Collaborative Research: Multiscale Physics and Feedback in the Real and Simulated Circumgalactic Medium Over Cosmic Time

1. Motivation and Overview: The circumgalactic medium (CGM) is now one of the vital frontiers of galaxy formation research. The CGM—between the disk and virial radius (~300 kpc)—is where accretion feeding star formation meets enriched feedback, where satellites are stripped and disrupted, and where ejecta may eventually be recycled. Recent observations indicate that the CGM at $z \sim 0.2$ hosts a share of galactic baryons comparable to the stellar masses($M_{CGM} \gtrsim M_\star$). It is also a massive reservoir of galactic metals, with galaxies having ejected at least as much metal mass as they have retained, which may be recycled into the ISM. These findings indicate that the mass flowing through the CGM have a major role in establishing galaxy properties, so it is likely that our picture of galaxy evolution will be incomplete until we understand it. These low-z observations bring great progress, but pose several puzzles:

The Mixing Problem: Dense circumgalactic gas (log [N_{HI}/cm^{-2}] ~ 16–18.5) appears bimodal in metallicity, with a metal-rich portion (~20–100% solar) potentially tracing enriched outflows, and a distinct metal-poor component (~5% solar) resembling "cold accretion" from the IGM (Lehner+2013). The apparent degree of mixing is surprisingly small.

The Density Problem: Cool, low-ionization CGM gas traced by lower H I column densities ($14 \lesssim \log N_{HI} \lesssim 16$) appears to have physical density several orders of magnitude lower than the densities implied by hydrostatic equilibrium with a virialized hot corona (Werk+2014).

The Timescales Problem: The CGM hosts a massive reservoir of highly ionized gas traced by the O VI ion, which is optimally produced at the peak of the cooling curve (Tumlinson+2011; Peeples+2014) where cooling times are short (~10^8 yr) and we would not expect much total mass. Yet if this material is a transient phase, perhaps cooling out of a more massive reservoir, it would imply a troubling surplus of halo baryons (Werk+2014).

The Fate Problem: Passive galaxies at $z \sim 0.2$ contain as much cool, bound gas in their CGM as their star-forming counterparts—gas masses comparable to their stellar mass (Thom+2012, Werk+2014). This presumably unstable gas presents a puzzle as to why they possess potentially star forming fuel but do not use it, and hints at small-scale physics that we do not understand.

We have the opportunity to learn far more about how galaxies acquire, process, and expel their gas if we can achieve a multiscale view of CGM gas physics that connects to data across cosmic time. The critical barriers to progress are *evolution* and *resolution*: further insight is limited by our inability to follow the time evolution of the CGM and its physical processes at small spatial scales, and to relate these to observables, from which we can estimate the critical mass flow rates. Our program combines observations and theory with the following key elements:

1) A new analysis of $z \sim 2$–3 CGM absorbers in archival Keck+HST data, to expand the "CGM timeline" over 10 billion years with data complementing the abundant $z < 1$ observations.
2) A multi-scale simulation effort combining cosmological boxes, zoomed galaxies, and idealized clouds to model CGM gas physics spanning over seven decades of spatial resolution.
3) Comparisons of data across cosmic time with multi-scale simulations using a new toolkit for mocking datasets that enables direct translation between underlying physics and observables.

We have designed our closely coupled program of data and theory so that they can continuously guide each other in a mutually reinforcing fashion. Our experience as separate investigators suggests that this integrated approach promises better intuition, deeper insights, faster progress, and possibly more breakthroughs, than working in a conventional decoupled way.

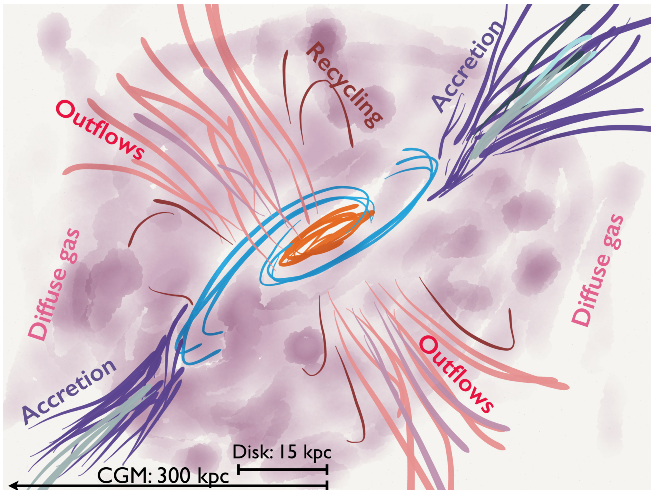

Figure 1: A cartoon of the CGM showing the diffuse gas of the varying density (shaded purple), accretion through filaments into the disk (darker purple), and outflows out of the disk (in orange). While simulations support this basic picture, and observations suggest it is largely correct, the actual density and temperature distributions, kinematics, and geometry of the gas, and thus the mass content and fate of these critical gas flows, are poorly understood.

2. Progress and Problems in the CGM, and the Need for Evolution and Resolution

In recent decades astronomers have made dramatic progress in elucidating the dark-matter underpinnings of cosmic structure formation and in characterizing the properties of galaxies as viewed in their starlight and ISM gas. It is fair to say that the basic outlines of galaxy formation are now drawn, and we must grasp the messy details of gas physics to explain how galaxies acquire their stellar masses, evolve in shape, and quench their star formation. Over the last decade, the gas surrounding galaxies—the CGM—has come under increasing scrutiny for its role in fueling galaxy accretion, mediating galaxy feedback, and promoting gas recycling. Because CGM gas is diffuse (~10^{-5}–10^{-2} cm^{-3}) it is difficult to see in emission and so is usually detected via absorption lines in the spectra of background QSOs—the classic "quasar absorption lines" method. For low-density gas the *observables* are absorption lines from common ions with electronic and fine-structure transitions in the rest-frame UV. Because of their wavelength and intrinsic strength, the most commonly used lines are from H I (Lyα), Si II / III / IV, C II / III / IV, O VI (all UV lines at low redshift) and Mg II (usually redshifted into the visible). These species possess a wide range of ionization potentials (from < 1 to > 5 Rydberg) and so they trace gas at a wide range of temperatures (hotter gas having higher ionization) and/or densities (more rarified gas having higher ionization). By observing multiple ionization stages from these elements, we can assess the density, temperature, ionization mechanism, and kinematics of the absorbing material. Since CGM gas is usually highly ionized (neutral fraction of $\lesssim$ 1%), these "multiphase" ions are used to constrain a model from which the "ionization correction" is derived. This correction is essential to deriving mass, mass density, and metallicity estimates of CGM gas.

