
Università degli Studi di Padova
Dipartimento di Astronomia
SCUOLA DI DOTTORATO DI RICERCA IN ASTRONOMIA
CICLO XXII

# Clustering Analysis of Cosmic Sources at High Redshifts

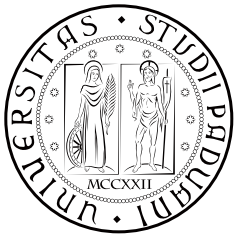
*Direttore della Scuola:* Ch.mo Prof. Gianpaolo Piotto

*Supervisore:* Ch.mo Prof. Alberto Franceschini

*Co-Supervisori:* Dr. Mattia Vaccari
Dr. Alexey Vikhlinin

*Dottorando:* Svetlana Starikova

Padova, 2011



Università degli Studi di Padova
Dipartimento di Astronomia
SCUOLA DI DOTTORATO DI RICERCA IN ASTRONOMIA
CICLO XXII

# Clustering Analysis of Cosmic Sources at High Redshifts

*Direttore della Scuola:* Ch.mo Prof. Gianpaolo Piotto

*Supervisore:* Ch.mo Prof. Alberto Franceschini

*Co-Supervisori:* Dr. Mattia Vaccari
Dr. Alexey Vikhlinin

*Dottorando:* Svetlana Starikova

Padova, 2011

# Riassunto

La tesi e' dedicata allo studio delle proprieta' di clustering di galassie normali e attive, in un ampio intervallo di redshift, dall'Universo locale fino a $z > 2$. Queste sorgenti, selezionate nellinfrarosso con *Spitzer* e nei raggi X con *Chandra*, rappresentano diverse classi di oggetti ad alto redshift — AGN a $z > 1$ che emettono nell'X, massicce galassie evolute a $z \sim 2$ e galassie con formazione stellare a $z \simeq 1.7$ e 0.7. Calcoliamo alcune misure di funzioni di correlazione angolare per questi oggetti e le combiniamo con dei modelli di clustering per gli aloni di materia oscura allo scopo di inferire le proprieta' di occupazione degli aloni per le diverse popolazioni di galassie.

Inizialmente si sono discusse le proprieta' di *clustering* delle galassie con indicazioni di attivita' di formazione stellare, rilevate a 24 $\mu$m da *Spitzer*. Usando i colori nell'ottico/vicino-IR, abbiamo suddiviso il campione in una popolazione a basso-redshift e in ($z_{\rm mean} = 0.68$) una popolazione ad alto-redshift ($z_{\rm mean} = 1.72$), e abbiamo misurato una lunghezza di correlazione comovente di $r_0 = 4.74 \pm 0.16\, h^{-1}$ Mpc e $r_0 = 7.87 \pm 0.63\, h^{-1}$ Mpc, rispettivamente. Da questo abbiamo potuto derivare le masse degli aloni di materia oscura $M_{\rm tot} \gtrsim 2 \times 10^{12}\, h^{-1}\, M_\odot$ per la popolazione ad alto redshift, da confrontare con quella stimata per la popolazione a basso redshift $M_{\rm tot} \gtrsim 5 \times 10^{11}\, h^{-1}\, M_\odot$ — in linea con le stime delle luminosita' IR $L_{\rm IR} > 10^{12} L_\odot$ ("ULIRG") e $L_{\rm IR} \sim 10^{11} L_\odot$ ("LIRG").

Si e' quindi presentata l'analisi di galassie massicce (massa stellare stimata di $M^* \approx 1.6 \times 10^{11}\, M_\odot$), ad alto redshift ($z_{\rm mean} = 2.2$) preliminarmente selezionate dalla presenza del picco di emissione stellare a 1.6 $\mu$m, che si trova spostato alla lunghezza d'onda osservata a nella 5.8 $\mu$m banda *Spitzer* (i cosiddetti "IR-peakers"). Si e' trovato che le proprieta' osservate degli IR-peakers - la loro massa stellare, densita' numerica e proprieta' di clustering - sono spiegate al meglio se si assume che essi risiedano in quel $10-25\%$ di aloni di materia oscura con $M_{\rm tot} \approx [0.5-1] \times 10^{13} M_\odot$, in cui l'efficienza di conversione stellare raggiunge

il $\sim 10-20\%$. Si e' inoltre dimostrato che gli IR-peakers non possono trovarsi solamente nelle galassie centrali dei corrispettivi aloni di materia oscura, o solo nei sottoaloni.

