Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **534,** 1247–1256 (2024) https://doi.org/10.1093/mnras/stae2115
Advance Access publication 2024 September 10

The HYENAS project: a prediction for the X-ray undetected galaxy groups

Weiguang Cui [1,2,3★†] Fred Jennings,[1] Romeel Dave [1,4] Arif Babul[1,5,6‡] and Ghassem Gozaliasl [7,8]

[1]*Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh, EH9 3HJ, UK*
[2]*Departamento de Física Teórica, M-8, Universidad Autónoma de Madrid, Cantoblanco, E-28049, Madrid, Spain*
[3]*Centro de Investigación Avanzada en Física Fundamental (CIAFF), Universidad Autónoma de Madrid, Cantoblanco, E-28049 Madrid, Spain*
[4]*Department of Physics and Astronomy, University of the Western Cape, Robert Sobukwe Rd, Cape Town, 7460,South Africa*
[5]*Department of Physics and Astronomy, University of Victoria, 3800 Finnerty Road, Victoria, V8P 1A1, BC, Canada*
[6]*Infosys Visiting Chair Professor, Indian Institute of Science, Bangalore, 560012, India*
[7]*Department of Computer Science, Aalto University, P. O. Box 15400, Espoo, FI-00076, Uusimaa, Finland*
[8]*Department of Physics, University of Helsinki, P. O. Box 64, Helsinki, FI-00014, Uusimaa, Finland*

Accepted 2024 September 5. Received 2024 September 4; in original form 2024 June 14

ABSTRACT

Galaxy groups contain the majority of bound mass with a significant portion of baryons due to the combination of halo mass and abundance (Cui 2024). Hence they serve as a crucial missing piece in the puzzle of galaxy formation and the evolution of large-scale structures in the Universe. In observations, mass-complete group catalogues are normally derived from galaxy redshift surveys detected through various three-dimensional group-finding algorithms. Confirming the reality of such groups, particularly in the X-rays, is critical for ensuring robust studies of galaxy evolution in these environments. Recent works have reported numerous optical groups that are X-ray undetected (see, e.g. Popesso et al. 2024), sparking debates regarding the reasons for the unexpectedly low hot gas fraction in galaxy groups. To address this issue, we utilize zoomed-in simulations of galaxy groups from the novel HYENAS project to explore the range of hot gas fractions within galaxy groups and investigate the intrinsic factors behind the observed variability in X-ray emission. We find that the halo formation time can play a critical role–we see that groups in haloes that formed earlier exhibit up to an order of magnitude brighter X-ray luminosities compared to those formed later. This suggests that undetected X-ray groups are preferentially late-formed haloes and highlights the connection between gas fraction and halo formation time in galaxy groups. Accounting for these biases in galaxy group identification is essential for advancing our understanding of galaxy formation and achieving precision in cosmological studies.

Key words: galaxies: evolution – galaxies: groups: general – galaxies: haloes – X-rays: galaxies: clusters – X-rays: general.

1 INTRODUCTION

Galaxy groups, typically with a halo mass in the range of $\sim 10^{12.5}$–$10^{14}\,M_\odot$ (Liang et al. 2016), occupy the most mass and contain the majority of galaxies in the Universe (Cui 2024). They also serve as the primary environment for key galaxy evolution processes such as galaxy transformation driven by the interplay between galaxies and their surrounding gaseous haloes (O'Sullivan et al. 2017). Note that the definition of galaxy group is not clear in observation for several reasons: the blended boundary between group and cluster; different observables yield different definitions; and the uncertainties in halo mass estimation. As such, we adopted a slightly wider halo mass range to select the sample. However, we argue that this study is not affected by the specific group definition. This is because our results are mostly shown as a function of halo mass. Galaxy groups offer valuable insights into fundamental physical processes such as galaxy quenching via active galactic nuclei (AGN) feedback (see, e.g. Bahar et al. 2024; Eckert et al. 2024; Yang et al. 2024), heating and cooling of the intragroup medium (Oppenheimer et al. 2021), the diverse kinematical and morphological properties of their central galaxies (e.g. Loubser et al. 2018; Jung et al. 2022), and the departure from self-similarity observed in clusters (see, e.g. Yang et al. 2022). Furthermore, galaxy groups offer constraints on galaxy formation, cosmological parameters, black hole–galaxy co-evolution, and environmental transformation (see, e.g. Eckert et al. 2021; Lovisari et al. 2021; Oppenheimer et al. 2021, for recent reviews).

Despite their importance, galaxy groups have not received as much attention as galaxy clusters, mainly due to the difficulty in their detection. They are faint in the X-rays, with the temperature of the diffuse gas typically ranging from approximately 0.3–2 keV (e.g. Mulchaey 2000; Liang et al. 2016), resulting in X-ray luminosities typically ranging from $\sim 10^{40}$–10^{43} erg s^{-1} which is several orders of magnitude lower than clusters (Lovisari et al. 2021, and references therein). Additionally, they contain far fewer member galaxies than clusters, typically ranging from a few to several dozen (George

★ E-mail: cuiweiguang@gmail.com
† Talento-CM fellow
‡ Leverhulme Visiting Professor

et al. 2011), which makes them more challenging to identify robustly via group-finding algorithms. This can further lead to false identifications due to chance projections (see, e.g. Pearson et al. 2017; Li et al. 2022). Detecting hot gas in low-mass galaxy groups (i.e. $M_h \in 10^{12.5-13.5}\,\mathrm{M}_\odot$) is often regarded as the gold-standard for validation, but their X-ray faintness limits the number of verified low-mass groups and biases them towards lower redshifts ($z < 0.4$) (see, e.g. O'Sullivan et al. 2017; Gozaliasl et al. 2019). Although optical and X-ray surveys continue to improve, it is critical to understand any biases introduced by group selection in order to interpret observations properly.

As an example, the scatter in their properties can introduce Malmquist biases in the inferred physical characteristics (Gozaliasl et al. 2020). Damsted et al. (2023) noted a significant increase in the scatter of L_X compared to other mass proxies below a redshift of 0.15, primarily in low-mass clusters, which hampers the effectiveness of X-ray observations in providing a comprehensive understanding of these groups. Recent studies by Khalil et al. (2024) corroborated these findings using the AXES-2MRS galaxy groups, which combined data from the ROSAT All-Sky Survey (RASS) with the Two Micron Redshift Survey (2MRS) Bayesian Group Catalogue. They further suggested that both feedback mechanisms and halo concentration are the reasons for the substantial scatter in the properties of X-ray groups, emphasizing that the scatter of scaling relations offers valuable insights into the underlying physics of galaxy groups.

The large variations in the X-ray brightness among galaxy groups can result in legitimately significant groups remaining undetected in X-ray surveys. It has been proposed that only galaxy groups with a central elliptical galaxy tend to exhibit diffused X-ray emission (e.g. Mulchaey et al. 2003, and references therein), a phenomenon contingent upon the detection limits of X-ray telescopes. Utilizing the *Chandra* X-ray Observatory, Pearson et al. (2017) investigated 10 relaxed galaxy groups carefully selected from the GAMA optical galaxy catalogue to mitigate spurious and projection effects. They observed that nine out of ten groups were underluminous in X-rays by a mean factor of approximately 4 compared to typical X-ray-selected samples. Hence, the converse practice of identifying X-ray samples and then seeking their counterparts in optically selected group catalogues may also introduce biases. Recent work by Damsted et al. (2024) expanded the findings of Manolopoulou et al. (2021) from galaxy clusters to galaxy groups–galaxy clusters/groups in overdense environments tend to have higher X-ray luminosities, which they hypothesized is driven by halo assembly bias. Another recent work by Popesso et al. (2024) directly combined data from eROSITA with the updated GAMA catalogue, revealing that 157 out of 189 systems with $M_{200} \geq 10^{13}\,\mathrm{M}_\odot$ and $z < 0.2$ remained undetected in X-rays. Hence, there are significant biases introduced either when selecting groups in the optical or the X-ray (see recent findings, e.g. O'Sullivan et al. 2017; Popesso et al. 2024). Quantifying these biases and understanding their physical origin is essential for groups to be leveraged for galaxy formation and cosmological studies.