These techniques are now routinely applied to CGM absorption detected in low-z QSO / galaxy pairs, where sample sizes reach 50–100 for individual studies and ~300–500 total with z < 1 (depending on selection criteria) using *Hubble* UV spectroscopy. The rapid growth in this subject was triggered by Hubble's Cosmic Origins Spectrograph (COS) and the vast number of QSOs and galaxies surveyed by SDSS that it brought within reach. Since 2009 this leap in capability has populated a new discovery space with a wealth of findings in the $z \lesssim 0.3$ universe (Figure 2). HST/COS observations have revealed that: (1) the CGM at $z \sim 0.2$ hosts a large share of galactic baryons, with the CGM mass approximately equal to the stellar mass ($M_{CGM} \sim M_\star$) at around ~L^* (Stocke+2013; Werk+2014), (2) the CGM is a massive reservoir of galactic metals, with galaxies having ejected at least as much metal mass as they have retained (Tumlinson +2011; Peeples+2014; Bordoloi+2014), which may ultimately be recycled back to the ISM (Keeney+2013; Ford+2014), (3) dense CGM gas appears bimodal in metallicity, with a portion tracing enriched outflows (~20–100% solar) and a distinct metal-poor component (~5% solar) which resembles "cold accretion" entering galaxies from the IGM (Lehner+2013; Kereš+2005), (4) galaxies can eject a large fraction of their gas in multi-phase superwinds from 10^4–10^6 K (Tripp+2011), and (5) the CGM of passive galaxies at $z \sim 0.2$ contains as much cold, bound gas as their star-forming counterparts, raising the question of why they do not use it and how the CGM affects (or is affected by) the quenching of star formation (Thom+2012; Tumlinson+2013).

The findings described above highlight the vital role of the CGM in galaxy formation, but the COS surveys on which they are based concentrate strongly at $z < 1$, by which time the red sequence is well-established and cosmic star formation is well off its peak. These low-z studies are complemented across the cosmic SFR peak by studies of CGM around the star-forming Lyman-break galaxies (LBGs), which drive outflows seen in common FUV lines redshifted into the optical at $z \sim 2$–3 (Steidel+2010; Rudie+2013; Matejek & Simcoe 2013; Turner+2014). A recent study by our group (Lehner+2014) finds that, just as at low z, highly ionized gas traced by O VI at $z \sim 2$ is common in strong CGM absorbers, implying that the host galaxies have metal-rich halos. Despite this good progress, high-z CGM metallicity distributions, density fields, and mass budgets are note yet characterized at the level as they have been for $z < 1$ because there is not yet a full database of comprehensive line measurements on which to base physical models.

These interesting results complete the first major intellectual cycle of CGM research. Yet in each case, the finding raises immediate puzzles, or conflicts with another finding, so that we cannot yet see a complete and accurate picture of the CGM and its role in galaxy evolution. We will now review these findings and the problems they raise, and from there consider how we will address them with new data (§3) and in state-of-the-art simulations (§4).

The Mixing Problem: COS observations of strong CGM absorbers (called Lyman limit systems or LLSs) at $z < 1$ show that relatively dense CGM gas has a bimodal metallicity distribution with gas at ~20–100% solar observed about as frequently as the distinct metal-poor component at ~5% solar (Lehner+2013; Fig. 2). These two peaks are interpreted as metal-enriched feedback from the galaxy and metal-poor accretion from the IGM, respectively. Theorists have struggled to recover these peaks in CGM simulations that tend to wash out widely varying metallicities into an indistinct smear, hinting that significant metal transport and mixing occur below their usual resolution limit. Conversely, in observations it can be difficult to know if multiple clouds with different metallicities occur along the same line of sight within a halo. The apparent inefficiency of metal mixing poses problems for having all the cooler gas form by condensing out of hotter gas, which is one of the conventional pictures of CGM evolution (Maller & Bullock 2004). Characterizing the true distribution of CGM metallicity, and in turn understanding galactic feedback, requires us to constrain CGM metal transport at higher spatial resolution.

The Density Problem: Converting observed ionic column densities to physical conditions requires an ionization model and in turn assumptions about the mechanism(s) causing the ionization (such as photoionization or shocks). The simplest assumption, almost always made but rarely verifiable, is that the absorbing gas is in thermal and ionization equilibrium at all times. But this assumption obviously conflicts with the clearly dynamic and evolving state of simulated CGM gas. Still more problematic is the tendency of absorbers to appear "multiphase"—a euphemism for the commonly observed phenomenon that widely separated ionization states (e.g. Mg II and O VI) appear kinematically aligned within the limits of instrumental resolution (e.g., Tripp+2011; Werk+2013). Either this gas is either far from equilibrium or it possesses ionization structure below the level of resolution (e.g. ionized skins surrounding cooler interiors). If so, the

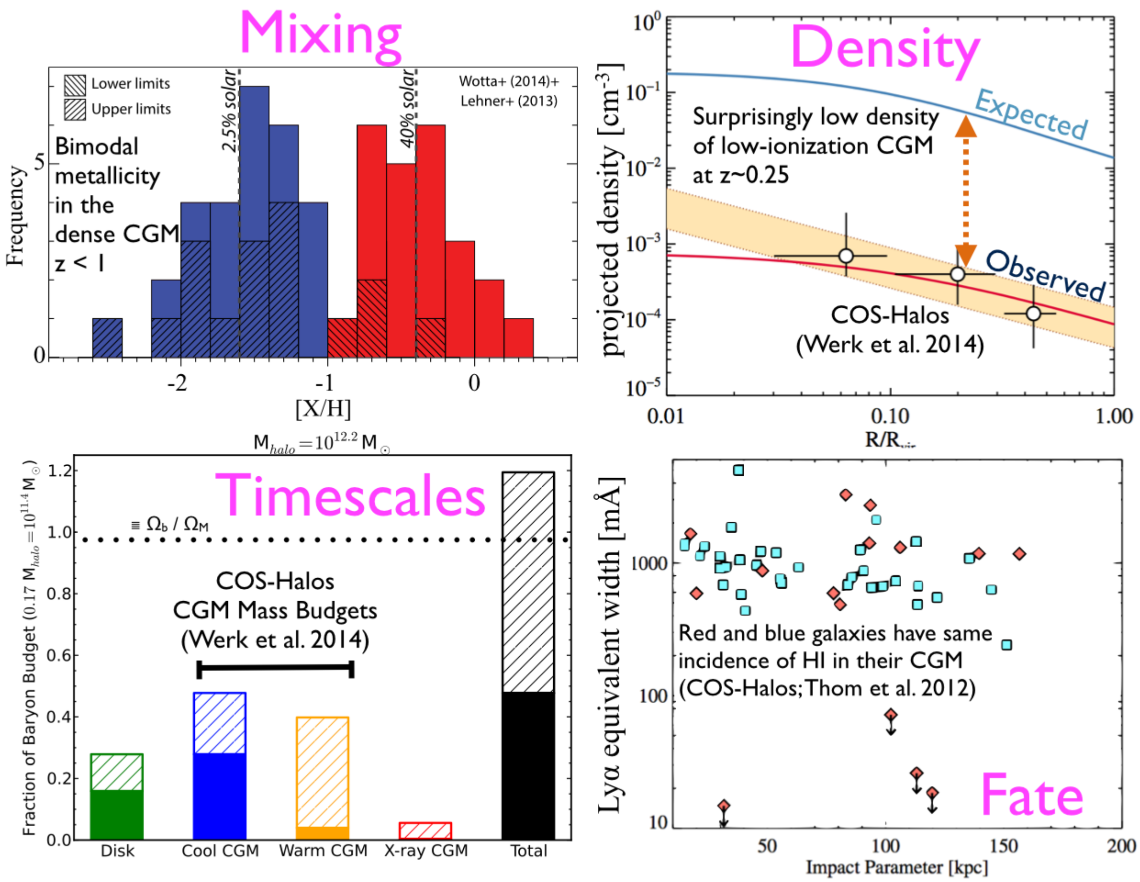

Figure 2: Highlights of low-redshift CGM studies. Upper left: The "mixing problem" is posed by the bimodal metallicity comprising metal-enriched feedback and metal-poor accretion (Lehner+2013). Upper right: The "density problem" is posed by the COS-Halos finding that cool CGM gas follows the density profile (not the pressure profile) of hot coronal gas. Lower left: The "timescale problem" is posed by the CGM baryon budgets compiled by Werk+2014, including cool and warm CGM gas. Lower right: The "fate problem" is posed by the widespread presence of cool CGM HI in the halos of passive galaxies (Thom+2012; Tumlinson+2013).