Infine, si e' misurato il clustering spaziale degli AGN rilevati dall'osservatorio spaziale a raggi X *Chandra* nel campo Boötes, in un intervallo di redshift da $z = 0.17$ a $z \sim 3$. Le lunghezze di correlazione derivate non indicano nessun andamento con il redshift, allinterno di questo campione. La disponibilita' di accurati redshift spettroscopivi ci ha permesso di utilizzare le funzioni di correlazione a due punti proiettate sul piano del cielo, e lungo la linea di vista, per mostare che gli AGN selezionati nei raggi X sono principalmente collocati nei centri degli aloni di materia oscura con $M_{\rm tot} > 3.7 \times 10^{12}\, h^{-1}\, M_{\odot}$, e tendono ad evitare galassie satelliti in aloni di masse di quest'ordine di grandezza, o maggiori. Le proprieta' di occupazione derivate dai dati di clustering degli AGN *Chandra* - la scala delle masse dei corrispettivi aloni di materia oscura, la mancanza di una significativa evoluzione con il redshift della lunghezza di correlazione del clustering, e la bassa frazione di satelliti - sono in linea di massima consistenti con lo scenario dell'attivita' di quasar attivata dai *mergers* di galassie di massa simile.

Di particolare interesse, si e' trovato che le tre classi di oggetti studiate in questa tesi hanno una lunghezza di correlazione simile ($r_0 \approx 7-8\, h^{-1}\,$Mpc), e quindi risiedono in aloni di materia oscura di massa comparabile, alcuni $\times 10^{12}\, h^{-1}\, M_{\odot}$. Tuttavia, essi rappresentano delle popolazioni distinte e largamente non sovrapposte. Le galassie star-forming si trovano nel $\simeq 20\%$ degli aloni di questa massa, probabilmente quelli che hanno subito un innescamento della formazione stellare nel passato recente. Le galassie normali molto massicce ("IR-peakers") risiedono nel $10-20\%$ degli aloni che hanno avuto un aumento dellefficienza della formazione stellare a redshift molto alti. Gli AGN selezionati nei raggi X occupano una piccola frazione ($\sim 1-5\%$) degli aloni, e si e' trovata evidenza di una differenza nelle proprieta' di clustering di AGN ed IR-peakers: mentre questi ultimi includono sia galassie centrali massicce e galassie satellite, gli AGN sono preferenzialmente situati nelle galassie centrali. Considerati insieme, gli oggetti studiati in questa tesi popolano il $\sim 30-50\%$ degli aloni di materia oscura piu' massicci fra $z = 1$ e $z = 2$.

# Abstract

This thesis is dedicated to the investigation of clustering properties of both normal and active galaxies over a wide range of redshifts, from local to $z > 2$. These sources, selected in the infrared with *Spitzer* and X-rays with *Chandra*, represent different classes of objects in the high-$z$ Universe — X-ray emitting AGNs at $z \gtrsim 1$, massive evolved galaxies at $z \sim 2$, and star-forming galaxies at $z \simeq 1.7$ and 0.7. We report on the measurements of the correlation functions for these objects and combine them with the clustering models of dark matter halos in order to infer the halo occupation properties of the various galaxy populations.

We first discuss the clustering of galaxies with signs of active star formation detected at 24 $\mu$m by *Spitzer*. Using optical/near-IR colors, we separate the sample into a lower-redshift ($z_{\rm mean} = 0.68$) and higher-redshift ($z_{\rm mean} = 1.72$) galaxy populations, and measure comoving correlation lengths of $r_0 = 4.74 \pm 0.16\,h^{-1}$ Mpc and $r_0 = 7.87 \pm 0.63\,h^{-1}$ Mpc, respectively. From this we derive the masses of parent dark matter halos $M_{\rm tot} \gtrsim 2 \times 10^{12}\,h^{-1}\,M_\odot$ for the high-$z$ population, compared to $M_{\rm tot} \gtrsim 5 \times 10^{11}\,h^{-1}\,M_\odot$ at the lower redshifts — in line with the estimated IR luminosities $L_{\rm IR} > 10^{12} L_\odot$ (corresponding to "ULIRGs") and $L_{\rm IR} \sim 10^{11} L_\odot$ ("LIRGs").

We then present the analysis of massive (estimated stellar mass $M^* \approx 1.6 \times 10^{11}\,M_\odot$), high-redshift ($z_{\rm mean} = 2.2$) galaxies previously selected by the presence of the 1.6 $\mu$m stellar peak redshifted into the 5.8 $\mu$m *Spitzer* band ("IR-peakers"). We find that the key observable characteristics of IR-peakers — their stellar mass, number density, and clustering, are best explained if they reside in those $10 - 25\%$ of dark matter halos with $M_{\rm tot} \approx [0.35 - 0.7] \times 10^{13}\,M_\odot$, in which the star conversion efficiency reaches $\sim 10 - 20\%$. We also show that IR-peakers can not be located only in the central galaxies of parent dark matter halos, or only in the subhalos.

Finally, we measure the spatial clustering of AGNs detected by the *Chandra* X-ray Observatory in the Boötes field over a redshift interval from $z = 0.17$ to $z \sim 3$. The derived correlation lengths are consistent with no redshift trend within the sample. The availability of accurate spectroscopic redshifts allows us to use the two-point correlation functions projected on the sky plane and in the line of sight to show that the X-ray AGNs are predominantly located at the centers of dark matter halos with $M_{\rm tot} > 3.7 \times 10^{12}\,h^{-1}\,M_{\odot}$, and tend to *avoid* satellite galaxies in halos of this or higher mass. The halo occupation properties inferred from the clustering data of *Chandra* AGNs — the mass scale of the parent dark matter halos, the lack of significant redshift evolution of the clustering length, and the low satellite fraction — are broadly consistent with the scenario of quasar activity triggered by mergers of similarly-sized galaxies.

