Scientific Justification

The majority of the baryons in the universe are found in the intergalactic and circumgalactic medium (IGM, CGM) (e.g. Danforth & Shull 2008, Lehner et al. 2007, Cen 2013 and references therein). Galaxies reside within large filaments of IGM gas, and draw new material from it for continuing star formation, as well as eject enriched material back into it. Studying this complex process observationally 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 (CITE), linewidth (e.g. Wakker & Savage 2009), column density (e.g. Rudie et al 2012), and metallicity (e.g. Kacprzak et al YEAR?) decreasing with the distance to the nearest bright galaxy.

Nevertheless, many open questions remain concerning the details of how gas near to galaxies (the CGM, generally considered to extend to $\sim 2R_{vir}$) interacts with and affects the evolution of the galaxies. Some key questions we would like to answer are: 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)?

Probing gas both very near to as well as physically far from a galaxy is essential to developing a full understanding of the galaxy-IGM interface. 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 in trying to address these questions. We propose a program to deal with the following 3 issues in particular:

1) **Incompleteness**: As the completeness of known galaxies decreases sharply with redshift, many CGM studies suffer form limitations due to inhomogeneity and incompleteness in the galaxy data. For example, Mathes et al. (2014) and Werk et al. (2014) are only complete to $\sim L^*$ (at $0.12 < z_s < 0.67$ and $z \sim 0.2$, respectively). This complicates the process of associating absorption with a nearby galaxy, and can 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. Aside from general incompleteness, distant galaxies also tend to have less detailed information (e.g. inclination, position angle, size), and are more prone to misclassification. To correct for this, we are limiting our study to the redshift range $400 \leq cz \leq 10,000$ km/s, where available galaxy data is complete to $\sim 0.2L^*$ at 10,000 km/s, and progressively better towards lower redshifts. We have created a nearby galaxy catalog to aid in our analysis 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

1

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.

~~We have also 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 Lyα absorption detected in their halos.~~

[margin note: Move to end of "analysis plan"]

2) **Reproducibility**: ~~In order to correlate CGM absorber properties with a galaxy, a match must be made between 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~~ *[pick the nearest]* galaxy ~~to~~ *[closest]* the detected IGM absorber, while others (e.g. CITATION) try to take galaxy size and other properties into account ~~as well.~~ However, different authors continue to use different, subjective criteria, and tend to deal with ambiguities in an ad-hoc manner.

[margin note: It is often unclear and sometimes arbitrary how absorption feature are matched with a galaxy.]

[margin note: Our solution:] **Likelihood Method:** 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:~~

[margin note: You could save a tiny bit of space by not using TeX's $$eq$$ system, but by doing "as: \smallskip \par .... (eq) \smallskip \par]

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

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~~ *[to]* 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.**~~ *[with previous practice, but is rigorously defined.]*

3) **Sample size**: Probing the CGM in absorption requires a serendipitously located background source, ~~which means that a particular galaxy halo generally only can be probed once.~~ *[Usually only one 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, ~~the largest~~ 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).

[margin note: Our solution:] **Archival data:** At the time of writing 550 QSO targets have been observed with COS. 300 spectra have S/N$\geq$10; of these 120 have $0.03 \leq z \leq 0.2$, such that intrinsic Lyα is in the G130M wavelength range, 90 have $0.2 \leq z \leq 0.45$, putting Lyα 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 (French et al. 2016, in prep), we measured 176 Lyα absorption lines, 42 of which we paired with nearby galaxy using our likelihood method. Hence, we predict 300 total spectra should produce over 1500 absorption lines and 360 absorber-galaxy pairings.

