Astro2020 Science White Paper
The hidden circumgalactic medium

Thematic Areas:

☐ Planetary Systems
☐ Star and Planet Formation
☐ Formation and Evolution of Compact Objects
☐ Cosmology and Fundamental Physics
☐ Stars and Stellar Evolution
☐ Resolved Stellar Populations and their Environments
☑ Galaxy Evolution
☐ Multi-Messenger Astronomy and Astrophysics

Principal Author:
Name: Claudia Cicone
Institution: INAF - Osservatorio astronomico di Brera, Milano (Italy)
Email: claudia.cicone@inaf.it
Phone: +39 02 72320381

Co-authors: Carlos De Breuck (ESO); Chian-Chou Chen (ESO); Eelco van Kampen (ESO); Desika Narayanan (U. Florida); Tony Mroczkowski (ESO); Paola Andreani (ESO); Pamela Klaassen (UK ATC); Axel Weiss (MPIfR); Kotaro Kohno (U. Tokyo); Jens Kauffmann (MIT); Jeff Wagg (SKAO); Dominik Riechers (Cornell/MPIA); Bitten Gullberg (Durham); James Geach (U. Hertfordshire); Sijing Shen (U. Oslo); J. Colin Hill (IAS/CCA); Simcha Brownson (U. Cambridge).

Abstract: The cycling of baryons in and out of galaxies is what ultimately drives galaxy formation and evolution. The circumgalactic medium (CGM) represents the interface between the interstellar medium and the cosmic web, hence its properties are directly shaped by the baryon cycle. Although traditionally the CGM is thought to consist of warm and hot gas, recent breakthroughs are presenting a new scenario according to which an important fraction of its mass may reside in the cold atomic and molecular phase. This would represent fuel that is readily available for star formation, with crucial implications for feeding and feedback processes in galaxies. However, such cold CGM, especially in local galaxies where its projected size on sky is expected to be of several arcminutes, cannot be imaged by ALMA due to interferometric spatial scale filtering of large-scale structures. We show that the only way to probe the multiphase CGM including its coldest component is through a large (e.g. 50-m) single dish (sub-)mm telescope.
1 Scientific context and motivation

Galaxies grow and evolve within a network of dark matter through a finely-tuned exchange of baryons with the cosmic web, a process that is challenging to understand and model. The importance of describing baryons, their interplay and behaviour in and out of galaxies is two-fold. On the one hand, if our research were restricted to the cosmological framework, we would not be able to comprehend how the bright components of galaxies - the stars - formed out of dense gas clouds, hence how planetary systems originated, and eventually how life began. On the other hand, since the only observables against which we can test theoretical predictions are produced by baryons, a proper understanding of the complex mechanisms involving baryonic matter is essential to interpret the observable data and so test our cosmological paradigm. Hence, understanding the baryon cycle is an endeavour at the forefront of modern astrophysics, and it is relevant to many different fields, from Galactic studies to Cosmology.

The baryon cycle problem can be visualised as follows: galaxies are embedded in large gas haloes that we define “circumgalactic medium (CGM)”, extending by hundreds of kpc (up to the virial radius). A small portion of CGM material, the interstellar medium (ISM), has already cooled and flowed towards the central regions of galaxies, and it is directly involved in the processes of star formation and accretion onto supermassive black holes (SMBHs). However, most of the CGM mass lies outside the ISM realm and represents the interface between the ISM and the matter distributed outside galaxies, the intergalactic medium (IGM). The CGM is fed - from the inside - by galactic outflows and fountains [1, 2] - and, from the outside - by cosmic streams [3, 4] and galaxy mergers [5]. Furthermore, the CGM gets depleted via accretion onto galaxies, and through expulsion into the IGM by energetic feedback processes [6–8]. The CGM is therefore intrinsically linked to the feedback and accretion processes in galaxies, and carries the imprint of their evolution.