In this work, we examine the physical origin of the scatter in properties of galaxy groups using the HYENAS suite of group-scale zoom simulations (see Section 2 for details). HYENAS is a new suite that re-simulates 120 group-size haloes drawn from a large-volume cosmological simulation employing the successful SIMBA galaxy formation model (Davé et al. 2019). Its novelty lies in its selection, which is based on bins in both halo mass and halo formation time. The latter is often implicated as a key driver in the scatter in galaxy group properties (see, e.g. Cui et al. 2021). Here, we investigate what implications the variations in group halo formation times can have on their detectability in X-ray and optical surveys, and thereby quantify associated selection biases.

Table 1. The cosmology parameters used in the SIMBA simulations.

H0 [km s^{-1} Mpc^{-1}]	Ω_Λ	Ω_m	Ω_b	σ_8	n_s
68	0.7	0.3	0.048	0.82	0.97

2 THE HYENAS PROJECT

The HYENAS project is a branch of the SIMBA with its focus on galaxy groups using the zoom re-simulation technique. We refer to a companion HYENAS paper, Jennings et al. (2024), for the X-ray cavity studies. While there have been many cluster-scale zoom projects, for example, the 300 project (Cui et al. 2018), group-scale zooms are less common. One reason is that large-volume cosmological simulations already contain many groups. However, when selected carefully, zooms can sample outliers in the distribution that are not well represented in a random sample. Also, zooms offer the opportunity to achieve higher resolution at a modest computational cost, enabling resolution convergence studies, though, in this introductory work, we do not employ that aspect of HYENAS.

Besides selecting zoom haloes in the group mass regime, HYENAS further selects objects with a wide range in halo formation times. This is motivated by Cui et al. (2021) who argued that halo formation time is the key determinant of the scatter in the stellar-to-halo mass relation, as well as the cold versus hot gas content of haloes. This will presumably also impact the X-ray properties of these systems, which is relevant for this work. Next, we describe the sample selection and X-ray analysis of the HYENAS zoom suite.

2.1 The HYENAS sample and initial condition generation

The simulation code and parameter choices used are identical to that in the SIMBA simulation, described in Davé et al. (2019) and many subsequent papers. For brevity, we do not repeat these here but focus on the aspects novel to the HYENAS zoom suite.

To increase the sample of galaxy groups, we first run a $200\,h^{-1}$ Mpc dark-matter-only simulation (8× SIMBA's volume), with the same dark matter particle mass resolution and *Planck*-concordant cosmology (see Table 1) as SIMBA using Gadget-4 (Springel et al. 2021). From this, we select 120 (out of ∼10 k) haloes with $M_{200c} \in 10^{12.5}$–$10^{14}\,\mathrm{M}_\odot$ (where '200c' denotes 200 times the critical density). However, these are not selected randomly; rather, within each 0.5-dex mass bin, we select the galaxies covering a spread halo formation times in percentile bins. The formation time is calculated as the time when half the $z = 0$ halo mass has assembled within the halo's main progenitor. Gadget-4's on-the-fly merger tree generation makes this calculation straightforward.

Fig. 1 shows the sample selected in 5 halo mass bins and 8 formation time bins, with the latter chosen with percentiles bounds of 0-2-5-20-50-80-95-98-100 within each mass bin, and are marked as green-dotted horizontal lines in Fig. 1. Inside each region in this space, we randomly select three haloes to re-simulate. The central galaxy's density map of one example within each region is shown as an inset image. Note that this is only for illustration, as the image covers 4 times the galaxy's half-mass radii by blending both stars and dark matter with arbitrary normalization using the Py-SPHViewer.

The large N-body volume's random initial conditions are generated using MUSIC (Hahn & Abel 2011), which conveniently allows us to generate these zoomed-in initial conditions (ICs) for the HYENAS

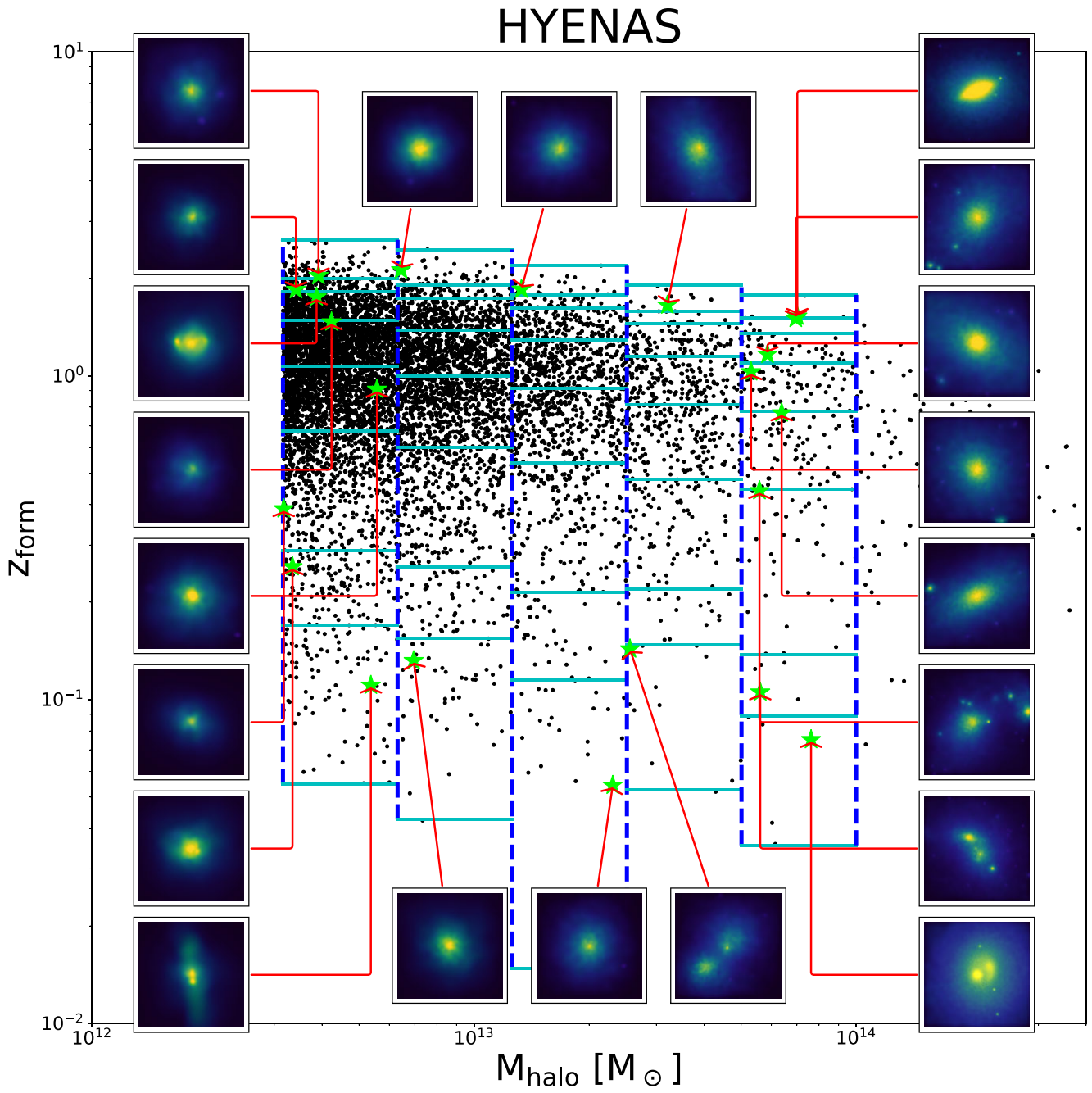

Figure 1. The halo mass–formation time relation from the parent dark-matter-only simulation with illustrations of the selected HYENAS sample. The vertical blue dashed lines mark the five halo mass bins, while the horizontal cyan lines indicate the $z_{\rm form}$ percentile bins. The inset images are the blended gas and star distributions around the central galaxy (highlighted by stars in lime colour) of the selected HYENAS samples using the Py-SPHViewer package.

sample using the same underlying white noise file. For each selected halo, we track all these dark matter particles within the halo at $z = 0$ to their initial condition positions. To precisely identify the centre of the resimulation region in the IC, we also track these particles lying in the centre (minimum potential positions) of the halo ($5\,h^{-1}$ kpc), and use the mean position of these central particles within the IC as the 'ref_center' parameter for MUSIC. Then, we calculate the distances from the IC central position to all the halo particle positions. The 2 times maximum distance rescaled to the simulation boxsize is used for the 'ref_extent' parameter for MUSIC to make sure the interested central halo is out of contamination.