equilibrium, plane-parallel, "single-phase" models are in doubt. These issues dominate the systematic errors in baryonic mass estimates for CGM gas (see Werk+2014 for a discussion). The Werk analysis also reached the surprising conclusion that cool CGM gas follows a density profile that is closer to that of the expected hot corona, which suggests it is far from thermal equilibrium (Figure 2). Our ability to diagnose densities for dynamic, highly structured, non-equilibrium CGM gas is a critical hindrance to assessing the mass flow rates we want to know.

The Timescale(s) Problem: The vital question of how galaxies acquire their gas comes down to the rate at which CGM gas can cool and fall in. The relevant timescales are set by the balance between gravity acting on a free-fall time, $\tau \propto 1/\sqrt{(G\rho)}$ and thermal pressure, $P \propto kT$. A fluid element must cool to become dense and fall in on timescales (0.1–1 Gyr) short enough to sustain star formation over long periods. But the simulated CGM exhibits a huge range of density ($n_H \sim 10^{-5}$–10^{-2} cm^{-3}) and temperature ($T \sim 10^{3-7}$ K), and as stated above these quantities are only crudely extracted from absorption line data obtained at finite resolution and signal-to-noise ratio. Most densities drawn from observations are model-dependent. Thus the density problem leads to a timescale problem—without stronger constraints on CGM densities, we cannot estimate timescales, and therefore the *mass flow rates* that influence galaxy fueling and feedback.

Another form of the timescale problem arises when we consider the COS-Halos finding that highly ionized gas traced by O VI is nearly ubiquitous in the halos of low-z star-forming galaxies

(Tumlinson+2011). Any ion detected along nearly every sightline (like O VI) implies a large fraction of the CGM mass and/or volume is contained in the gas phase that it traces. The O VI abundances indicate that CGM oxygen is near its highest ionization fraction, but the temperature where this occurs ($10^{5.5}$ K) is near the peak of the cooling curve where the cooling time is short. If the observed gas is in a short-lived state (rapidly heating or cooling), its widespread presence in the CGM indicates a source of replenishment on similar timescales (winds from the disk, inflow from the IGM, or cooling out of a massive hot medium). These processes and their timescales are essential to understanding the mass flows between the IGM, CGM, and ISM.

The Fate Problem: Low-z observations raise a serious puzzle about CGM physics that might be resolved if solutions to the preceding problems were better understood. The COS-Halos survey (Tumlinson+2013) studied a modest sample of passive (red and dead) galaxies and found that they possess cool H I in the CGM gas about as frequently as do their star-forming counterparts at the same mass (Thom+2012; Fig. 2). This is a major puzzle: at the temperatures and densities we infer this gas "should" cool and fall into the galaxies, but it clearly does not, as passive galaxies have cold ISM gas masses of $M_{HI} \lesssim 10^{6\text{-}7}$ $M_{\odot}$. It might be that passive galaxies are able to "store" cold CGM gas within their (presumed) hot halos through some physical mechanism that we do not understand. This is a recast of the timescale problem—how can cold gas exist but evolve and move only over very long timescales, so that it avoids accretion only in the halos of passive galaxies? It is also possible that, since passive galaxies are not using their diffuse cool CGM as star-forming fuel, star-forming galaxies are not either, and we are misled into thinking that this diffuse CGM is a reservoir of star-forming gas. Thus, low-z CGM data has posed a problem that our poor understanding of small-scale physical processes prevents us from solving.

All these problems emerge when we try to work out the detailed interpretations of low-z CGM discoveries and relate these carefully to galaxy properties. The problem stems from the finite kinematic resolution in data and finite spatial resolution in simulations, combined with the difficulty of following CGM evolution over cosmic time. The stakes are high: understanding the CGM addresses some of the thorniest outstanding questions in galaxy evolution: how do galaxies sustain their star formation over >10 Gyr when they have ~1 Gyr worth of fuel? How is the mass-metallicity relation produced? What happens to the gas when galaxies turn passive—is it consumed or ejected? Where are the 80% of metals produced to date that are not in the galaxies that produced them (Peeples+2014)? Answers to these questions depend on our ability to pin down the mass, metallicity, and fate of the CGM, and to relate these to galaxy properties.

The problems of CGM *evolution* and *resolution* are major hindrances to rigorous estimates of CGM mass flows and thus a more complete picture of galaxy formation in general. Solving the problems posed by the low-z data will benefit enormously from progress on two fronts: (1) we must expand the CGM measurements past $z > 1$ to establish a 10 Gyr baseline of CGM measurements, and (2) we must apply new, highly resolved multiscale simulations to data from these epochs to examine CGM physics at the relevant scales.

At $z \sim 2$–3, we can obtain a set of baseline CGM measurements that is comparable to the low-z data, but at a time when the mean density was about ~10–30 times higher. At these redshifts, the UV lines commonly used at low z become accessible to ground-based optical telescopes, which have generated a rich and varied dataset just waiting to be mined. As we will detail in the next section, the extension to high z will address many of the open problems by establishing a baseline for the evolution of the CGM.

For simulations, the challenge is to connect microphysics to cosmological context in physically rigorous ways. This means including the appropriate physical processes in the simulations, but also ensuring that all relevant physical scales (from ~ 50 Mpc to < 1 pc) are covered, *and* that comparisons between data and theory are done as closely to the observable quantities as possible. Section 4 details our plan for multi-physics, multi-scale simulations that address the first two issues, and our plan to leverage a toolkit for deriving "mock spectra" from simulations for which our group is separately funded. Together these simulations and mock datasets promise a big advance on the problem of CGM *resolution*.

3. Solving CGM Evolution: Extending Discoveries to $2 < z < 4$:

To properly understand CGM physics over cosmic time, we must have CGM samples that are directly comparable in size and physical diagnostics across cosmic time. The low-z databases (e.g. COS-Halos) comprise ~50 galaxies with coverage of all the relevant diagnostic ions. At high redshift, the reservoir of relevant data is vast but has been relatively un-exploited so far. Our group has begun to remedy this situation by mining archival data for CGM absorbers.