Interestingly, the three classes of objects studied in this thesis have similar correlation lengths ($r_0 \approx 7 - 8\,h^{-1}$ Mpc), and hence reside in dark matter halos of similar mass, a few $\times 10^{12}\,h^{-1}\,M_{\odot}$. However, they represent distinct and mostly non-overlapping populations. The star-forming galaxies are found in $\simeq 20\%$ of halos of this mass, presumably those which experienced a star formation trigger in the recent past. The very massive normal galaxies ("IR-peakers") reside in $10 - 20\%$ of the halos which had an enhanced star formation efficiency at very high redshifts. The X-ray selected AGNs occupy a small fraction $\sim 1 - 5\%$ of the halos, and we find an evidence for a difference in the clustering properties of AGNs and IR-peakers — while the latter include both massive central galaxies and satellite galaxies, the AGNs are preferentially located in the central galaxies. Taken together, the objects studied in this thesis populate $\sim 30 - 50\%$ of the massive dark matter halos at $z = 1 - 2$.

# Ringraziamenti

Non ho esperienza nello scrivere ringraziamenti, ma non temo di essere giudicata e, al meno in questa parte della tesi, non seguiró le regole standard. Pertanto, poiché sono state molte le persone che hanno fornito il loro contributo a questo lavoro, non solo per le loro brillanti idee, i consigli e il loro tempo prezioso, ma anche per il loro supporto e il loro semplice sorriso, non scriveró ringraziamenti personali e ufficiali separati.

Innanzitutto, non sarei giunta a questo livello di studi senza l'amore e il supporto incondizionato di mia madre. L'incoraggiamento, l'entusiasmo e la fiducia da parte sua mi hanno impedito di interrompere ciò che mi è sempre piaciuto fin dalla quinta elementare e di rinunciare quindi alla scienza. Le parole non bastano per esprimere tutta la gratitudine che provo per lei; voglio solo dire che, se un giorno avrò la fortuna di avere figli miei, mi piacerebbe essere per loro la madre che lei è stata per me.

Desidero ringraziare in particolar modo la Fondazione Cassa di Risparmio di Padova e Rovigo (CARIPARO) per essere stata cos generosa nell'aver finanziato in toto la mia borsa di studio, il vitto, l'alloggio e i viaggi durante questi tre anni. Mi considero la persona piú fortunata del mondo: ho fatto quello che mi piaceva, vivendo nel Paese più bello della Terra e mi hanno anche pagato per questo. Grazie mille davvero. La mia gratitudine alla CARIPARO non sarebbe completa se non menzionassi la Prof. ssa Francesca Simion, una delle persone che hanno creato questo programma per gli studenti stranieri del Dottorato. Grazie per avermi scelta tra i numerosi candidati e quindi per avermi permesso di venire in una delle citt di Galileo Galilei, sede di una delle università più antiche del mondo. Grazie a questo programma ho incontrato persone fantastiche provenienti da ogni parte del mondo. Mariana, Sanja, James, Surachai, Leon, Hedzer, Van Sy e Khanh sono stati (e spero lo siano ancora) i migliori amici che io avessi mai potuto sognare di avere. Mi mancano moltissimo e spero tanto

di incontrarli di nuovo un giorno. Credo che questo programma sia davvero incredibile e spero che, nonostante la crisi finanziaria, i giovani studenti appassionati di tutto il mondo continuino a venire a Padova ogni anno per dare il proprio contributo alla scienza e avvicinarsi non solo alla cultura italiana ma anche alle culture di altri Paesi. Vorrei inoltre ringraziare il Servizio Accoglienza Ospiti Stranieri (SAOS). La situazione del mio permesso di soggiorno è stata davvero penosa i primi due anni della mia permanenza in Italia, ma sono consapevole che il Direttore del SAOS, dott. Babetto, e i suoi colleghi hanno fatto veramente di tutto per aiutarmi.

Per me il Dipartimento di Astronomia è sempre stato il posto migliore di tutta Padova. Qui mi sono sempre sentita al sicuro, a mio agio e benvenuta. Vorrei ringraziare tutte le persone che ho incontrato all'interno del Dipartimento per il loro sorriso, la loro disponibilità ad aiutarmi e per aver reso il Dipartimento stesso un luogo colmo di ospitalità. Sono davvero riconoscente al Prof. Francesco Bertola ed Elena Dalla Bontà per la loro immensa cordialità e attenzione. Grazie a loro sapevo dove potevo trovare aiuto e calore. Vorrei ringraziare i docenti del Dipartimento di Astronomia e dell'Osservatorio per le loro interessanti ed utili lezioni. Mi è davvero dispiaciuto (e mi dispiace tuttora) di non aver saputo Italiano. Mi sono sentita davvero in imbarazzo quando, a causa mia, in classe i docenti hanno svolto le lezioni in lingua inglese con paziente sopportazione da parte degli studenti. Pertanto un grazie va al 21, 22 e 23 cicli per la loro comprensione.