Each zoomed-in halo has its ICs generated with three different levels of resolution: Level 0 cuts out the zoomed-in region in the IC, adds gas particles, and decreases the resolution outside of the zoomed-in region, which is controlled by the MUSIC code. This results in a dark matter particle mass of $6.513 \times 10^7\,h^{-1}\,{\rm M}_\odot$ and gas element mass of $1.241 \times 10^7\,h^{-1}\,{\rm M}_\odot$, and a minimum gravitational softening of $0.5\,h^{-1}$ kpc. Level 1 increases the zoomed-in region's resolution by 1 level higher with a new white noise at this resolution level. The white noise is consistent as Level 0 for these low-resolution levels decreased outside the zoomed-in region. The Level 1 suite has $8\times$ lower particle masses and $2\times$ lower minimum softening. Level

2 has a resolution that is higher than Level 1, with the outside low-resolution region being consistent with Level 1. Thus, each of the 120 selected HYENAS halo has 3 ICs with different resolutions. Though all the 120 HYENAS haloes have both dark-matter-only and hydro ICs, we only run the hydrodynamic for the 40 selected 'elite' halo for Level 1 and Level 2, which is still not fully finished due to their very high computation cost. The 'elite' sample is the one out of the three random samples with the smallest number of high-resolution particles.

2.2 The HYENAS catalogue and analysis

We output 151 snapshots for each simulation run from $z = 20$ to 0. Besides the on-fly FoF catalogue from GIZMO and the CAESAR catalogue based on it, we also run the AHF halo finder (Knollmann & Knebe 2009) to identify the haloes within R_{200c} and produce an AHF-Caesar catalogue following Chen et al. (2024).

Although there are other (uncontaminated) haloes inside the high-resolution zoomed-in regions, we only focus on these originally selected HYENAS haloes in this paper. It is also recommended for all papers using the HYENAS data. As such, we need to identify these selected haloes properly. Unfortunately, the particle IDs of these haloes chosen from the parent dark-matter-only simulation are scrambled within these zoom ICs generated by MUSIC. Thus, we adopted another method to cross-match these haloes based on the particle IDs of these zoom ICs. Since we know the exact halo centre position in these ICs, we use the IDs of dark matter particles within $300\,h^{-1}$ kpc (more-or-less corresponding to the $5\,h^{-1}$ kpc radius of haloes central regions at $z = 0$) and track them down to $z = 0$ to get their median positions. This position's distances to the $z = 0$ halo centres are used to select the closest halo. If the halo mass difference from the original N-body one is less than 1.5 times, the halo is matched. If not, we choose the halo within the distance of R200c with the closest mass as the matched halo. Only seldom does this happen for these late-formed haloes due to the slight evolution difference between dark matter and hydro runs. After finding the matched haloes for both FoF and AHF catalogues, we compared the matched halo masses, which basically agree with previous findings (for example Cui et al. 2012; Cui, Borgani & Murante 2014, minor effects due to the baryons). We further calculate the formation redshifts using both FoF and AHF halo catalogues, which are compared to the original halo formation redshift. There is a slightly larger scatter compared to the halo mass differences, primarily coming from the $z_{\rm form} \lesssim 0.3$ sample. We note here that there are only a few cases which show a clear difference on $z_{\rm form}$ and this doesn't affect our conclusions because all the quantities from the hydrosimulation are consistently used with no information from the dark-matter-only run required.

2.3 X-ray luminosities

X-ray properties are calculated using MOXHA package (Jennings & Davé 2023), which combines the yT-based PyXSIM (Biffi et al. 2012; Biffi, Dolag & Böhringer 2013; ZuHone et al. 2014) and CAESAR (Turk et al. 2011) software packages with the XSPEC spectral fitting package to provide an end-to-end pipeline for creating mock X-ray photon maps and analysing them to obtain mock observations such as X-ray luminosities, temperatures, and metallicities. First, we make a cut on the cold gas to remove ISM particles, which are artificially pressurized to resolve the Jeans mass (Davé et al. 2019). We use a cut such that only gas particles with a density $\rho < 0.1\,{\rm m_p\,cm^{-3}}$ and temperature $T > 2 \times 10^5$ K are included. Furthermore, we remove all wind particles and all particles with a non-zero star formation rate. We then use PyXSIM to generate X-ray emission fields in the source band of 0.5–2.0 keV, using a CIE APEC model (Smith et al. 2001) and using the SIMBA-tracked particle mass fractions for He, C, N, O, Ne, Mg, Si, S, Ca, Fe scaled to the Anders & Grevesse (1989) solar abundances table. The other elements are fixed at their solar abundance ratios. We finally sum the luminosity of the hot gas particles within a radius of R_{500} to give our value of $L_{X,0.5-2.0}$. For more details, we refer to Jennings et al. (2024).

3 RESULTS

3.1 Hot gas fraction

Since the X-ray is coming from hot gas, we first investigate the hot gas mass fraction in both the galaxy groups from both HYENAS (large stars) and SIMBA (small points) simulations and compare them to recent observations from Sun et al. (2009); Laganá et al. (2013); Lovisari, Reiprich & Schellenberger (2015); Akino et al. (2022) in Fig. 2. Using the X-ray images of 43 galaxy groups from *Chandra*, Sun et al. (2009) derived the enclosed gas fraction from the fitted gas density and temperature profiles. The galaxy group data from Laganá et al. (2013) comes from the *XMM–Newton* observation with the total halo mass based on the hydrostatic equilibrium (HE) assumption. Lovisari et al. (2015) selected their group sample (23) from NORAS and REFLEX catalogues, and used the *XMM–Newton* observation result to reconstruct these profiles. Again their gas mass was coming from the fitted gas density profile, and the halo mass is based on the HE assumption. The most recent work by Akino et al. (2022) also used an X-ray-selected XXL sample. However, their halo mass is based on the Hyper Suprime-Cam Subaru Strategic Program weak-lensing mass measurements, whereas only the fitting result with error bars is shown in Fig. 2. To our knowledge, these are the only data on galaxy group mass scale available from observation in the literature. This is simply due to the difficulties of observing them in X-rays, and as such, all the sample as primarily selected from the X-ray catalogue.

By comparing with the observation data, we see a decreasing trend for the $f_{\rm hot\ gas}$ with decreased halo mass for both simulation and Akino et al. (2022) fitting line. However, the data points from other observations show a large scatter without a clear trend. At $M_{500} \gtrsim 10^{13.5}\,{\rm M_\odot}$, the simulations agree better to the observation data points from Sun et al. (2009); Laganá et al. (2013) (see Liang et al. 2016, as well) than those of Akino et al. (2022). Note that the gas fraction in the massive halo seems to agree better with Akino et al. (2022) as shown in Cui et al. (2018), which could be due to the different AGN feedback strengths implemented. While in the intermediate halo mass range, the simulation data tends to agree with Akino et al. (2022) instead of others (see also Robson & Davé 2020). The disagreements between these observations' results at the galaxy group scale have been discussed in Eckert et al. (2021). Here, we would like to add an additional potential cause: Akino et al. (2022) and Eckert et al. (2016) (which also has a lower gas fraction compared to the others) are both based on the XXL survey (Pierre et al. 2016), which is a volume complete survey, while the others mostly preferentially selected the X-ray bright objects. At an even lower halo mass with $M_{500} \lesssim 10^{13}$, the simulation predicts a flattened trend.