3.1 The KODIAQ Survey

The Notre Dame / St. Michaels members of our group created the KODIAQ survey to open the frontier for detailed CGM studies at high z. KODIAQ (the Keck Observatory Database of Ionized Absorbers toward Quasars, P.I. Lehner) draws on a sample of over 400 $z > 2$ quasars with Keck HIRES (Vogt+1994) spectra obtained from the Keck Observatory Archive (KOA, http://www2.keck.hawaii.edu/koa/public/koa.php). KODIAQ was previously funded through a 4-yr NASA Astrophysics Data Analysis Program grant (June 2010–2014) for limited goals: to collect and reduce the data and to study the highly ionized gas traced by the O VI λλ1031,1037Å doublet associated with the dense CGM of galaxies at $z > 2$. A major portion of the KODIAQ grant funded our data reduction and co-addition of the individual spectral exposures of the Keck Observatory Archive (KOA) QSO database. HIRES observations of QSOs are not readily useable without intensive processing and strong knowledge of the instrument, owing to the non-uniform calibrations of the data taken by so many PIs. We have fully processed the data for 250 QSOs and are deep into the processing of the full dataset. This will provide the largest analysis-ready sample of high-z QSO spectra with high spectral resolution (6–8 km s^{-1}) and high signal-to-noise (S/N > 20–50). The full dataset is set for public release in Summer 2015.

The original goal of the KODIAQ sample was to identify highly ionized gas via O VI absorption; these results are summarized in Lehner+2014, in which we report on a survey for strong H I absorbers (i.e., log $N_{\rm HI} \gtrsim 17.3$) that have clean O VI lines. Just as COS-Halos found at $z \sim 0.2$, Lehner et al. demonstrate that strong OVI absorption is nearly ubiquitous in these strong H I absorbers associated with the CGM of $2 \lesssim z \lesssim 4$ galaxies. The strong and very broad O VI absorbers (log $N \gtrsim 14.7$, Δv = 200–400 km s^{-1}) probe massive outflows from actively star-forming galaxies, independently revealing the extremely high stellar activity and their impact on large-scale environments of high redshift galaxies. These same absorbers contain as much as 3–14% of the cosmic baryon budget at $z \sim 2$–4, i.e., almost all the baryons at that epoch are found in ionized gas in the IGM and CGM, not in galaxies. KODIAQ finds that 19–34% of the metals are in the form of ionized gas outside galaxies at these epochs, and hence a substantial fraction of metals have already been ejected by galaxy-scale outflows in the CGM at $z \sim 2$–4. These results are resonant with the low-z findings of COS-Halos, in which a large fraction of metals ever

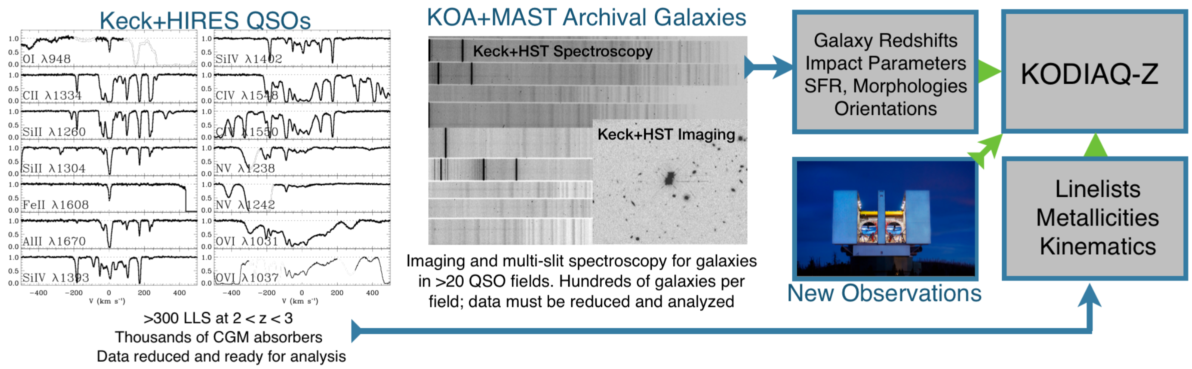

Figure 3: A summary of KODIAQ-Z (i.e. KODIAQ with metallicities). The expanded database will contain thousands of galaxies (many with HST imaging) with associated circumgalactic absorption and published linelists.

produced are found in the CGM (Peeples+2014). The KODIAQ results so far highlight, but do not solve, the density and timescale problems of small-scale physics raised by the low-z results.

3.2 The KODIAQ-Z Extension: Full Ionization Diagnostics at z = 2–4

The proposed KODIAQ-Z survey is the next generation of this project. To expand the analysis to take in the full set of diagnostic ions, and to extract from these the physical information needed to compare to low-z data and our multi-scale simulations, we must extend the KODIAQ survey in two ways. First, the new KODIAQ-Z survey will not restrict the analysis to systems showing O VI absorption, which will increase the number of strong CGM absorbers ($> 10^{16}$ cm^{-2}) by a factor of ten, to ~300 such systems as a conservative estimate.

We must complete KODIAQ-Z to address the physical problems raised at low z, particularly the mixing and density problems. KODIAQ-Z will be the first study to systematically examine the metallicity of high-z CGM gas, and therefore the first to compare it to the bimodal metallicities discovered at $z < 1$. At high redshift, the two low-z metallicity peaks could change their mean values, change in their relative populations, or merge into a unimodal distribution. The relative weighting of the two branches, and even the extent to which they mix, may have unique diagnostic power. At high redshift the star formation rate density of the universe is higher, reaching its peak at $1.5 \le z \le 3$ (e.g., Madau & Dickinson 2014). What does this imply for the metallicity of the CGM? Is the SFR increase fueled by a larger density of accreting matter? Is it matched by more powerful and numerous outflows? These questions in part depend on the star formation efficiencies and wind mass loading factors at these redshifts.

There have been very few metallicity determinations of high-redshift CGM gas to date; the total sample of ≲20 systems with reported metallicities at $z > 2$ from unbiased samples (e.g., Steidel 1990; Lehner+2014) is too small to reach firm conclusions about the metallicity distribution or its evolution with time. The small sample size at these redshifts is limited by the need for high signal-to-noise (S/N) and resolution spectra (for measuring accurate column densities of metals and H I) and the need for detailed photoionization models to determine the metallicity (Prochaska 1999; Lehner+2013; Crighton+2013). Thus, increasing the sample of high-z systems with bona-fide metallicity measurements is a primary goal of the KODIAQ-Z extension. This expanded sample, and the much larger database of low-ionization gas at $z > 2$ required to execute it, will be a major benefit to constraining the small scale physics we need to address the mixing, density, and timescale problems.