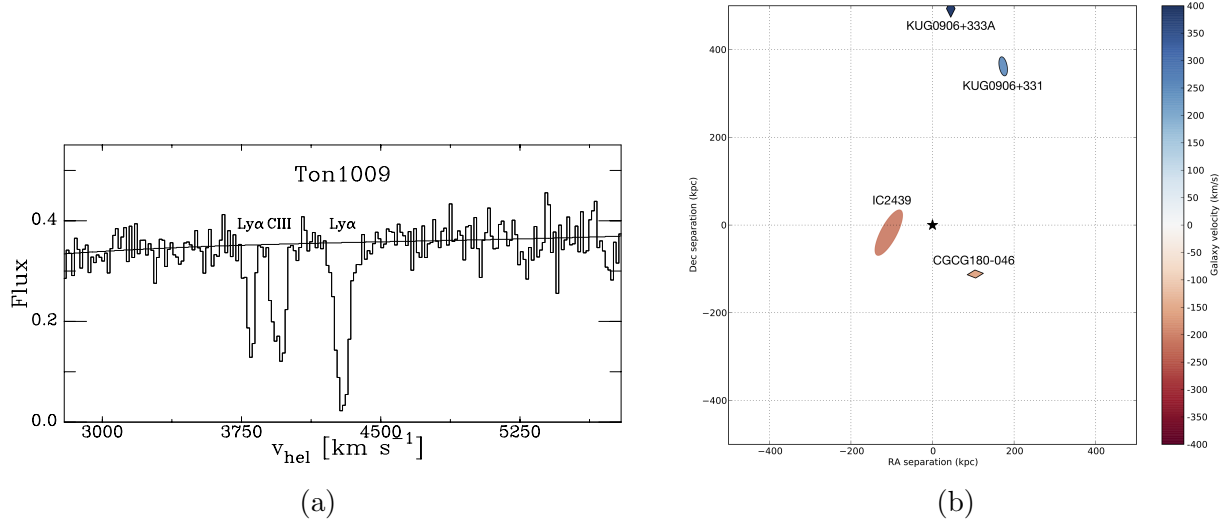Figure 1: a) An example Lyα line found in a sightline towards target TON1009 at 4295 km/s. b) A map of *all* galaxies within a 500 kpc impact parameter target TON1009 sightline and with velocity (cz) within 400 km/s of absorption detected at 4295 km/s (central black star). The galaxy IC2439 ($v = 4494$ km/s, inclination $= 71°$) can be unambiguously paired with the Lyα absorption feature at $v = 4295$ km/s following our selection criteria ($\mathcal{L}_{IC2439} = 0.45$, many times larger than all other nearby galaxies).

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. 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 Lyα 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.

Analysis Plan

1. Line Identifications and Measurements:

~~The analysis begins with~~ [We will begin by] aligning and combining multiple exposures to produce a single, clean spectrum. We ~~will~~ then apply the line identification code, the mechanics of which are relatively simple and robust. First, we ~~manually~~ fit [a] gaussian profiles to all spectral features [to determine their centroid wavelengths] above 2σ significance. Second, this list ~~of fits~~ is fed into our algorithm, which produces ~~best guess~~ [ID] IDs for each feature. ~~Finally,~~ we inspect each ~~set fits~~ and make ~~small~~ 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] linewidths, [estimates] ~~and column density estimates~~. The time needed to ID a sightline can be as little as half an hour for a~~n easy,~~ low-redshift target, to several days for one at $z > 1$; for the first 100 we have completed ~~already,~~ [already] the median is ~ 2 hours.

Once ID'd, we [carefully] identify and 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). ~~We then derive a linewidth by calculating the second moment of the apparent optical depth profile.~~

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 (~~and~~ progressively better towards lower redshifts). The data table contains the following entries for each galaxy: coordinates, redshift, heliocentric velocity, **CORRECTED VELOCITY**, 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 by ~~normalizing all measurements of inclination, position angle, and diameter to 2MASS K-band values.~~ 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 Lyα 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 and impact parameter), so that we can under-

preferentially using 2MASS Ks band values, and when those are not known, scaling the values in other bands (e.g. SDSS r) to the expected Ks band values based on correlations between r and Ks values.

stand the CGM-galaxy connection and how it evolves across a range of scales. For sightlines that lie nearby the SALT-observed galaxy disk velocity measurements, we will investigate if Lyα absorbers tend to co-rotate with the direction of the gaseous disk, and if so, out to what distance.   Insert the "SALT time" paragraph here

The results of these studies will be published in a series ~~of 2~~ papers, ~~published in order as ground observations become available and IDing COS spectra progresses. The first will be the study of SALT rotation curves vs Lyα halo absorption, the second will be the full analysis of absorber properties as functions of galaxy properties. This second paper will also include an electronic data set containing all COS IDs, absorption line measurements, and galaxy associations.~~

Rewrite, with HST emphasis. Don't give a number.
Include, in order: (1) list of COS IDs, (2) galaxy catalogue,
(3) absorber properties as function of galaxy properties
(4) SALT rotation curves vs Lya absorption
(5) Study of CGM density profiles and velocity fields.

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.
Need more words. Include TaeSun for line fitting.