It is worth noting that the HYENAS sample generally covers the distribution of SIMBA data points well at $M_{500} \gtrsim 10^{13}\,{\rm M_\odot}$. These outliers from SIMBA with $f_{\rm hot\ gas} \gtrsim 0.04$ at $M_{500} \lesssim 10^{13}\,{\rm M_\odot}$ could be due to the fact that they are close to a massive cluster, as discussed in Cui et al. (2022); in contrast, HYENAS groups are

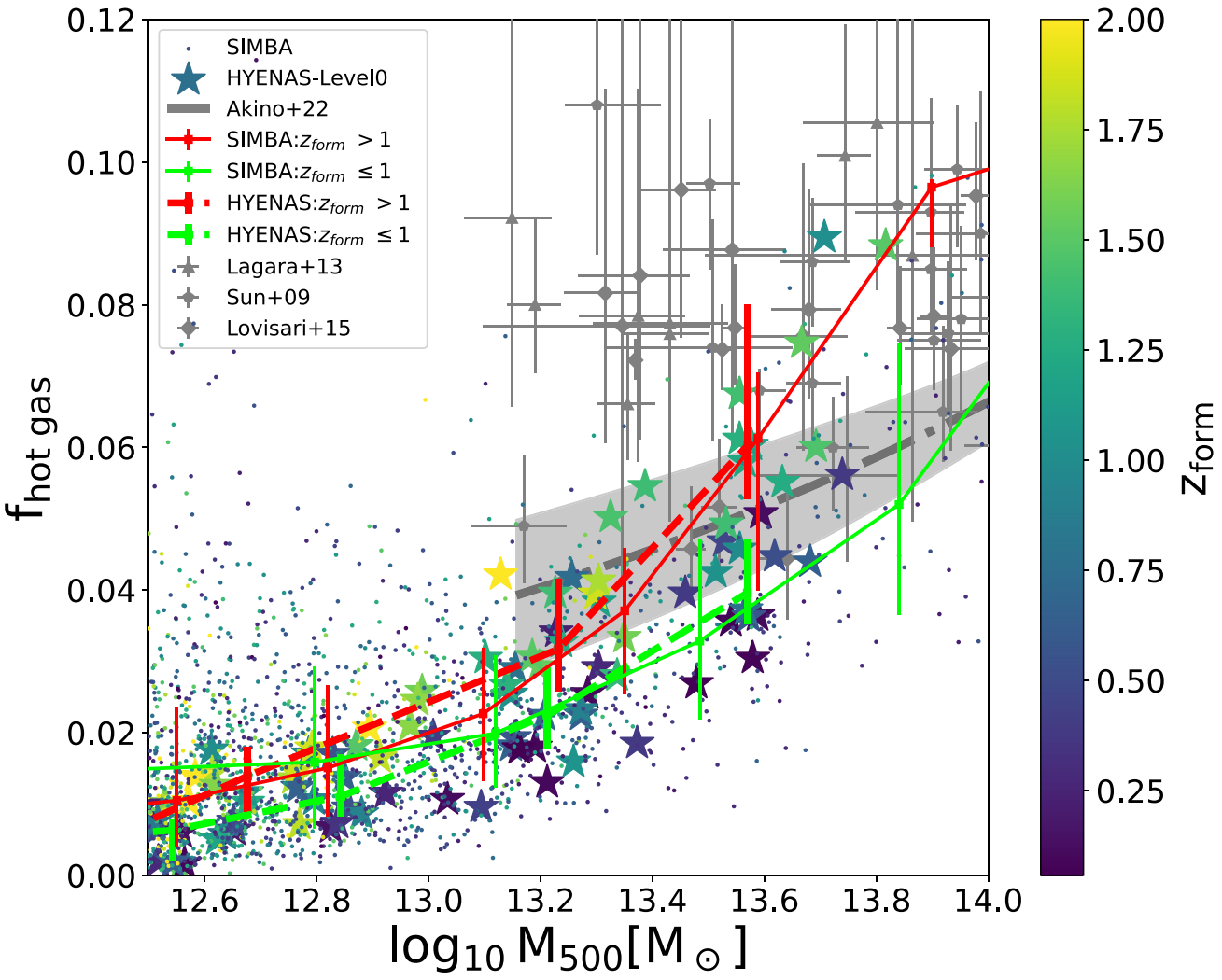

Figure 2. The hot gas fraction as a function of halo mass with colours coding to the halo formation time. The SIMBA simulation results are shown as small dots with the median values in solid thin red and cyan lines for early-formed ($z_{\rm form} > 1$) and late-formed groups ($z_{\rm form} \leq 1$), respectively. The error bars present the 16$^{\rm th}$–84$^{\rm th}$ percentile. While the HYENAS sample at Level 0 is indicated by stars with the medians in thick dashed lines. The observational results from Sun et al. 2009; Laganá et al. 2013; Lovisari et al. 2015; and Akino et al. 2022 are shown in grey symbols with error bars and a dotted-dashed line with a shadow region as indicated in the legend.

selected to be isolated haloes by design. Nevertheless, the agreement of overall distributions between SIMBA and HYENAS suggests that our HYENAS selection criteria are unbiased with respect to halo gas fractions.

Additionally, HYENAS points are colour coding to the halo formation redshift as shown by the colour bar in Fig. 2. This indicates an interesting trend – the gas fraction at a given halo mass correlates with its halo formation time. Late-formed haloes tend to have a smaller gas fraction. To statistically show this trend, we further separate the data points into $z_{\rm form} > 1$ (red lines) and $z_{\rm form} \leq 1$ (cyan lines). There is quite good agreement between the thick (HYENAS) and thin (SIMBA) lines, and both have a clear separation between early-formed and late-formed haloes.

The halo formation time dependence, we speculate, owes to accumulated heating processes from both AGN feedback and shock heating from structure formation. For early-formed haloes, it will not only have an early accretion of more cold gas at very high redshift (Cui et al. 2021), experience shock heating earlier and longer but also form its central galaxy earlier with a massive black hole according to the M_*–$M_\bullet$ relation, which leads to an earlier jet mode feedback in the SIMBA model.

At lower halo mass $M_{500} \lesssim 10^{13}\,{\rm M}_\odot$, the late-formed haloes tend to have a higher gas fraction than early-formed haloes, i.e. a reversed trend, albeit a large error bar for the SIMBA simulation. In comparison, the cross point is at a relatively lower halo mass, $M_{500} \approx 10^{12.5}\,{\rm M}_\odot$, for HYENAS, which could be due to the limited number of objects. At this lower halo mass, the jet mode AGN feedback will not turn on because the BH mass (see Cui et al. 2022, for the halo mass–BH mass relation) is lower than the threshold set in SIMBA which is around $10^8\,{\rm M}_\odot$. Thus, the heating process should be dominated by the supernovae feedback, which can be more dominant in these late-formed haloes with higher star formation. We will return to these explanations in the discussion part in detail.

3.2 X-ray luminosities

In Fig. 3, we show the HYENAS groups' X-ray luminosity within the soft band, [0.5, 2] keV, which has been adopted by many surveys