The KODIAQ-Z survey will also include archival data on galaxies along the sightlines to the maximum practical extent. We will include the full suite of data from other Keck instruments in

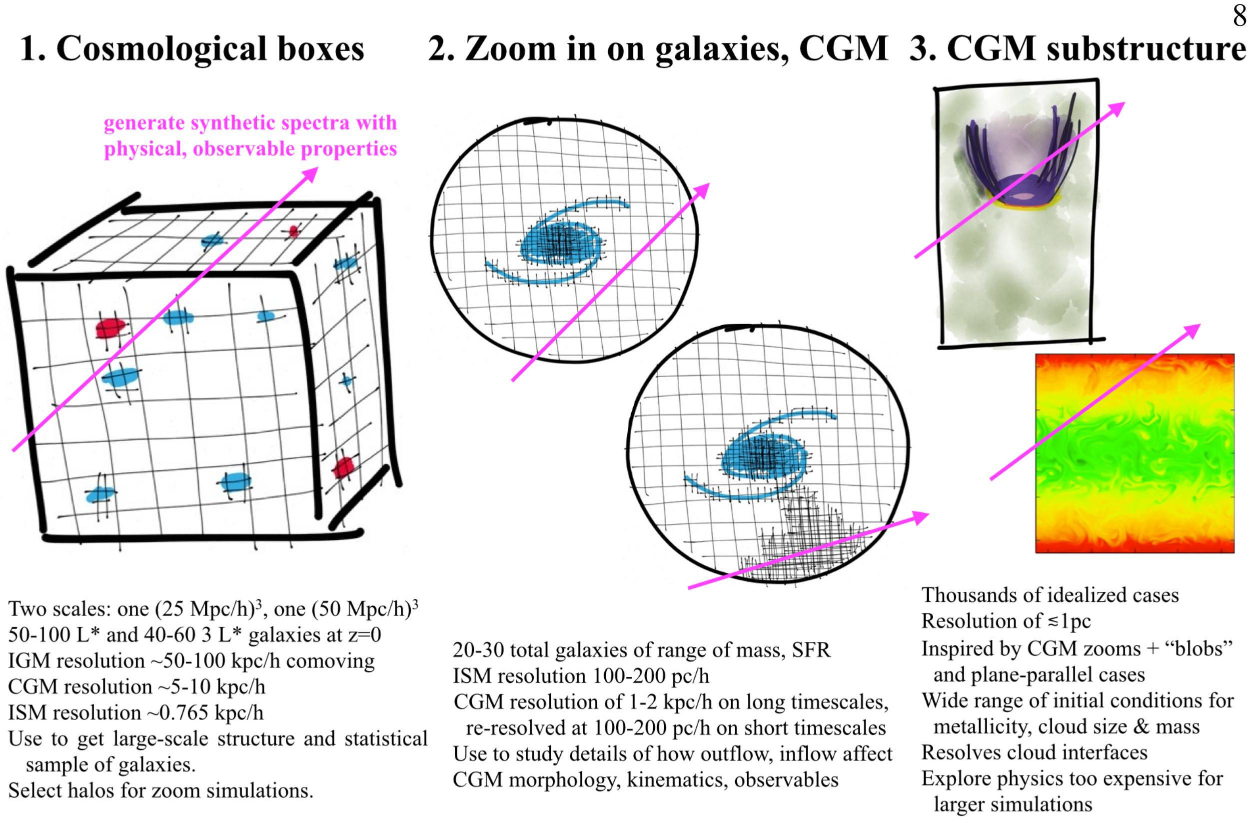

Figure 4: A schematic for our multi-scale simulations. Cosmological boxes will follow populations of galaxies and their CGM scaling with galaxy properties to kpc scales. Zoom simulations of selected galaxies within these boxes will resolve populations of CGM clouds and their surrounding medium at 100 pc scales. Idealized boxes extracted from the zooms will simulate individual CGM clouds and their interfaces down to sub-parsec scales.

the KOA, and from archival HST data to provide crucial information about the galaxies in those fields. One example is the set of KBSS (Rudie+2012) quasars, which are within KODIAQ, but which also have significant amounts of LRIS, DEIMOS and MOSFIRE low-res Keck spectra of galaxies within ~2 arcmin of the sightline, along with HST/WFC3 IR slitless spectra (Fig. 3). This will require a thorough search for complementary (HIRES, LRIS) data in the KOA, but we anticipate adding ≥ 10 more QSO fields with both Keck LRIS imaging and spectroscopy, where there will be ≥ 30 LLSs. The low resolution galaxy data is presently only in raw form on the KOA. The galaxy spectra will number in the thousands, so the reduction of the galaxy component of KODIAQ-Z will require significant effort. Moreover, the galaxies targeted in surveys like the KBSS were chosen to fit specific color/magnitude criteria, and thus likely exclude many galaxies whose CGM cause absorption in the spectrum. For these galaxies, which we will identify in the deep imaging available from KOA+HST, we will obtain additional spectra with the LBT (via ND proprietary time). The main tasks of reducing the quasar and galaxy data to science-ready levels will be carried out at St. Michaels, while the tasks of absorber identification and measurement will be done at Notre Dame.

Our main goals with our KODIAQ-Z study of the galaxy environment are two-fold. First, we will determine the covering factors (H I-weighted and metallicity-weighted) of H I, C II, Si II, Si IV, C IV, and O VI in the sample of LBGs at z~2–3 within 1 and 2 virial radii ($R_{\rm vir}$) and with velocity offsets between the galaxy and the absorbers of ≲ 300 km s^{-1}. We will determine if the covering factors of dense CGM gas (LLSs) in the metal rich and metal poor branches are different (if the bimodality even extends to these redshifts). Second, for the LBGs where a LLS is found within 1–2 R_{vir}, we will compare the metallicity of the absorber and the galaxy. LBGs at $z \sim 2$–3 have on average a 50% solar metallicity ([X/H]=−0.3; Erb+2006b), significantly higher than the log $N_{\text{HI}} > 19$ gas at similar epochs. With our proposed KODIAQ-Z study, we will increase the sample suitable for comparisons of gas and galaxy metallicity.

4. Solving CGM Resolution: Addressing the problems with multiscale simulations

We will address the multiscale puzzles of the evolving CGM using a combination of (1) physics-rich, high dynamic range *cosmological simulations* targeted at the scaling of CGM properties with galaxy mass and star formation, (2) *circumgalactic zoom* simulations to explore with high fidelity how galactic outflows and inflows affect the small-scale structure of the CGM, and (3) *"idealized" simulations* to explore specific physical phenomena of CGM substructure in a focused, constrained manner. This multi-scale approach is *essential*: idealized boxes give us physical conditions of gas and what it looks like when observed; zooms and cosmological boxes provide statistical samplings of clouds and show how they scale with galaxy properties. To this multiscale set of simulations, we will add a new toolkit (MISTY; separately funded) for generating mock spectra so that simulations can be "observed" with the same analysis techniques as the real Universe. This innovation will unlock a host of new capabilities to tie physical conditions directly to observables at the correct physical scales, to break projected and blended components into their real physical pieces, and to assess observational biases. In this section we describe these simulations (§4.1) and analysis tools (§4.2) in detail. Section 5 will describe how we will address the problems of the CGM using synthetic observations of these simulations.

4.1 A tiered approach to simulating multiscale physics of circumgalactic gas

Cosmological simulations: We will evolve six large galaxy formation simulations with the adaptive mesh refinement code Enzo (Bryan+2014) for two box sizes, 25 Mpc/h and 50 Mpc/h, with a 512^3 root grid and 512^3 particles, generated using the MUSIC cosmological initial conditions generator (Hahn & Abel 2011). This will result in a 7.3 x 10^6 $M_\odot$/h (5.85 x 10^8 $M_\odot$/h) dark matter particle resolution for the 25 Mpc (50 Mpc) volume and a root grid size of 48.8 (97.7) kpc/h comoving resolution; with 6 (7) levels of adaptive mesh refinement everywhere in the simulation volume, we will achieve a finest gridsize of 0.765 kpc/h comoving at all times. Resolving a given halo with 1,000 dark matter particles, this means that in the 25 (50) Mpc/h box all halos above roughly 10^{10} $M_\odot$ (10^{11} $M_\odot$) will be adequately resolved, and there will be roughly 50-100 L* (40-60 3 L*) galaxies by $z \sim 0$, and comparable numbers at higher redshifts. For each of the two box sizes, we will perform three simulations with variations in physics:

1. Moderate stellar and AGN feedback, and the Haardt & Madau 2012 ionizing background.
2. As above, but with a substantially higher level of feedback from stellar populations and AGN.
3. As #1, but with the Haardt & Madau 2001 background, which better produces the $z < 1$ H I column density distribution function (Kollmeier+2014).