Sono davvero grata al mio supervisore, il Prof. Alberto Franceschini, per avermi condotta alla scoperta dell'Astronomia dell'infrarosso e della Cosmologia osservativa, per avermi fornito fin dai primi giorni un interessante argomento di studio, con una scrivania e un computer (mi è stato detto che non è una prassi normale per gli altri Dipartimenti dell'Università) e per avermi presentato il suo team di ricerca composto da Mattia Vaccari, Stefano Berta, Giulia Rodighiero ed Elena Ricciardelli. Mi ricorderò sempre che, quando Mattia non era impegnato nelle sue uscite, il suo aiuto ha spronato il mio lavoro e la sua disponibilità nel fornirmi codici IDL, cataloghi, libri, ecc. mi ha sempre colpita. Non dimenticherò mai il contributo di Stefano e Giulia al mio lavoro (per i dettagli si vedano i Chapters 4 e 5). Il loro aiuto e le loro risposte alle

mie sciocche domande di Astronomia dell'infrarosso hanno reso i miei giorni di lavoro molto più facili. Elena mi ha aiutato e fornito così tanto supporto che è difficile farne una stima. Le ho chiesto aiuto milioni di volte e lei miliardi di volte mi ha aiutata. Non dimenticherò nemmeno la solare Lucia Marchetti per il suo ottimismo quotidiano e l'enorme aiuto con i cataloghi SWIRE, Topcat, le traduzioni dallitaliano all'inglese, ecc. Sentiti ringraziamenti a tutti voi e spero che continueremo a lavorare insieme.

Ho apprezzato enormemente tutto ciò che il Dott. Forman ha fatto per me. Grazie alla sua infinita accoglienza ed amicizia, al suo tempo prezioso e ai suoi consigli, mi sono sempre sentita la benvenuta al CfA avendo sempre la sensazione che il mio lavoro fosse necessario. Sono molto grata a Christine Jones, Steve Murray e Ryan Hickox per avermi aiutata a lavorare con i dati di *Chandra* e per avermi permesso di sentirmi parte di una piccola comunità all'interno dell'enorme Centro di Astrofisica.

Gran parte degli ultimi ringraziamenti sono davvero ufficiali. Vorrei ringraziare A. Klypin e A. Kravtsov per avermi lasciato usare i risultati delle simulazioni cosmologiche dei loro gruppi. Inoltre, il Chapter 6 sarebbe stato molto meno coerente senza le ampie discussioni e i preziosi consigli di A. Kravtsov. Vorrei ringraziare Mark Frost e Matthew Thomson dell'Astronomy Centre dell'University of Sussex per aver condiviso con me il "npt-code" e le maschere di campo: senza il loro aiuto perderei molto più tempo su queste componenti tecniche dell'analisi del *clustering*.

Da ultimo, ma non per importanza, grazie ad Alexey: come ha detto un mio amico, sei il supervisore della mia vita e spero tanto che tu non smetta mai di svolgere questo ruolo.

# Acknowledgements

I have no experience in writing acknowledgements but I am not afraid to be judged and at least in this part of my thesis I am not going to follow "standards". Therefore, as a lot of people contributed to this work, not only by their brilliant ideas, advices and time, but also by their support and just smiles, I will not separate the personal and official acknowledgements.

First of all, I would have never come to this level of my education without the love and unwavering support of my mother. Her encouragement, enthusiasm and belief prevented me to cease what I have liked since the fifth grade and go out of Science. Words are not enough to express all my grateful feeling for her; I just want to say that if one day I am blessed to have my own children, I would like to be the same mother for them as She is for me.

I am very thankful to Fondazione Cassa di Risparmio di Padova e Rovigo (CARIPARO) for being so generous to fully fund scholarship, food, accommodation and trips during three years. I considered myself the luckiest person in the world I was doing what I liked, I was living in the most wonderful country on the Earth and I was paid for it. Thank you, thank you indeed. My gratitude to CARIPARO would not be full without mentioning Prof. Francesca Simion, the person who created this program for foreign PhD students. Thank you for choosing me among a lot of other applicants and thereby allowing me to came to a town of Galileo Galilei with one of the oldest universities in the world. Thanks to the program I met fantastic people from the entire world. Mariana, Sanja, James, Surachai, Leon, Hedzer, Van Sy and Khanh were (hope they are) the best friends I would ever had dreamed to have. I miss them so much and I hope very much to meet them one day again. I think the program is something really amazing and I hope that despite of any financial crisis, young and eager people around the world will come to Padova every year to make their contribution to the Science and to become familiar not only with the Italian culture but also with

cultures of other countries. I also would like to thank the Servizio Accoglienza Ospiti Stranieri (SAOS). The situation with the permesso di soggiorno was miserable for the first two years of my staying in Italy, but I felt that the Director of the SAOS, Dr. Babetto and his colleagues made everything to help.