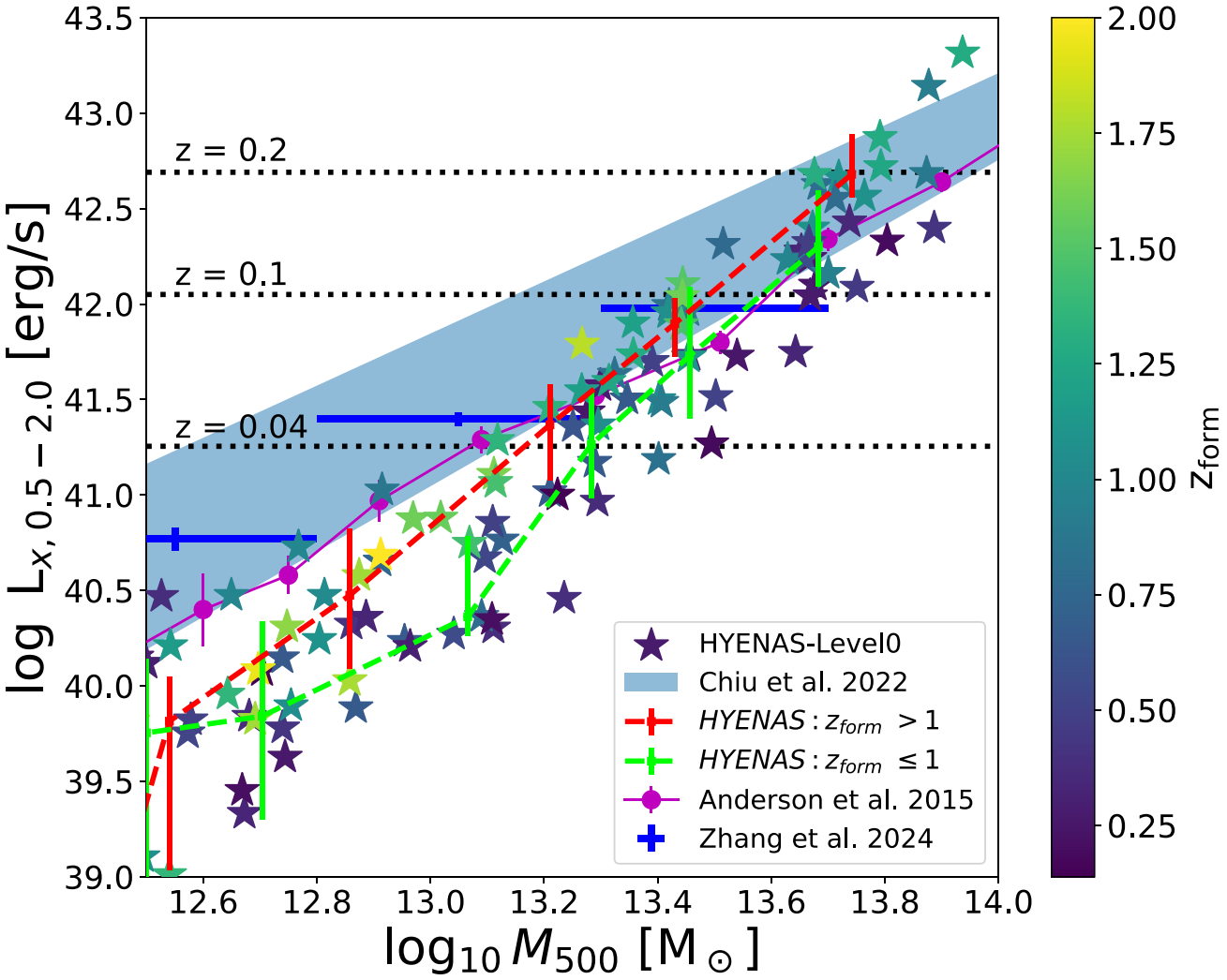

Figure 3. Similar to Fig. 2, but for the groups X-ray luminosity. Magenta lines and symbols show the result from Anderson et al. 2015 with the X-ray luminosity estimated at the same energy band – 0.5–2.0 keV. Blue error bars are the results from Zhang et al. 2024 using the latest eROSITA survey catalogue. We remind here again that the errorbars are different in simulation ($16^{\rm th}$–$84^{\rm th}$ percentiles) from observations.

for detecting X-ray galaxy groups. In comparison, we include the recent observation results from Anderson et al. (2015); Chiu et al. (2022); Zhang et al. (2024). The result from Anderson et al. (2015), shown in magenta line, is based on the stacked X-ray emission from the ROSAT All-Sky Survey around the local brightest galaxies with the halo mass computed using the simulated catalogue of these local brightest galaxies. Similarly, Zhang et al. (2024) (blue error bars) stack the X-ray luminosities of the CENhalo sample, which is binned in halo mass M_{200m} based on the group finder algorithm (Tinker 2021). Here, we used the corresponding M_{500c} values, which are derived from the concentration model (Ishiyama et al. 2021). Instead of stacking the galaxy groups, Chiu et al. (2022) (light blue region) used 434 groups and clusters, which are cross-confirmed via their weak lensing masses from the HSC survey. They removed these contaminated systems due to a random superposition and did an MCMC fitting, which yielded a similar result to Zhang et al. (2024) as shown in Fig. 3.

By separating the sample with their $z_{\rm form}$ using an arbitrarily selected threshold of $z = 1$, the late-formed haloes have a lower X-ray luminosity than early-formed ones, which is in agreement with the hot gas fraction shown in the previous section. It is worth noting that the differences between the two families at the same halo mass can be an order of magnitude. The magenta line (Anderson et al. 2015) is lying in between the HYENAS samples at $M_{500} \gtrsim 10^{13.2}\,{\rm M}_\odot$, roughly crossing the detection limit, and consistently lying on the top boundary of the HYENAS low mass sample. Note that the uncertainty on L_X in Anderson et al. (2015) is estimated from the quadratic sum of the Poisson error, which is not the same as what we are showing in HYENAS. While the other two observation results from Chiu et al. (2022); Zhang et al. (2024), show a slightly higher L_x than Anderson et al. (2015). As indicated by Bulbul et al. (2024), the detection limit of eRosita at $z \approx 0.2$ (the upper redshift limit in (Popesso et al. 2024)) is around 5×10^{42} ergs s^{-1}, which is indicated by the top horizontal dotted line in Fig. 3. If we naively use that as the detection limit of eROSITA at $z = 0.1$ (middle dotted horizontal line), it is clear that some of our simulated haloes, even with $M_{500} \sim 5 \times 10^{13}\,{\rm M}_\odot$ can not be seen. Therefore, these observation results in the galaxy group scale basically lie on the upper end of or higher than the HYENAS results, which could be explained if the observations are missing these X-ray faint groups.

Table 2. Predicted X-ray detection limits and fractions.

redshift z	min mass[1] log M_{500}	max mass[2] log M_{500}	mean mass[3] log M_{500}	fraction[4] 100 per cent
0.2	13.676	13.874	13.775	15.4
0.1	13.444	13.751	13.597	71.4
0.04	13.118	13.495	13.307	70.0

Note. [1]The minimum halo mass in $M_\odot$ above the X-ray luminosity limit at a given redshift, corresponds to 0 per cent detection if the halo mass is smaller than this one. [2]The maximum halo mass above the X-ray luminosity limit at a given redshift, corresponds to 100 per cent detection if the halo mass is larger than this one. [3]The mean of columns 2 and 3 for calculating the detection fraction in column 4. [4]The fraction of X-ray detected groups within the halo mass bin, column 3 ± 0.1.

Based on HYENAS data in Fig. 3, we further list the roughly predicted limits for eROSITA in Table 2, which are also shown by horizontal dotted lines in Fig. 3. At each of the three redshifts, we first select all HYENAS haloes within ± 10 per cent of the X-ray detection limit. The minimum and maximum halo masses within that L_x limit are listed in the second and third columns of Table 2, respectively. Using the mean (column 4) of columns 2 and 3, we select all haloes within a halo mass bin of ± 0.1, then give the detection fraction in the fifth column of Table 2 as $N_{L_x > L_{x,\rm lim}}/N_{\rm total\ in\ bin}$. We note here that the HYENAS sample is not a mass-complete one. Therefore, these limits and fractions only serve as a rough prediction. The clear drop of the fraction at $z = 0.2$ presents a good agreement to Popesso et al. (2024). However, we will need a larger sample to confirm this.

4 DISCUSSION AND CONCLUSION

As we have presented before, it is clear that the X-ray-detected galaxy groups are biased toward these gas-rich ones, which are closely linked to their early halo formation time. Although it is easy to understand that – early halo formation will bring more cold gas at high redshift, as shown in Cui et al. (2021), those gases will be heated up by either shock heating in structure formation or feedback in the process of galaxy formation. it is unclear how this fits into the picture of general expectations from galaxy formation. Therefore, we break this down into three aspects for elaboration:

(i) **[Connection to the central galaxy]** Observations have suggested that early-type elliptical galaxies in galaxy groups tend to be associated with diffuse X-ray emission, while late-type disc galaxies do not (Mulchaey et al. 2003). This is especially interesting because Cui et al. (2021) studied the connection between the central galaxy stellar mass and the halo mass and revealed the scatter in that relation is intrinsically driven by halo formation time. In their Supplementary Fig. 2, it is clear that early-formed haloes tend to host massive quenched galaxies at the same halo mass when $M_{\rm halo} > 10^{13}\,{\rm M}_\odot$, reversed from the low-mass haloes in which the red/quenched galaxies with lower stellar mass tend to live in late-formed haloes. This is further confirmed by the galaxy age (see Fig. D1) if we simply interpret those as the early-type galaxies tend to form earlier with older age. However, we understand that this issue is still in debate; see Scholz-Díaz et al. (2024) for the most recent discussions on that. We argue here that our results (including Cui et al. (2021)) are consistent with their claim that higher stellar mass galaxies at a given halo mass have characteristics of old, red, and passive systems at halo mass larger than$10^{13}\,{\rm M}_\odot$. While there is less data in Scholz-Díaz et al. (2024) at low-halo/stellar mass range to make a solid conclusion, and their total mass is only calculated within $3R_e$.