These simulation volumes are large enough to generate "light cone" synthetic absorption spectra from, e.g., a sightline from z=0→4 without having significant unphysical correlation of large scale structure in the spectrum. The chosen mass and spatial resolution will result in galaxies with reasonably converged star formation histories (e.g., Scannapieco+2012). Furthermore, the three physics variations listed above, combined with the different box sizes and resolutions, will constrain the uncertainty in our numerical results that results from model choices. We will use these simulations to probe the origins of the bimodal metallicity distribution in dense CGM, the ionization state of the intergalactic and circumgalactic medium, and more generally the distributions of ionized gas around galaxies and in the intergalactic medium, at all redshifts where relevant data is available.

Circumgalactic Zoom Simulations: We will find 10–15 individual galaxies of interest from our two cosmological boxes and re-simulate these at 4–8 times higher spatial resolution (100–200 pc/h) with a variety of different physics packages, including varied star formation and feedback algorithms (including different assumptions about the stellar initial mass function), ultraviolet backgrounds, magnetic fields, and radiation transport. This will provide important insights into the cooling times and mass budgets for gas in the CGM, the formation of low-ionization gas around passive galaxies, and the origin of Lyman Limit Systems at low redshift. In addition, we can locally increase the resolution of the circumgalactic medium by factors of 4–16 (to resolutions of ~10–20 pc/h) after the simulations have been run to low redshift in order to examine the formation of the inferred cold phase of circumgalactic gas. This will also bridge the gap between cosmological boxes and the idealized calculations aimed at CGM substructure.

Circumgalactic Substructure in Idealized Simulations: We will use idealized calculations to explore specific physical phenomena in a constrained manner (a "numerical experiment", e.g., Meece+2014; Parrish & Stone 2005). This approach complements cosmological calculations in two ways. First, an idealized calculation generally does not have to include the full range of physics modules needed in a more general simulation, so it is often easier to understand the physical origins of an emergent phenomenon. Second, idealized calculations do not have to simulate a representative volume of the universe over a Hubble time, so they can be used to explore much smaller physical or temporal scales than a cosmological calculation. These calculations will resolve the formation of a multiphase medium in the CGM, and will be tuned to have properties that are analogous to either Milky Way-like galaxies or early-type galaxies. We will also experiment with physics that is too expensive (e.g., thermal conduction, non-equilibrium chemistry, radiation from local sources) or exploratory (non-ideal MHD) for cosmological boxes. These calculations are simple to design, execute, and analyze, and are inexpensive (10–1,000 CPU-hours each). We anticipate using several generations of such simulations to explore specific physical phenomena of interest, pursuing a range of calculations:

1. Idealized simulations of the CGM at high spatial ($\lesssim$ 1 pc) and time resolution, with initial densities, metallicities, temperatures, and velocity fields drawn from the zooms. We will do this by measuring the spherically-averaged density, temperature, metallicity, and gravitational acceleration from the zoom simulations and use these as initial conditions. We will then measure the power spectrum of density, temperature, and metallicity variations in the regions in question in the cosmological boxes, and use those to seed the initial fluctuations in the idealized calculations. These simulations will explore how metal mixing, cooling times, and the fate of this gas is affected, and how it appears when observed via absorption lines.
2. Idealized simulations of spherical clouds ("blobs") passing through an ambient medium with varying temperature, pressure, and metallicity contrasts and at varying Mach numbers. These are also aimed at exploring the circumgalactic medium at the resolution necessary to resolve individual clouds—and how the physics and physical conditions affect the observables.
3. Simplified models of the CGM: 2D Cartesian AMR calculations of a stratified, thermally unstable medium in a fixed gravitational potential at extremely high spatial and mass resolution, analogous to McCourt+2012 or Meece+2014.

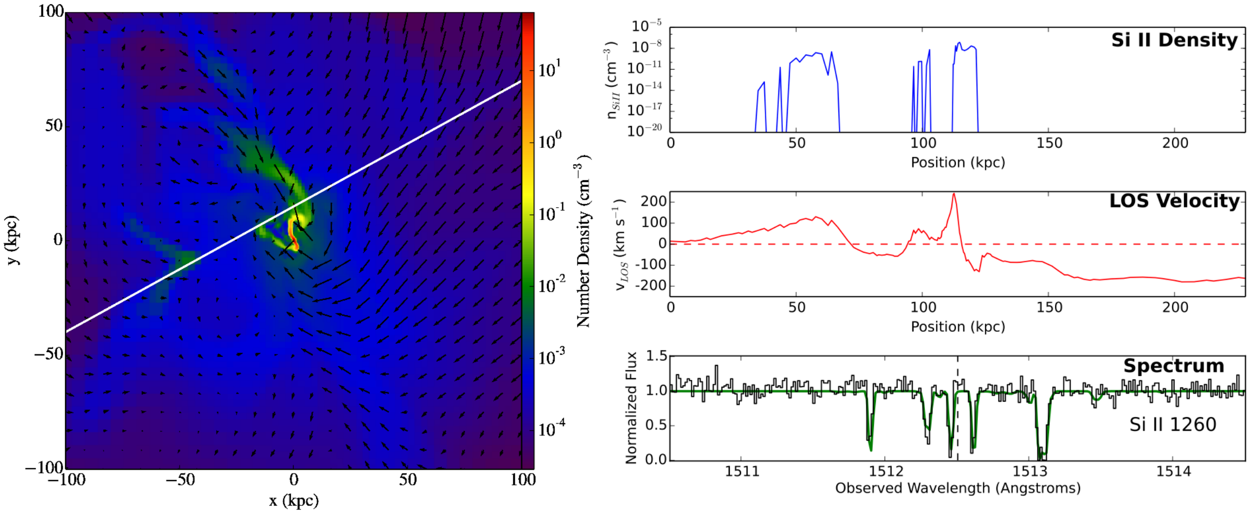

Figure 5: An illustration of synthetic spectrum generation using simulations and the yt software. This technique relates detected absorption (eg. Si II) back to the underlying physical variables. It is an essential piece of properly comparing small-scale CGM physics to observables. Our project will use the MISTY pipeline (Hubble Theory program 13919, PI Peeples) to generate mock spectra for analysis beside real data. (Figure from C. Hummels.)