The Department of Astronomy was always the best place for me in Padova. I always felt safe, comfortable and welcome there. I would like to thank all people I met within the Department for their smiles, wish to help and for making the Department a place filled of hospitality. I am grateful from all my heart to Prof. Bertola and Elena Dalla Bontà for their immense cordiality and attentiveness. Thanks to them I knew where I could look for help and warmness. I would like to thank Professors from the Department of Astronomy and the Observatory for their interesting and useful lectures. I was deeply sorry (and I am) for not knowing Italian. I felt really embarrassed when because of me in a classroom Professors gave lectures in English and students patiently born it. Therefore, thank you, the 21st, 22nd and 23d circles, for being understandable.

I am sincerely grateful to my Supervisor, Prof. Alberto Franceschini, for taking me in Infrared Astronomy and Observational Cosmology, for providing me from the first days with an interesting topic to study, with a desk and a computer (I was told it is not normal for other University Departments), and for introducing me to his research team consisting of Mattia Vaccari, Stefano Berta, Giulia Rodighiero, Elena Ricciardelli. I will always remember that, when Mattia was out of his trips, his help pushed my work forward, and his enormous will to provide me with IDL codes, catalogues, books, etc. every time astonished me. I will always remember the contribution of Stefano and Giulia to my work (see Chapters 4 and 5 for details :). Their help and willingness to answer to my silly IR Astronomy questions made my working days much more easier. Elena gave me so much help and support that it is hard to evaluate. I asked for her help million times and billion times she helped me. I also will not forget sunny girl Lucia Marchetti for her every day optimism and huge help with SWIRE catalogues, Topcat, translations from Italian to English, etc. Many, many thanks to all of you, and I hope we will continue working together.

I appreciate greatly everything Dr. Forman has done for me. Thanks to his unlimited hospitality and friendship, time and advices, I always felt that I was

welcome to CfA and my work was needed. I am very grateful to Christine Jones, Steve Murray and Ryan Hickox for helping me work with Chandra data and letting me feel that I was a part of a small society among that huge Center for Astrophysics.

Almost the last acknowledgements are indeed official. I would like to thank A. Klypin and A. Kravtsov for letting me use the outputs of the cosmological simulations belonging to their groups. Moreover, Chapter 6 was much less coherent without extensive discussions and invaluable advices by A. Kravtsov. I would like to thank Mark Frost and Matthew Thomson from the Astronomy Centre of the University of Sussex for sharing with me "npt-code" and field masks: without their help I would spend much more time on these time-consuming technical components of clustering analysis.

And, last but not least, to Alexey: as a friend of mine said, you are a supervisor of my life and I only wish you will never resign your position.

# Contents

# I Introduction

From the first extragalactic surveys (e.g. Shapley & Ames map (1932), the Zwicky catalogue [267], the Lick map (Shane & Wirtanen (1967); for details see, e.g., [27]) it has been recognized that galaxies in our neighborhood are not randomly distributed on the sky, but appear to be arranged into a complex network of clumps, filaments and voids, which is often called the "Cosmic Web" (Fig. 1). Understanding how and when the structures formed from the initial conditions presented in the early universe is one of the principal challenges in modern cosmology. Of particular importance are the studies focused at the galaxy populations as close as possible to the epoch of peak of cosmic star formation.

According to the current concensus from both theoretical and observational studies, major developments in the history of the star formation in the Universe happened at high redshifts, $z > 1$. The observational evidence puts the epoch of the peak star formation near $z \approx 2 - 2.5$ (Madau et al. [147]). The observations also point towards close co-evolution of the star-formation history of galaxies and the emissivity of active galactic nuclei at $z < 2$ (see Fig. 2 and Franceschini et al. [79], Silverman et al. [228], Shankar et al. [221]), implying that there is a common mechanism triggering these activities and playing a major role in galaxy formation. What exactly is responsible for this co-evolution, however, remains unclear. Very different processes, such as major galaxy mergers and secular evolution, both can potentially funnel gas to the galactic nuclear region, thus simultaneously powering concurrent star formation and accretion onto the central supermassive black holes [229]. New multiwavelength observations of the high-redshift objects are crucial for unraveling the details of the physical processes in the $z > 1$ galaxies, which most likely shaped the properties of the most massive galaxies observed today (see, e.g., Franceschini et al. [81]). The