Due to being driven by halo formation time in both relations, we expect a positive connection between the gas fraction and central galaxy stellar mass at group halo mass scales – more massive galaxies tend to be surrounded by more hot gas. This is proved in Fig. B1, see more discussions regarding the hot gas fraction in the next item. As suggested by Correa & Schaye (2020), disc galaxies are less massive than elliptical galaxies in same-mass haloes when the halo mass is larger than $10^{13}\,{\rm M}_\odot$, which confirms the previous suggestion that elliptical galaxies tend to associate with X-ray emissions.

We further find a positive correlation between the central galaxy mass-weighted age and halo formation time, as shown in Fig. D1. This again confirms our results at $M_{500} \gtrsim 10^{13}\,{\rm M}_\odot$ are consistent with Scholz-Díaz et al. (2024) and lead to a positive correlation with the scatter in central galaxy stellar mass. This is contrary to the findings of Kulier et al. (2019), which could be because they used all galaxies within the EAGLE simulation, so low-mass haloes dominate the sample. At low halo masses $M_{\rm halo} \lesssim 10^{13}\,{\rm M}_\odot$, such an anticorrelation is also found in Cui et al. (2021). We also note that a crossing point is shown in Fig. 2, which should be consistent with the reversed trend at low-halo mass, though at slightly different halo masses when comparing SIMBA and HYENAS.

(ii) **[The abundance of cold gas]** In previous figures, we only focus on the hot gas mass fraction; it is unclear how the cold gas abundance will contribute to the full picture, i.e. whether the low hot gas fraction is due to a high cold gas fraction or not. This is because galaxy groups, unlike galaxy clusters, tend to host a noticeable fraction of gas mass in cold as well. Investigating that will help us to form a full picture of how they are formed. As shown in Fig. A1, the cold gas gradually contributes more (from about 5 to 20–40 per cent) to the total gas mass for the halo with $M_{500} \approx 10^{13.2}$–$10^{12.5}\,{\rm M}_\odot$. It is further interesting to see that there is more cold gas in late-formed haloes than in early-formed ones in that figure, which we will discuss the reasons for in the following section. That reveals that the history of halo formation also affects the history of gas thermalization (e.g. Sereno et al. 2021). However, we do not think this small change is the main reason for the $f_{\rm hot\ gas}$ difference between early- and late-formed haloes, especially for massive galaxy groups.

(iii) **[Connection to central BH]** Although there is more gas in early-formed haloes, the gas must be hot to be seen in the X-ray band. Therefore, the heating processes are key to understanding why there is more hot gas in these early-formed haloes than in late-formed ones. As shown in the previous section, the early-formed haloes not only have more gas but also more hot gas than these late-formed haloes (see Fig. A1). Thus, early-formed haloes should have more heating sources/energies than late-formed ones. One possible reason is shock heating, which should happen earlier in early-formed haloes, yielding a hot gas fraction. The other reason is AGN feedback. For example, Liang et al. (2016) suggested that the winds ejected from the group galaxies interact with and heat the hot halo gas, which not only reduces the rate at which the halo gas cools and accumulates in the group's central galaxies but also causes its distribution to remain more extended. More importantly, from Fig. C1, we found that the massive galaxies tend to host a more massive BH at the same halo mass (see Davies et al. 2019, 2020, at the more massive halo mass end; see also Ma et al. in prep.). This is not surprising since the early-formed haloes tend to form the central galaxy earlier, and as such, the central BH mass is seeded earlier and grows faster. For the case of SIMBA model, it enters the jet mode earlier to quench the central galaxy (Cui et al. 2021) with the higher hot gas mass as a by-product. This picture is supported by Fig. C1, which shows

that early-formed haloes tend to host massive BHs. We are currently working on another paper to record the heating energy from different sources to determine which is more important for gas heating in galaxy groups.

Our findings in this work are based only on the SIMBA baryon model. However, we also investigated the TNG-300 simulation, which shows the same gas fraction trend with a clear separation between early- and late-formed haloes, albeit with a little systematically higher values than what is shown in Fig. 2 (see also Davies et al. 2020, for the higher gas fraction in TNG than EAGLE). Recent observation work by Popesso et al. (2024), which compared the X-ray detected and undetected groups, also suggested a similar conclusion, i.e. halo assembly bias is the cause. Furthermore, Andreon, Trinchieri & Moretti (2022) showed that under X-ray luminous clusters populate the low concentration of dark matter end of the distribution for a given mass, suggesting that they are late-formed as well. However, the halo formation redshift is very hard to measure in observations. There are ways to approximate it, such as the galaxy magnitude/stellar mass difference in fossil groups (e.g. Jones et al. 2003; Gozaliasl et al. 2014) and the connection between galaxy/gas dynamical state and halo formation time (e.g. Mostoghiu et al. 2019), but all have a substantial uncertainty. Though there are claims that the fossil groups show no difference to normal groups in X-ray scaling relations (e.g. Girardi et al. 2014; Kundert et al. 2015), their lowest X-ray is still above $\sim 10^{42}$, which is much higher compared to the limit shown in this study. In this theoretical investigation, we don't probe into details but suggest these connections to galaxy and BH properties can be tested in observations as discussed in previous paragraphs.

Another possible explanation for these X-ray faint or undetected galaxy groups is the projection effect when they are generally identified through the galaxy catalogues (see Hernquist, Katz & Weinberg 1995, for example). If two small haloes are lying along the same line of sight but have a large separation, neither will have large enough hot gas to shine in X-ray, see Fig. 3 for how quickly the X-ray luminosity drops with halo mass. However, this projection issue may be solved by highly accurate spectroscopic redshift measurement with proper galaxy velocity distribution modelling.

Lastly, the role that baryonic physics models play in this result is still not very clear, especially the AGN feedback, which may affect the X-ray luminosity. For example, Kar Chowdhury et al. (2022) showed that the different versions of SIMBA run turning on and off different SIMBA' models, especially the X-ray AGN feedback and radiative mode, result in different surface brightness profiles at different radii. However, we argue that this will only systematically shift our result, while the effect of halo formation time on the X-ray luminosity will be unchanged, which we have confirmed with the TNG-300 result. On the other hand, many studies with the EAGLE simulation show consistent predictions, as we have discussed before.

ACKNOWLEDGEMENTS

We thank the anonymous referee for his/her valuable comments.

WC is supported by the STFC AGP Grant ST/V000594/1, the Atracción de Talento Contract no. 2020-T1/TIC-19882 was granted by the Comunidad de Madrid in Spain, and the science research grants were from the China Manned Space Project. He also thanks the Ministerio de Ciencia e Innovación (Spain) for financial support under Project grant PID2021-122603NB-C21 and HORIZON EUROPE Marie Sklodowska-Curie Actions for supporting the LACEGAL-III project with grant number 101086388.

FJ would like to acknowledge the support of the Science and Technology Facilities Council (STFC).

AB acknowledges support from the Natural Sciences and Engineering Research Council of Canada (NSERC) through its Discovery Grant program, the Infosys Foundation, via an endowed Infosys Visiting Chair Professorship at the Indian Institute of Science and the Leverhulme Trust.

These HYENAS simulations were performed using the DiRAC@Durham facility managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk) with the DiRAC Project: ACSP252, titled Simba Zoom Simulations of Galaxy Groups. The equipment was funded by BEIS capital funding via STFC capital grants ST/P002293/1, ST/R002371/1, and ST/S002502/1, Durham University, and STFC operations grant ST/R000832/1. DiRAC is part of the National e-Infrastructure. The analysis reported in this paper was also partly enabled on the Niagara supercomputer by SciNet, WestGrid and the Digital Research Alliance of Canada (alliancecan.ca). SciNet (Locken et al. 2010) is funded by Innovation, Science and Economic Development Canada; the Digital Research Alliance of Canada; the Ontario Research Fund: Research Excellence; and the University of Toronto.