Although traditionally the CGM is thought to be dominated by hot and warm gas [9], with little or no cold gas beyond the ISM scales, this picture has been challenged by recent breakthroughs in this field. On the one hand, the lack of cold gas in most simulated galaxy haloes appears to be a bias of hydrodynamical simulations rather than a theoretical prediction. More specifically, several studies are now finding that increasing the spatial resolution on halo scales (i.e. beyond the star forming ISM) naturally leads to an enhanced cold gas fraction in the CGM, because it allows to properly capture cooling processes and to spatially resolve smaller cold clumps embedded in a warmer medium [10–12]. On the other hand, a massive cold ($T < 10^4$ K) CGM reservoir may result from cosmic feeding, which at high redshift is dominated by cold filamentary inflows, but also from powerful feedback. Indeed, recent studies have shown that intense episodes of star formation and active galactic nuclei (AGN) can drive massive ($\dot{M}_{\text{out}} \sim 1000 \, M_\odot \, yr^{-1}$) outflows of H\textsc{i} and H\textsc{2} gas. These can extend by several kpc [13, 14] and embed very dense ($n > 10^4 \, \text{cm}^{-3}$) clouds [15, 16]. However, although H\textsc{2} outflows can be powerful enough to clear the central regions of galaxies within a few Myr, they are likely incapable of expelling all of the molecular ISM from the halo, both in starbursts [17] and in luminous quasars [18, 19]. Instead, cumulative effects of such outflows may be: (i) the transportation of material to CGM scales, where most of the (metal-enriched) H\textsc{2} and H\textsc{i} gas stalls without escaping the halo [2, 20], and (ii) the enhancement of gas cooling on halo scales [1, 21, 22]. Therefore, we expect the CGM to be multiphase and entrain not only diffuse ionised gas, but also colder and denser clumps. Such cold and non-pristine gas represents fuel that, once (re)-accreted onto the ISM, would be readily available to feed star formation, and hence it would have a crucial role for galaxy evolution.
As we argue in this paper, only (sub-)millimeter ((sub-)mm) and far-infrared (FIR) observations can probe the multiphase CGM including its coldest component, and so deliver a complete census of baryons in galaxy haloes. Such measurements are however extremely challenging even for ALMA because they require high sensitivity to large-scale emission and wide field of views (FoVs) which only a large (e.g. 50m) single-dish can provide.

2 State-of-the-art and forthcoming experiments

Currently, we have very little observational constraints on the mass and physical properties of cold gas on CGM scales. The CGM of high-z (radio quiet) AGNs has so far been investigated predominantly in its diffuse ionised and atomic phases through absorption features detected in rest-frame UV spectra. Interestingly, some of these works hinted at a higher cold gas fraction in the CGM of luminous quasars at $z \sim 1 - 3$ compared to inactive galaxies at similar redshifts and with similar masses, especially at large projected velocities of $|v| > 300$ km s$^{-1}$ [23]. Recently, the ubiquitous detection of low surface brightness $\gtrsim 50 - 100$ kpc-size Ly$\alpha$ nebulae, permitted by instruments such as MUSE at the VLT and KCWI at the Keck telescope, has enabled direct imaging of the CGM at $3 < z < 6$ [24–26]. The properties of Ly$\alpha$ haloes around $z \gtrsim 3$ quasars suggest entrainment of cold ($T \lesssim 10^4$ K) and dense (density, $n \gg 1$ cm$^{-3}$) clumps up to $r \sim 50$ kpc in these sources [27]. However, the absorption of Ly$\alpha$ photons and the resonant nature of Ly$\alpha$ emission prevent a detailed kinematical study of the CGM using this tracer. Furthermore, the presence of a ionising source (such as a quasar) can significantly boost the Ly$\alpha$ flux, compared to what is expected, due to the ubiquitous fluorescent signal stimulated by the high UV background [28]. Therefore, Ly$\alpha$ imaging cannot provide an unbiased census of the high-z CGM.

The ionised and neutral CGM, including any cold and dense H$_2$ gas, can be directly imaged through FIR fine structure lines (such as [CII] 158\,$\mu$m, [C1] 609\,$\mu$m & 370\,$\mu$m, [NII] 205\,$\mu$m & 122\,$\mu$m, [OIII] 88\,$\mu$m, etc) and molecular rotational transitions (e.g. CO, HCN, CN, CS). At high-z, FIR lines are redshifted into the (sub-)mm atmospheric windows, allowing ground-based observations. Thanks to the bright and multiphase (HII, H1, H2) nature of [CII] [29], sensitive [CII] line observations hold the potential to directly image faint and diffuse CGM components at $T < 10^4$ K. Cicone et al. [30] demonstrated the existence of copious amounts of [CII]-emitting gas in the CGM of a luminous quasar at $z = 6.4$, which was also found to host the most extended ($r \sim 30$ kpc) high-velocity ($v \sim 1400$ km s$^{-1}$) outflow of cold gas known. By modelling the $uv$ visibilities, 70% of the [CII] flux at low projected velocities was found to trace a 20 kpc-scale nebula, while only 30% of [CII] arises from the compact ISM. This result completely overturned the scenario favoured by earlier interferometric [CII] observations of the same source, which, lacking short antenna spacings, resolved out all of the extended emission.