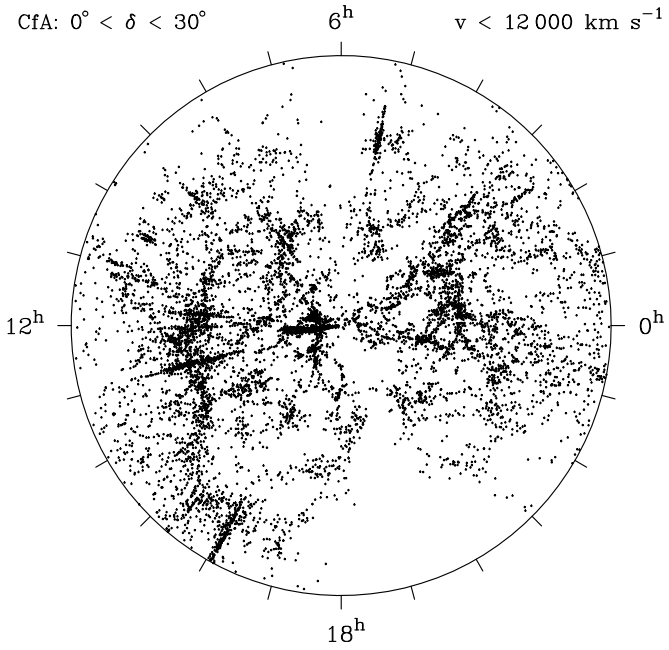



Figure 1. The distribution of galaxies in the nearby Universe as derived from the CfA redshift survey [62]. Even though these data are not perfect by modern standards (there is a gap due to the Galactic plane; the rich clusters are over-represented [181], etc.), one can clearly see the main features of the large scale structure — the "Great Wall" on the left, huge empty void regions, and the radial "fingers of God" caused by virialized motions of galaxies within the clusters.

observations of high-redshift galaxies and AGNs have only recently started to be "routine" with the advent of sensitive, wide-field surveys with *Spitzer* infrared space telescope and the *Chandra* X-ray observatory. *The main topic of this thesis is to use the clustering properties of high-z objects detected in those surveys to constrain their mass scale and environmental properties.*

The basic statistical methods required to analyze galaxy distribution were introduced in the 70s [186] and since then various types of cosmic sources were explored with a goal to describe and compare their clustering properties. These studies have revealed that galaxy cluster depending on their type, luminosity, morphology, that is a direct insight into the underlying physics of galaxy formation and evolution. The results presented in this thesis are based on the clustering analysis of normal and active galaxies at high redshifts, implemented through measurements of the two-point correlation function for each class of objects and matching these measurements to the clustering models for dark matter halos. The outline of the thesis is as follows.

In Chapter II, we provide a short summary of the main statistical tool widely

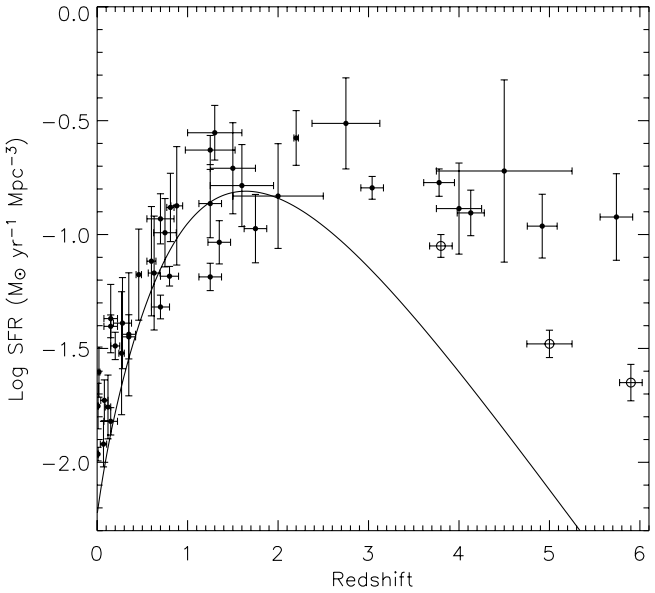



Figure 2. The evolution of the star formation history of galaxies (data points) and the mass accretion rate ($M_\odot\,\mathrm{yr}^{-1}\,\mathrm{Mpc}^{-3}$) of SMBHs (solid line). SFRs are the compilation from Hopkins [104] (filled circles) and Bouwens et al. [30] (open squares). The mass accretion rates have been modeled based on luminosity function of X-ray selected AGNs with $L_{2-8\mathrm{keV}} > 10^{42}\,\mathrm{ergs}^{-1}$. Reproduced from Silverman et al. [228].

used to describe clustering properties of galaxies, the two-point correlation function. We discuss the essential steps in the estimation of the correlation function from the data, and how the spatial and angular correlation functions can be linked by the Limber equation. The correlation function measurement for different classes of sources presented later in the thesis, are contrasted with the clustering models for dark matter halos. In Chapter III we describe the cosmological simulations used in these analysis, and also the techniques for "populating" the outputs of such simulations with model galaxies.

We then proceed to the description of results for three different classes of high-redshift sources. In Chapter IV, we analyzed clustering properties of star-forming galaxies emitting at 24 $\mu$m and detected by the *Spitzer*/MIPS camera in the SWIRE Lockman Hole field. A selected sample of $\sim$ 20000 objects with fluxes $S_\nu(24\,\mu\mathrm{m}) > 310\,\mu$Jy was divided into low-redshift ($z_{\mathrm{mean}} = 0.7$) and high-redshift ($z_{\mathrm{mean}} = 1.7$) subsamples using an optical/near-IR color criterion calibrated with GOOODS surveys. We then derive the angular correlation function for each subsample and use the Limber equation to infer the spatial correlation lengths. The spatial correlation length is used to estimate the mass scale of the dark

matter halos associated with the 24 $\mu$m sources, and the fraction of halos which host an mid-IR bright galaxy.