4.2 Technical details of simulations

Enzo is an open-source, block-structured, Cartesian adaptive mesh refinement (AMR) code for the simulation of astrophysical phenomena (Bryan+2014; http://enzo-project.org). The code features a range of physics: prescriptions for star formation and feedback, different chemistry and cooling models (including non-equilibrium chemistry and ionization; Abel+1997, Anninos +1997), cosmic ray injection and transport, conduction, and radiation transport using both implicit flux-limited diffusion and a ray-casting method (Reynolds+2009; Wise & Abel 2011).

yt and MISTY: We will analyze our simulations with yt, an open source analysis and visualization toolkit for numerical simulations (Turk+2011 and http://yt-project.org). We will use yt as a platform for several critical analysis tasks, including multiwavelength synthetic observations of the galaxies and idealized calculations simulated with Enzo, leveraging tools previously developed by our collaboration (Smith+2011, Hummels+2013, Egan+2014). A significant piece of our analysis will be the generation of synthetic absorption-line spectra for direct comparison to the observations (see §5). Our team is developing the MAST Interface to Synthetic Telescopes with yt (MISTY) pipeline (Hubble Theory #13919, PI Peeples; co-I's incl. Hummels, O'Meara, O'Shea, and Tumlinson) with which enables extraction of absorption-line spectra from hydro simulations. For each of these synthetic spectra, the MISTY pipeline will deliver the corresponding physical properties (density, temperature, metallicity, kinematics, nearby galaxy information, etc.), along with line spread function-convolved noiseless and noisy spectra. These data products will be made available *and searchable* in MAST (**http://archive.stsci.edu**). The construction of the MISTY pipeline is fully funded. The JHU postdoc, supervised by Peeples and Collaborator Cameron Hummels will be leading the MISTY pipeline development; as part of this program, they will add new capabilities to this pipeline, such as the inclusion of the Keck HIRES line spread function and the ability to model the spectral break caused by Lyman-limit systems.

5. How data and theory combine to solve the problems

Our program is designed to apply time evolution and high resolution to the problem of constraining CGM mass flows. To accomplish this, we must have some way of combining simulations with data. Our approach is two-fold: (1) using the cosmological boxes and the zoom-in simulations, we will determine what the CGM looks like given our best understanding of galactic evolution, and how that compares to the observed evolution of the CGM, and (2) based on the idealized CGM simulations, we will determine what physical conditions are required to obtain CGM gas with the same distribution of observables as the data, using the results of the MISTY pipeline to compare synthetic absorbers with the real ones. We outlined in §3 how we will use KODIAQ-Z to establish a longer observational baseline for the four gas physics problems discovered at low z. Here, we describe how we will combine our multiscale simulations, and the synthetic spectra extracted from them, to address each problem:

The mixing problem: At its core, the mixing problem is one of scale: on what scales is dense circumgalactic gas well-mixed, and on what scales is it heterogeneous? The data at $z < 1$ shows that the $16 < \log N_{\rm HI} < 18.5$ gas is bimodal: about half is high-metallicity, half low-metallicity (Lehner+2014; Wotta+2014; Figure 2). At higher H I column densities, the metallicities are more smoothly distributed; at lower H I column densities, the situation is unclear at low-redshift. With KODIAQ-Z, we will extend this measurement to higher redshift and lower H I columns. We will use our suite of multi-scale simulations to address metallicity evolution as a function of H I column density (and the less observable but more physical 3D density) from $z = 4 \rightarrow 0$ to connect the KODIAQ-Z and COS measurements. We will use the zoom simulations to determine the occurrence and metallicity of dense H I gas in the CGM of galaxies of different masses, and then statistically populate the larger cosmological simulations with the zooms to piece together how often a random line of sight would pass through each of those occurrences. We will further use the CGM-zooms to 1) ensure that there is no dense gas that our traditional zooms are missing owing to their relatively lower resolution, and 2) add Lagrangian tracer particles to the gas and trace in detail mixing works in gaseous halos. This work will be led by the MSU graduate student and include collaboration with Cameron Hummels, who is a CoI on the MISTY team. Finally, we will "tune" a subset of our idealized CGM simulations to produce gas with an H I column density of $> 10^{16}$ cm^{-2}; these will be able to track in detail how gas mixing depends on pressure, metallicity substructure, cooling times, etc.; this part of the work will be led by the JHU postdoc. By combining these results, we will be able to directly answer the question of the origin and fate of the dense CGM: does the metal-rich gas originate in outflows, and will the metal-poor gas accrete onto galaxies?, or is there some other explanation?

The density problem: The mass flows powering galaxy evolution are the objects of CGM studies. To obtain masses and timescales, we must have a physical density, which we convert from observed column densities of ions by assuming an ionization model, usually photoionization (Werk+2014). The presence of singly-ionized magnesium, silicon, and carbon strongly disfavors collisional ionization; though the low-z UV background is uncertain (Kollmeier+2014), this does not affect the qualitative result that the low-ionization CGM appears to be at a much lower density than predicted by a two-phase model (e.g., Maller & Bullock 2004) in which this gas is in hydrostatic equilibrium with a virialized corona. Our tiered simulations will address this problem in two ways: 1) with their varying densities, temperatures, and metallicities, do the zoom simulations produce absorbers (i.e., sets of coincident absorption lines from different ions) resembling the data? If so, how do the physical conditions compare to what is derived from the Cloudy analysis? and 2) we will use the idealized simulations to determine what range of physical conditions can lead to the observed absorber properties. With the circumgalactic zooms and the CGM substructure simulations, we can "turn on" non-equilibrium chemistry and ionization in order to track how these processes affect the observables given a certain physical structure. This work will be led by the JHU postdoc.

The timescale problem: We will tune a subset of the idealized CGM substructure simulations to result in O VI column densities similar to data at both low (Tumlinson+2011) and high (Lehner +2014) redshifts and vary the physical conditions to determine what is required to have this material be long-lived. As with the density problem, we will explore the effects non-equilibrium chemistry and ionization have on our interpretation of the observables (e.g., Oppenheimer +2013). Combining these results with the galaxy and CGM zoom simulations to capture the broader dynamics, we will address whether or not the massive O VI-traced reservoirs found observationally are long-lived, or if there is some other explanation for the observed high column densities of O VI. As with the density problem, this work will be led by the JHU postdoc.

The fate problem: From the simulation side, the fate problem and the density problem are intimately connected, as the density problem is one of the origin and physical structure of circumgalactic low-ionization gas, while the fate problem is an issue of how it manages to not accrete back onto the galaxies, particularly passive galaxies. Therefore, we will use the same set of simulations we will use to address the density problem, but take it one step further by studying in detail the kinematics of this material and its interactions with surrounding gas. This work will be divided between the JHU postdoc, who will be leading the CGM substructure efforts, and the MSU postdoc, who will have developed the tools for analyzing the Lagrangian tracer particles needed to understand the fates of the dense CGM gas when studying the mixing problem.

The scale problem: Finally, we will tie what we learn from the higher resolution, idealized simulations back to the lower resolution simulations that capture fully cosmological scales. For example, though we are primarily focused here on the physics of gas in the CGM, the conditions of that gas will be affected by the details of the material flowing out of galaxies (its momentum, temperature, metallicity, density, etc.), but those details are determined not by circumgalactic physics but by interstellar. However, by comparing the idealized CGM substructure simulations to the more realistic zoom simulations, we will be able to place new constraints on feedback models that could be implemented in future generations of Enzo and other codes.