The paper was written with the help of the search and cite extension for Overleaf, which is publicly available at: https://chromewebstore.google.com/detail/search-and-cite/opadkkneiklbpkpdglojbcgdfaanmkoh.

DATA AVAILABILITY

The data output of this paper is available on GitHub: The HYENAS simulation is currently available upon request, but it will become publicly available very soon.

REFERENCES

Akino D. et al., 2022, PASJ, 74, 175
Anders E., Grevesse N., 1989, Geochim. Cosmochim. Acta, 53, 197
Anderson M. E., Gaspari M., White S. D. M., Wang W., Dai X., 2015, MNRAS, 449, 3806
Andreon S., Trinchieri G., Moretti A., 2022, MNRAS, 511, 4991
Bahar Y. E. et al., 2024, preprint (arXiv:2401.17276)
Biffi V., Dolag K., Böhringer H., 2013, MNRAS, 428, 1395
Biffi V., Dolag K., Böhringer H., Lemson G., 2012, MNRAS, 420, 3545
Bulbul E. et al., 2024, A&A, 685, A106
Chen M., Cui W., Fang W., Wen Z., 2024, ApJ, 966, 227
Chiu I. N. et al., 2022, A&A, 661, A11
Correa C. A., Schaye J., 2020, MNRAS, 499, 3578
Cui W. et al., 2018, MNRAS, 480, 2898
Cui W. et al., 2022, MNRAS, 514, 977
Cui W., 2024, preprint (arXiv:2406.03829)
Cui W., Borgani S., Dolag K., Murante G., Tornatore L., 2012, MNRAS, 423, 2279
Cui W., Borgani S., Murante G., 2014, MNRAS, 441, 1769
Cui W., Davé R., Peacock J. A., Anglés-Alcázar D., Yang X., 2021, Nat. Astron., 5, 1069
Damsted S. et al., 2023, A&A, 676, A127
Damsted S. et al., 2024, preprint (arXiv:2403.17055)
Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, MNRAS, 486, 2827
Davies J. J., Crain R. A., McCarthy I. G., Oppenheimer B. D., Schaye J., Schaller M., McAlpine S., 2019, MNRAS, 485, 3783

Davies J. J., Crain R. A., Oppenheimer B. D., Schaye J., 2020, MNRAS, 491, 4462
Eckert D. et al., 2016, A&A, 592, A12
Eckert D., Gaspari M., Gastaldelo F., Brun A. M. C. L., O'Sullivan E., 2021, Universe, 7, 142
Eckert D., Gastaldello F., O'Sullivan E., Finoguenov A., Brienza M., the X-GAP collaboration, 2024, Galaxies, 12, 24
George M. R. et al., 2011, ApJ, 742, 125
Girardi M. et al., 2014, A&A, 565, A115
Gozaliasl G. et al., 2019, MNRAS, 483, 3545
Gozaliasl G. et al., 2020, A&A, 635, A36
Gozaliasl G., Khosroshahi H. G., Dariush A. A., Finoguenov A., Jassur D. M. Z., Molaeinezhad A., 2014, A&A, 571, A49
Hahn O., Abel T., 2011, MNRAS, 415, 2101
Hernquist L., Katz N., Weinberg D. H., 1995, ApJ, 442, 57
Ishiyama T. et al., 2021, MNRAS, 506, 4210
Jennings F. J., Babul A., Dave R., Cui W., Rennehan D., 2024, preprint (arXiv:2407.14415)
Jennings F., Davé R., 2023, MNRAS, 526, 1367
Jones L. R., Ponman T. J., Horton A., Babul A., Ebeling H., Burke D. J., 2003, MNRAS, 343, 627
Jung S. L. et al., 2022, MNRAS, 515, 22
Kar Chowdhury R., Chatterjee S., Paul A., Sarazin C. L., Dai J. L., 2022, ApJ, 940, 47
Khalil H., Finoguenov A., Tempel E., Mamon G. A., 2024, preprint (arXiv:2403.17061)
Knollmann S. R., Knebe A., 2009, ApJS, 182, 608
Kulier A., Padilla N., Schaye J., Crain R. A., Schaller M., Bower R. G., Theuns T., Paillas E., 2019, MNRAS, 482, 3261
Kundert A. et al., 2015, MNRAS, 454, 161
Laganá T. F., Martinet N., Durret F., Lima Neto G. B., Maughan B., Zhang Y. Y., 2013, A&A, 555, A66
Li Q. et al., 2022, ApJ, 933, 9
Liang L., Durier F., Babul A., Davé R., Oppenheimer B. D., Katz N., Fardal M., Quinn T., 2016, MNRAS, 456, 4266
Loken C., et al., 2010, JPhCS, 256, 012026
Loubser S. I., Hoekstra H., Babul A., O'Sullivan E., 2018, MNRAS, 477, 335
Lovisari L., Ettori S., Gaspari M., Giles P. A., 2021, Universe, 7, 139
Lovisari L., Reiprich T. H., Schellenberger G., 2015, A&A, 573, A118
Manolopoulou M., Hoyle B., Mann R. G., Sahlén M., Nadathur S., 2021, MNRAS, 500, 1953
Mostoghiu R., Knebe A., Cui W., Pearce F. R., Yepes G., Power C., Dave R., Arth A., 2019, MNRAS, 483, 3390
Mulchaey J. S., 2000, ARA&A, 38, 289
Mulchaey J. S., Davis D. S., Mushotzky R. F., Burstein D., 2003, ApJS, 145, 39
O'Sullivan E. et al., 2017, MNRAS, 472, 1482
Oppenheimer B. D., Babul A., Bahé Y., Butsky I. S., McCarthy I. G., 2021, Universe, 7, 209
Pearson R. J. et al., 2017, MNRAS, 469, 3489
Pierre M. et al., 2016, A&A, 592, A1
Popesso P. et al., 2024, MNRAS, 527, 895
Robson D., Davé R., 2020, MNRAS, 498, 3061
Scholz-Díaz L., Martín-Navarro I., Falcón-Barroso J., Lyubenova M., van de Ven G., 2024, Nat. Astron., 8, 648
Sereno M., Lovisari L., Cui W., Schellenberger G., 2021, MNRAS, 507, 5214
Smith R. K., Brickhouse N. S., Liedahl D. A., Raymond J. C., 2001, ApJ, 556, L91
Springel V., Pakmor R., Zier O., Reinecke M., 2021, MNRAS, 506, 2871
Sun M., Voit G. M., Donahue M., Jones C., Forman W., Vikhlinin A., 2009, ApJ, 693, 1142
Tinker J. L., 2021, ApJ, 923, 154
Turk M. J., Smith B. D., Oishi J. S., Skory S., Skillman S. W., Abel T., Norman M. L., 2011, ApJS, 192, 9
Yang T., Cai Y.-C., Cui W., Davé R., Peacock J. A., Sorini D., 2022, MNRAS, 516, 4084
Yang T., Davé R., Cui W., Cai Y.-C., Peacock J. A., Sorini D., 2024, MNRAS, 527, 1612
Zhang Y. et al., 2024, preprint (arXiv:2401.17309)
ZuHone J. A., Biffi V., Hallman E. J., Randall S. W., Foster A. R., Schmid C., 2014, preprint (arXiv:1407.1783)

APPENDIX A: HOT GAS ABUNDANCE

As the halo mass decreases, gas heating becomes weak for various reasons. As such, galaxy groups, unlike clusters, may contain a certain fraction of cold gas, which doesn't emit X-ray photons. As such, it would be interesting to understand the cold gas content in galaxy groups. In Fig. A1, we show the hot gas mass fraction with respect to the total gas mass as a function of halo mass. Furthermore, the simulation data is coloured and split by their halo formation time. Although hot gas still occupies the most mass in galaxy groups, the decreasing fraction is clear along halo mass, especially with $M_{500} \lesssim 10^{13.2}\,\mathrm{M}_\odot$. By separating the haloes into early- and late-formed families, we found that the late-formed haloes have systematically more cold gas than early-formed haloes. This becomes much clearer at the lower halo mass end.