The cold CGM may be dense and bright enough to be detected also in CO line emission. The so-called ‘Spiderweb’ galaxy, a protocluster at $z \sim 2.2$ with a central bright radio galaxy, hosts a 70 kpc-size halo of molecular gas [31, 32], which does not arise from individual galaxies but from the extended CGM. Surprisingly, its carbon abundance and excitation properties are similar to a normal star forming disk, hence supporting a ‘recycled’ origin of the CGM - due to a combination of outflows, mass transfer among galaxies, gas accretion, and mergers. Interestingly, the CO(1-0) emission seen in the Spiderweb galaxy is so diffuse that it was resolved out by the JVLA, while being detected at high significance by the shorter baselines of ATCA [see also 33]. Molecular haloes may not be unique to powerful quasars. Ginolfi et al. [34] discovered a 40 kpc-size CO(4-
3) structure around a massive galaxy at \( z = 3.5 \), where most of the molecular line emission is not associated with the ISM of the main galaxy or its satellites. These pilot studies suggest that FIR and (sub-)mm line observations of the high-z CGM carried out with a high sensitivity to large-scale structures hold a huge discovery potential.

The variety of (sub-)mm and FIR lines would allow us not only to detect any hidden massive CGM reservoir, but also to constrain its physical properties such as density, temperature, ionised-to-neutral gas fraction, ionisation parameter, and metallicity [35, 36]. Moreover, the unique chemical properties of some molecular ions (OH\(^+\), CH\(^+\), H\(_2\)O\(^+\)) make them suitable to explore physical processes that can critically affect galaxy growth, such as turbulence and shocks [37, 38]. These mechanisms are strictly associated to the baryon cycle (feedback through outflows and feeding by inflows) and so are particularly relevant to CGM gas. For example, ALMA imaging of CH\(^+\) lines, detectable in absorption from diffuse molecular gas [39], has recently revealed hidden massive reservoirs of highly turbulent cool (\( T \sim 100 \) K) gas, extending to \( > 10 \) kpc, far beyond the \( \sim 1 \) kpc-size ISM of (lensed) dusty star forming galaxies at \( z \gtrsim 2 \) [40]. The same observations detected extremely broad (FWZI \( \sim 2500 \) km s\(^{-1}\)) CH\(^+\) emission tracing very dense gas (\( n \sim 10^5 \) cm\(^{-3}\)) entrained in powerful galactic outflows, which are probably feeding the large-scale diluted turbulent H\(_2\) CGM reservoir seen in absorption.

In parallel, high sensitivity, wide FoV, broadband continuum observations will open interesting new avenues to the investigation of the hidden CGM. Bright high-z FIR/(sub-)mm continuum emitters can be used as background sources to detect diffuse, low column density CGM gas in absorption through virtually any molecular tracer (e.g. CO, HCN, HCO\(^+\), OH, OH\(^+\), CH\(^+\), [41, 42]). Such experiments require a large FoV to sample several continuum emitters and so overcome the likely low filling factor of the CGM absorbers. They also require high spectral resolution, since molecular line absorbers usually have line widths of \( \sim 10 \) km s\(^{-1}\).

Finally, sensitive continuum observations will enable the detection of the thermal Sunyaev-Zeldovich (tSZ) effect from the warm/hot phase of the CGM. The tSZ signal, being proportional to the integral of electron pressure, provides calorimetry of the ionised gas [e.g. 43, 44]. This, in turn, provides strong constraints on galaxy formation models, whose predictions span a wide range of tSZ signals depending on the AGN feedback prescriptions (see e.g. [45], and white paper by Battaglia & Hill et al.). In addition, the integrated tSZ signal can give an estimate of the total mass of a system. Stacking of Planck, ACT, and SPT data [46, 47] has allowed to reach down to galaxy mass scales, with typical fluxes of a few tens of \( \mu \)Jy\(^1\). Thus tSZ measurements of the CGM require good sensitivity and likely next generation (sub-)mm instruments.