In Chapter V, we turn to a galaxy population in many respects opposite to the actively star-forming galaxies. These objects are massive, passively evolving galaxies which already formed a high stellar mass by a redshift of $\sim 2$. They were previously selected in the SWIRE ELAIS-S1 field by the presence of the 1.6 $\mu$m stellar peak redshifted into the 5.8 $\mu$m *Spitzer*/IRAC band. These objects, called "IR-peakers", likely represent the most massive galaxies at $z \sim 2$, with the estimated stellar mass $M^* \approx 1.6 \times 10^{11}\, M_\odot$. We present the angular correlation function for these sources detected at small scales, and discuss how using a power-law modeling of these data leads to the biased results on the spatial clustering amplitude. Much more consistent results follow from a more complete analysis based on direct projection of the spatial correlation function model derived from cosmological simulations.

The third population of sources, whose clustering we discuss in Chapter VI, is the X-ray selected AGNs. The sample of $\sim 1500$ objects with the median redshift $\sim 1$ was derived from the *Chandra* survey in the Boötes field. For these sources, accurate spectroscopic redshifts were available from the previous *MMT*/Hectospec observations, so we are able to study their spatial clustering more directly. In particular, using the spatial correlation functions along and perpendicular to the line of sight, we are able to constrain simultaneously the mass scale and halo occupation properties of the *X-ray* AGNs.

In the analyses below, we adopt a $\Lambda$CDM cosmological model, in which dark matter (DM) is composed of cold, weakly interacting, massive particles. Throughout the thesis, the following conventions are used to report observational results. All cosmology-dependent quantities are computed assuming a spatially flat model with parameters $\Omega_M = 0.268$ and $\Omega_\Lambda = 0.732$ (best-fit $\Lambda$CDM parameters obtained from the combination of CMB, supernovae, BAO, and galaxy cluster data, see [252]). The masses in Chapter V are reported for a Hubble constant of $h = 0.71$ (e.g., [124]), while in the rest of the thesis the masses are quoted for $h = 1$ and the explicit $h$-dependence is given. All distances are comoving. The parameter uncertainties are quoted at a confidence level of 68%.

# II Two-point correlation function

In the Chapter we provide a summary of the two-point correlation function formalism commonly used in analyzing the large scale distribution of galaxies in the Universe. Section 2 summarizes how the correlation function can be estimated from the data. We then discuss the spatial and angular correlation functions (§ 3, 4), and how they can be related through the Limber equation (§ 5). The mathematical framework summarized in this Chapter is used throughout the Thesis.

## 1 Introduction

In striking contrast with the extremely high level of isotropy observed in the temperature of the cosmic microwave background (e.g. [59]), galaxies are not distributed throughout the Universe in a random manner. According to the gravitational theory of instability the present structures originated from tiny fluctuations in the initial mass density field. This has shaped the large scale structure of the Universe (Fig. 1), which consists of vast empty regions (voids), and strings of dark and luminous matter (walls) where billions of galaxies are found [184, 181, 67]. The first map of the sky revealing convincing clustering of galaxies was the Lick survey undertaken by Shane and Wirtanen (1967). However, only in early 80s, the Center for Astrophysics (CfA) survey, in which 3D-position of galaxies were recorded, revealed the whole wealth of large-scale structures in a "slice of the universe" extended up to $150\,h^{-1}$Mpc (e.g. [58], [62]). Together with improving observational strategy, technology and technique, various statistical tools have been developed to quantify distribution of galaxies. They are methods of nearest neighbor, count in cells, power spectra,

and correlation functions [184, 158, 28]. Particularly, the two-point correlation function $\xi(d)$ is a simple, but powerful tool that first was applied by Totsuji and Kihara [243] in 1969 to analyze angular data taken from the Lick survey. Following the works of Peebles [185, 186], the two-point correlation function has become *a standard* for studying clustering properties of galaxies [117, 178, 157]. It can be defined in terms of the probability $dP$ of finding a galaxy within a small volume $dV$ lying at a distance $d$ from an arbitrary chosen galaxy:

$$dP = n[1 + \xi(d)]\, dV, \tag{2.1}$$

where $n$ is the mean number density over the whole sample volume (e.g. [184]). For a completely random distribution $\xi(d) = 0$. Positive $\xi(d) > 0$ and negative $\xi(d) < 0$ values indicate clustering and anti-clustering, respectively. This definition assumes isotropy and homogeneity of the point (i.e. galaxy) distribution, otherwise the function $\xi(d)$ should depend on a vector quantity. Totsuji and Kihara [243] also were the first to fit their angular correlation function measurements by a power-law (see reproduction of their plot in [157]) and to derive spatial clustering strength, $r_0$, by using the Limber equation [137], [184] (§ 5). It is remarkable but only recently some attempts have been made to explain this power-law shape of the correlation function, which is not predicted by numerical CDM simulations (see e.g. [12], [127] and Chapter III) and, moreover, the first deviations from the pure power-law behavior have been detected due to the largest up-to-date extragalactic survey, SDSS (the Sloan Digital Sky Survey), [263].