6. Intellectual Merit

The intellectual merit of our program rests on its potential to surmount the two key barriers—CGM *evolution* and *resolution*—to enable a major advance in understanding the circumgalactic medium of galaxies and the mass flows that drive their evolution. As a massive reservoir of galactic baryons and metals, the mediator of galactic feedback, and the fuel for star formation, the CGM plays a major role in galaxy evolution, which is one of the grand unsolved problems in astrophysics. Our program promises two major advances. First, we will pin down the physical state and metal content of the CGM at $z \gtrsim 2$, extending the timeline for galaxy/CGM co-evolution to 10 Gyr. Second, our CGM simulations will make a major advance in the physical insights we can draw from comparing data to theory at all the relevant physical scales. This work promises to help answer some key questions about how galaxies form and evolve as mediated by their halo gas. Among these are: how do galaxies sustain their star formation over > 10 Gyr when they have ~ Gyr worth of fuel? How is the galactic mass-metallicity relation produced? Is ISM gas consumed or ejected when galaxies turn passive? Where are the metals that are not seen in the galaxies that produced them? How much star formation is powered by gas recycled through the CGM? Answers to these questions address still broader questions of interest to all of astrophysics: what sets the masses, shapes, and colors of galaxies? What is the cosmic history of star formation? What are the origins of the chemical elements that form planets and life? Our program will make significant contributions to this hierarchy of cosmic origins questions by breaking through the key barriers that prevent us from understanding the circumgalactic medium.

7. Broader Impacts

The broader impact of this program will span four states and many levels of educational and public engagement. We will support the development of new visualizations of the hydrodynamical simulations to be included in a new planetarium show at MSU that will present the findings of this research to the public, first via the MSU planetarium and then more broadly through regular distribution channels (this work is budgeted at MSU). The show will aim to make spectroscopic data more approachable by showing the transition from the observed and simulated universe in emission (galaxies) to their larger environments (the CGM) to their imprints in lines of sight (spectroscopy). These visualizations will be used in public talks in Baltimore, South Bend, Burlington, and Lansing. The simulation data produced as a result of this project will be used in computational science courses at Michigan State University, where it will be used to train students in scientific visualization and data analysis techniques. The resulting curricular materials will be made available to the public. Materials from this program will be used in public observatory talks at JHU and at STScI's Youth for Astronomy and Engineering, which aims to engage children in science, especially those from groups underrepresented in STEM (neither requires funding). At Saint Michael's, the broader impacts of this research will be focused on undergraduate education and student research. Co-PI O'Meara has used KODIAQ data in the past to mentor SMC student research, with one undergraduate appearing in the Lehner et al (2014) publication. The research proposed here will facilitate student access to research in a simpler, more engaging way. With reduced data products in hand, undergraduates can go straight to analysis, minimizing startup times. Co-PI O'Meara will continue his history of science outreach by bringing our planetarium show to various public venues, such as the ECHO science center in Burlington, along with K-12 schools throughout northern Vermont. The MSU component of this project will include undergrad research students in its REU program and the MSU Honors College, which both target female and under-represented minority students. Finally, the proposed project will also have impact on scientists in training by supporting a graduate student and postdoc, who will gain experience doing cutting-edge research that combines theoretical and observational techniques. These young scientists will receive active mentoring as part of formal programs established at JHU/STScI and MSU.

8. Results from Prior NSF Support

Profs. Howk, Lehner, and O'Meara are jointly supported by NSF grant AST-1212012 ($460,058; 10/1/2012–9/30/2015) entitled Building Up and Tearing Down Galaxies: the Impact of Infalling and Outflowing Gas on Galaxy Evolution. **Intellectual Merit**: This program aimed to test fundamental predictions of currently preferred theories for the on-going assembly of galaxies. It focused on the distribution of CGM metallicities at $z<1$ and connecting the properties of the low-redshift CGM absorbers with the galaxies they feed. **Broader Impact**: This program will lead to a better qualitative and quantitative understanding of feedback and infall processes in galaxies and of the distribution and quantity of metals in galaxies and the Universe. It supports undergraduate and graduate educational initiatives and work on a planetarium movie about the cosmic history of the elements. This grant has led to publications discussed above, including Lehner+2013, Burchett+2013, and the in-preparation Wotta+2014. The work led by ND graduate student Christopher Wotta is based on new LBT spectroscopy coupled with a new approach to determining CGM metallicities providing a larger sample than previously obtainable.

Prof. O'Shea has been the PI or co-PI of several NSF-funded projects, most recently an NSF Division of Advanced Cyber-Infrsastructure Petascale Computing Resource Allocation grant (PRAC; OCI 0832662; “Formation of the first galaxies: predictions for the next generation of observatories,” $40K). **Intellectual merits:** This grant supported improvements to the scaling of the Enzo code to run a series of simulations relating to high-redshift galaxy formation on Blue Waters. This has resulted in several papers that have been published or are in the review process (Xu+2014, Chen+2014, Ahn+2014), as well as several additional papers that are in preparation and will be submitted in the near future. The data from this project will soon be made available via the National Data Service, heavily leveraging this investment in computational resources. **Broader impacts:** Simulation data from this project will be used in planetarium and television shows about astronomy, thus educating a broad international audience about this topic. Improvements made to the Enzo code’s infrastructure to improve its scaling performance are generalizable to other disciplines, enhancing technical understanding of high performance computing. The simulation data created in this project have been used by graduate students and postdoctoral researchers, including by female and minority researchers, as a training resource.

9. You might be wondering. . .

Why one big proposal instead of separate observational and theoretical proposals? We want to interpret observations in terms of physical conditions and to connect simulations to reality in the most direct and insightful way possible. To accomplish this, our team must include both observers and theorists who can bridge the gap between data and simulations by comparing the two in the observable space of detailed absorber statistics. We aim to guide the simulations with a “close read” of the data, and vice versa, which requires regular collaborative contact.

Do you really need a postdoc instead of a graduate student at Johns Hopkins? Yes. The postdoc will need to quickly ramp up to running and analyzing a large suite of idealized simulations and must become conversant with observations to make detailed comparisons between data and simulations. Mastery of all this is not a reasonable expectation for a beginning graduate student.

Won’t you need to secure a lot of computing time to run all these simulations? O’Shea has a track record of obtaining this level of computing time (21 million CPU-hours on a Stampede-level machine over three years), via the NSF XSEDE program and on the Blue Waters supercomputer through NSF PRAC or the Great Lakes Consortium, where he was allocated 124.8 million CPU hours last year. O’Shea will lead applications for computing time in support of this project. The inexpensive idealized simulations (~10,000 CPU hours) can be run locally by the JHU postdoc on JHU or STScI computing facilities to which Tumlinson and Peeples have access.

By mining the Keck and HST archives for data, aren’t you scooping the scientists of all of the hard work they put into collecting those data? KODIAQ and the KODIAQ-Z extension leverage the last two decades of Keck data in new ways. The HIRES sample in KODIAQ alone combines data from over 25 PIs in thousands of observations since 1995, placing them on a uniform footing to search for ionized gas in the CGM, a search which had not been performed to date. By extending to the full suite of ions and including information about the galaxies, KODIAQ-Z continues this philosophy. Our goal is not to supplant the science goals of the original PIs, but to integrate publicly available data into a readily usable form (reduced spectra, linelists, etc) both for our science objectives and to facilitate new discoveries by the community at large.