3 Opportunities to make a big step forward

Improving over ALMA: the quest for a large (sub)-millimetre single dish:

The cold neutral CGM, and in particular the molecular CGM, is an almost completely unexplored avenue, due to the lack of (sub-)mm facilities that are able to detect and trace large-scale, low surface brightness structures. The latter are indeed filtered out by interferometers, as explained below, and undetected by current single dish facilities because of their low sensitivity (e.g. the 12-meter APEX telescope or the ALMA Total Power antennas). Larger dishes such as the LMT 50-m or the IRAM 30-m telescopes have a limited frequency coverage, imposed by the atmospheric

\(^1\)For example, a total mass of \( M \sim 10^{11} \) \( M_\odot \) produces a negative continuum signal of \( \sim -70 \) \( \mu \)Jy at \( z = 0.15 \).
conditions at the telescope site as well as by the antenna surface accuracy which make observations at $\nu > 345$ GHz almost prohibitive.

ALMA has transformed our understanding of galaxies at both high and low redshift, but there are limitations to its capabilities. Due to spatial scale filtering, an interferometer only detects a fraction of the total flux density for sources with emission on scales larger than its shortest antenna spacing (baseline), even in mosaicked observations. The maximum recoverable scale (MRS) of an interferometer is $\sim 0.6\lambda/b_{\text{min}}$, where $b_{\text{min}}$ is the shortest baseline (close to the dish diameter), although the exact MRS is best determined through simulations because it depends on both the array configuration and the details of the large-scale structure. For the ALMA 12-m array, MRS $\lesssim 3''$ at 850 GHz (MRS $\lesssim 8''$ for the ACA 7-m array) [48]. An MRS of $3''$ corresponds to a largest detectable angular scale of $\sim 45$ pc at the distance of M 82 ($z = 0.00073$, distance of 3 Mpc), and of $\sim 0.6 - 3$ kpc at $z = 0.01 - 0.05$, hence preventing the study of not only the CGM, but also of any large-scale ISM structure (such as a galactic outflow or a tidal tail) in nearby galaxies. However, at higher-$z$, a MRS $\sim 3''$ corresponds to $\sim 26 - 16$ kpc at $z = 2 - 7$, which in principle allows some CGM science with ALMA (see §2) but only limited to much smaller structures compared to those detected in Ly$\alpha$ emission (with sizes of $\sim 100$ kpc, e.g. [24]).

To further illustrate the advantage of a large single dish (SD) antenna vs ALMA in observing extended sources, we show here a sensitivity comparison. More specifically, we consider the ALMA 50 element array of 12-m diameter antennas, and compare it with a 50-m diameter SD antenna hypothetically placed near the ALMA site. The rms noise for an interferometric observation scales as $\sigma_{\text{rms}}^{\text{ interf}} \sim T_{\text{sys}}/[A_{\text{eff}}(N_{\text{base}} \Delta \nu t)^{1/2}]$, where $T_{\text{sys}}$ is the noise contribution of the atmosphere and receiver, $A_{\text{eff}}$ is the effective collecting area of each element, $t$ is the integration time, $\Delta \nu$ is the spectral bandwidth, and $N_{\text{base}}$ is the number of baselines, which for $M$ array elements equals to $N_{\text{base}} = M (M-1)/2$ (for example, for the ALMA 12-m array and $M = 50$ antennas, $N_{\text{base}} = 1225$). For a SD telescope, equipped with a heterodyne focal plane array (an instrument that delivers spectro-imaging information, i.e. 3D datacubes), the relation is similar, but we redefine $N$ to be the number of independent beams on the sky, so that $\sigma_{\text{rms}}^{\text{SD}} \sim T_{\text{sys}}/[A_{\text{eff}}(N_{\text{beams}} \Delta \nu t)^{1/2}]$. The effective area $A_{\text{eff}}$ of an antenna scales with its geometric area, hence $A_{\text{eff}}^{\text{SD}} = 17.36 A_{\text{eff}}^{\text{ALMA}}$, and so $\sigma_{\text{SD}}$ matches $\sigma_{\text{ALMA}}$ for $N_{\text{beams}} = 4$, if one assumes the same $\Delta \nu, T_{\text{sys}},$ and $t$. However, one should consider that the FoV of a 50m-antenna is smaller by a factor of $(50/12)^2$ than the FoV of ALMA. Hence, the sensitivities per FoV of ALMA and of a 50-m SD become comparable when the source area fills at least $\sim 70$ beams. At 850 GHz (resolution of 1.45″) and for typical beam spacings (i.e. $2f\lambda$, $f=$focal ratio), this is true for sources $> 24''$ across, hence at larger sizes the mapping speed of the 50-m SD outperforms the full ALMA array. Furthermore, we note that a SD telescope is much easier to keep at the forefront of upgraded technologies compared to a 50-element array. For instance, the new generation heterodyne receivers mounted on the APEX telescope already provide instantaneous spectral bandwidths (per polarization and per pixel) that are twice as large as the current ALMA bandwidth.