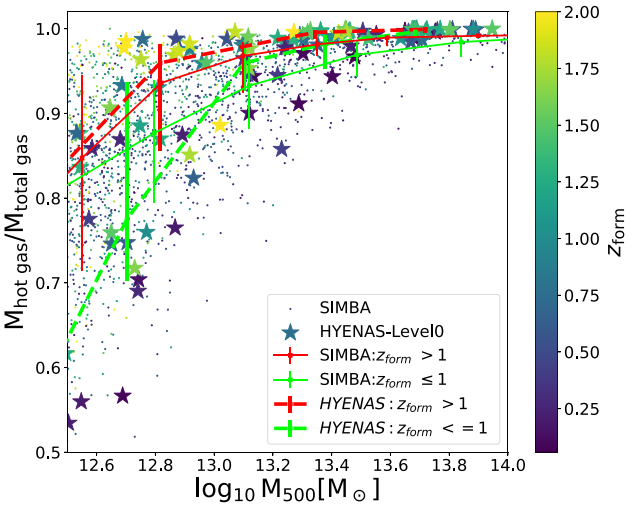

Figure A1. The hot gas mass fraction with respect to the total gas mass. The symbols and lines share the same meanings as Fig. 2. The hot gas dominates ($\gtrsim$0.95) the total gas mass in haloes with $M_{500} \gtrsim 10^{13.4}\,\mathrm{M}_\odot$. Cold gas starts to contribute more mass as the halo mass drops. It is also interesting to see that the hot gas fraction is higher in these early-formed haloes than in later-formed ones at the same halo mass, which is clearer at the lower halo mass.

APPENDIX B: GAS FRACTION SEPARATED BY CENTRAL GALAXY FRACTION

Instead of halo formation time, which directly affects the central galaxy properties (e.g. Cui et al. 2021), we investigate the central galaxy (or brightest group galaxy in observation, BGG) stellar mass fraction's influence on the hot gas fraction in Fig. B1. The symbol colours of simulated objects are to the BGG's stellar mass fraction – M_*/M_{500}. To clearly show its effects, we first estimate the median line in the M_{500}–M_*/M_{500} relation. Then, separate these objects in Fig. B1 into two groups: above the median line or below the median line in the M_{500}–M_*/M_{500} relation. After that, we show the median values of the two groups in red and blue, respectively. As shown in Fig. B1, the high gas fraction is generally associated with a massive BGG for haloes with $M_{500} \gtrsim 10^{13}\,\mathrm{M}_\odot$.

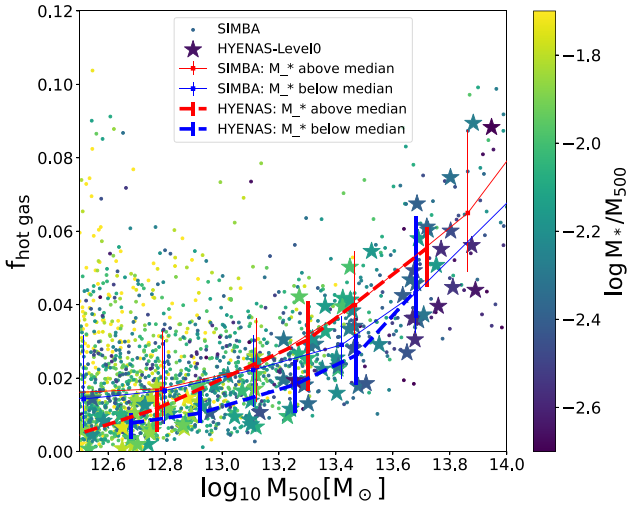

Figure B1. Similar to Fig. 2 but colour coding to the central galaxy mass fraction. Due to the stellar mass fraction depending on halo mass, we used its median value in the $M_*/M_{\rm halo}$–$M_{\rm halo}$ to separate the two families. At M_{500} above $10^{13}\,{\rm M}_\odot$, More massive central galaxies tend to have higher hot gas fractions than less massive galaxies at the same halo mass.

APPENDIX C: GAS FRACTION SEPARATED BY CENTRAL BH MASS

Similar to Fig. B1, we highlight the effect of black hole mass, $M_\bullet$,

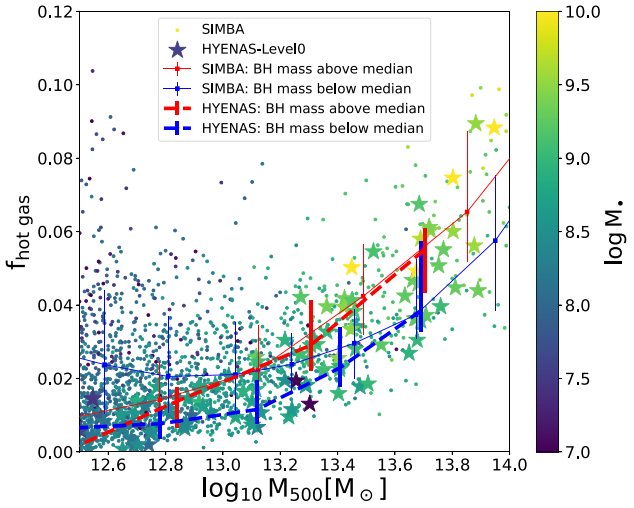

Figure C1. Similar to Fig. B1 but colour coding to the BH mass in the central galaxy. Again, we use the median line in the $M_\bullet$–$M_{\rm halo}$ relation to split the massive and low mass BH families. A more massive central BH tends to have a higher gas fraction at the same halo mass range when $M_{500} \gtrsim 10^{13}\,{\rm M}_\odot$.

in Fig. C1. It is not surprising to see the halo with a higher black hole mass tends to have more hot gas. This is consistent with Fig. B1 because we know that the black hole mass is primarily scaling with its host galaxy's stellar mass. On the other hand, this hints the AGN feedback may play a role in the higher hot gas fraction, because the higher BH mass, the longer AGN feedback time.

APPENDIX D: GAS FRACTION SEPARATED BY CENTRAL GALAXY AGE

Instead of BGG's stellar mass fraction, we show the connection to BGG's mass-weighted stellar age in Fig. D1. Again, this fits into the consistent picture of BGGs formed earlier with older age to have more stellar mass in the early-formed haloes. To summarize, for haloes with $M_{500} \gtrsim 10^{13}\,{\rm M}_\odot$, the early-formed halo initializes its star formation earlier, with an early-formed central galaxy, SIMBA seeds the BH at a higher redshift, which is based on galaxy stellar mass. As such, both the central galaxy and BH have a longer time to grow, resulting in a higher mass than in late-formed haloes. If the early- and late-formed haloes have the same mass, we expect the energy for heating gas from gravity should be similar. Therefore, we suspect the feedback from baryon models is the key responsible for the difference in gas heating.

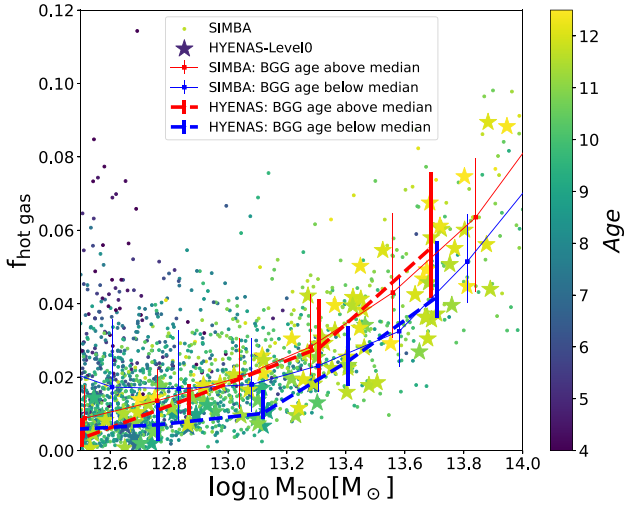

Figure D1. Similar to Fig. B1 but colour coding to the stellar age of the central galaxy. Again, we use the median line in the age–$M_{\rm halo}$ relation to split the sample into two galaxy age families. At the same halo mass range with $M_{500} \gtrsim 10^{13}\,{\rm M}_\odot$, an older galaxy tends to have a higher gas fraction.

This paper has been typeset from a TEX/LATEX file prepared by the author.