of velocity, impact parameter, and likelihood), so that we can understand the CGM-galaxy connection and how it evolves across a range of scales.

To address our original question, "b) does the CGM follow the rotation direction and velocity of the associated galaxies?", we have secured Priority 1 time on the Southern African Large Telescope (SALT) to measure the rotation curves of 15+ galaxies with sightlines lying within $2R_{vir}$. With this data in hand we will compare the rotation and orientation of the galaxies with that of Ly$\alpha$ absorption detected in their halos, investigating if the Ly$\alpha$ absorbers tend to co-rotate with the direction of the gaseous disk, and if so, out to what distance.

The results of these studies will be published in a series of papers. First, the list of COS IDs and line measurements, as well as our new nearby galaxy catalog, will be published and made available electronically once this step is completed. Next, we will complete and publish our study of absorber properties as a function of galaxy properties (including new SALT rotation curves). Finally, with the robust statistics provided by our completed datasets, we plan to publish our results on the global CGM density profile and velocity fields in the nearby universe.

## 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.