Figure 1 compares the capability of such hypothetical 50-m SD telescope (see, e.g. [49–53]) with that of the ACA and ALMA arrays in mapping large-scale emission from the cold CGM at $z = 0.02$. The input image is drawn from the state-of-the-art SIMBA cosmological simulations with a $(25/h\text{Mpc})^3$ volume [54]. We selected a star-forming (SFR = 11 M$_\odot$ yr$^{-1}$) disk galaxy with stellar mass $M_*$ = 1.3 $\sim 10^{11}$ M$_\odot$, and total CGM gas mass $7.75 \times 10^{10}$ M$_\odot$. The equilibrium molecular line luminosities are calculated simultaneously by employing the thermal-chemical-radiative equilibrium code DESPOTIC [55]. We then projected the particle information
onto a uniform $1024^3$ grid, and calculated the line-of-sight velocity-integrated intensities. The latter have been re-normalised to 1 Jy km s$^{-1}$, since our purpose is to compare how much flux density is recovered by the different facilities in $t = 10$ hours. As shown in Fig. 1, the molecular line flux is spread across a large projected area of $\sim 200''$, corresponding to 80 kpc at $z = 0.02$, which is not recovered by the interferometers. Clearly, for such low-$z$ galaxies, ALMA and ACA are not optimised to probe the CGM because of (i) their small beam (even in the most compact configuration) which requires large mosaic observations to probe the full source extent (implying a lower integration time per pointing), and of the (ii) spatial scale filtering of large-scale emission, which affects even mosaic observations. The best results are obtained with the 50-m SD antenna, which we have conservatively assumed to be equipped with a single-pixel ALMA detector (but technology advancements will soon deliver much more sensitive multi-pixel detectors hence significantly shortening the SD observing time, [53]).

Of course, the most important factor to consider for ALMA vs a 50-m SD in the Atacama desert is their complementarity: on the one hand, as shown by Fig. 1, ALMA will never access scales large enough to constrain the full extent - and mass - of the CGM - for this, single dish mapping is required. On the other hand, a 50-m SD will be limited to 1.5'' spatial resolution at 850 GHz. Only joint SD and ALMA mapping can thus provide a complete view of the CGM over a broad range of spatial scales and cosmic times.

Figure 1: The expected molecular CGM of a star forming galaxy at $z = 0.02$ as seen by different (sub-)mm facilities in 10 hours of integration. The observing frequency corresponds to the CO(3-2) line (Band 7). Panel (a) shows the input image (details in main text). Panels (b-d) show CASA simulations corresponding to: (b) an on-the-fly map obtained with a 50-m SD telescope equipped with a single ALMA detector; (c) a mosaic map obtained with the ACA 7-m array; (d) a 576 pointings-mosaic map obtained with the ALMA 12-m array in compact configuration.

**Synergies with astronomical facilities planned beyond 2030s:**

Capabilities to map low surface brightness H\textsc{i} 21cm emission from galaxies up to high-z, realised by (near-)future projects including the SKA [56] and the ngVLA [57], will be highly complementary to the (sub-)mm-based studies of the denser CGM described here. However, until the full development of the SKA, FIR tracers of H\textsc{i} gas such as the [C\textsc{ii}] line will remain the only way to probe such phase at $z > 1$. Promising avenues to detect the CGM at $z > 3$ will be through [C\textsc{ii}] line intensity mapping surveys, to be carried out, e.g., with CCAT-prime ([58], see also Kovetz et al. A2020 WP), TIME [59], or CONCERTO [60]. These experiments will detect [C\textsc{ii}] both from the host and from the diffuse CGM, without resolving individual structures, hence requiring higher-resolution follow ups. The CGM is also one of the main science goals of Athena (2030s), which will trace the warm/hot phase, of the Simons Observatory [61] and CMB-S4 [62] (2020s), which will utilize the thermal and kinematic SZ effects to probe ionized gas, and of LUVOIR (2030s), which will detect the CGM in absorption in the UV band.
References


[57] E. J. Murphy, A. Bolatto, S. Chatterjee, C. M. Casey, L. Chomiuk, D. Dale, et al. The ngVLA Science Case and Associated Science Requirements. In Eric Murphy, editor, Science with a Next