## 2 General definitions

*Estimator for the two-point correlation function*

In order to estimate the two-point correlation function $\xi(d)$ for a sample of $N_\mathrm{d}$ galaxies, one needs to determine separations of all source pairs and to bin these separations to form the data-pair count $DD(d)$, which is the number of source pairs with separations within the interval $[d - \Delta d/2, d + \Delta d/2]$. As the clustering is characterized by the *excess* probability of finding a galaxy at a certain distance from another galaxy, the number of data pairs $DD(d)$ is compared with

the expected number of pairs $RR(d)$, calculated from large artificial catalogues of intrinsically randomly distributed objects. This random catalogue should faithfully reproduce all observational distortions presented in the data sample. Example of such distortions are boundaries of the survey region; gaps in the data due to the presences of nearby stars, fiber collisions in a spectrograph or spatial variations of the sensitivity in a detector; variations of the selection efficiency with redshift; variations in magnitude/flux limit within the survey boundaries due to the different coverage of the field area by observations, etc.

Several estimators have been used to obtain $\xi(d)$ from a given data set [58, 134, 98, 193, 118]. The difference between them lies mainly in their method of boundary correction. The most widely used is the Landy & Szalay [134] estimator, which has been constructed to have Poissonian variance and which appears to outperform other popular estimators (see [118] and [255], [158] for details):

$$\xi(d) = \frac{[DD] - 2[DR] + [RR]}{[RR]}. \tag{2.2}$$

[DD], [DR] and [RR] are the normalized pair counts, i.e.,

$$\begin{aligned} [DD] &\equiv DD(d)\,\frac{N_\mathrm{d}(N_\mathrm{d}-1)}{2}, \\ [RR] &\equiv RR(d)\,\frac{N_\mathrm{r}(N_\mathrm{r}-1)}{2}, \\ [DR] &\equiv DR(d)\,\frac{1}{N_\mathrm{r}N_\mathrm{d}}, \end{aligned} \tag{2.3}$$

where $DR(d)$ is the number of pairs between cross-correlated data and random catalogues, $N_\mathrm{d}$ and $N_\mathrm{r}$ the total number of objects in the data and random catalogues, respectively. It was noted by Gilli et al. [90] that the difference in the ways of the pair counting may lead to a confusion in definition of [DD], [DR] and [RR]. The notations used above (eq. 2.3) assumes that the number of *DD* and *RR* pairs are counted only *once*. If one counts *DD* and *RR* pairs *twice*, the total number of data and random pairs are $N_\mathrm{d}(N_\mathrm{d}-1)$ and $N_\mathrm{r}(N_\mathrm{r}-1)$, respectively. This results in dropping the factor of 2 in the denominators of eq. 2.3 and in a different calculated formula for the Landy & Szalay estimator (e.g. [261]). Despite of these differences, both ways of pair counting lead to the same measured $\xi(d)$.

We note that the estimation of the correlation function is a time consuming process, especially if a "naive" method of pair counting is used. If there are $N$ objects in a sample, $N(N-1)/2$ distances between each pair have to be calculated. Thus, when $N$ is large, this direct pair counting method requires $N^2$ of the computation time. Fortunately, there are the so-called hierarchical tree methods that significantly reduce the time needed for the distance calculation. In particular, we use a pair-counting code, *npt-code*, utilizing the k-dimensional (kd-tree) algorithm [167, 84], in which the computational time is proportional to $N^{3/2}$. The *npt-code* was provided by The Auton Lab at Carnegie Mellon University School of Computer Science[1].

*Error estimation*

Statistical uncertainties which are assigned to $\xi$ measured using the Landy & Szalay estimator are

$$\delta\xi(d) = \frac{1+\xi(d)}{\sqrt{DD(d)}}, \tag{2.4}$$

[186]; this equation includes standard deviation in the count of pairs for the Poisson distribution (Poisson shot noise $\propto \sqrt{DD}$) and intrinsic variance term taking into account correlation of points [58, 255].

There is no concensus in the literature about an ideal method which fairly estimates statistical and systematical uncertainties in the correlation function (e.g. [28], [14], [192]). It is considered that errors given by Eq. 2.4 are underestimated [77, 192, 90] when source pairs are not independent. Moreover, this uncertainties do not account for cosmic variance. The former issue is alway the case, simply because a single object appears in many different separation bins through the numerous pairs in which it participates [255]. These correlations are incorporated by computing the covariance matrix using one of the widely used methods: the bootstrap resampling of data [9], the field-to-field and the jackknife subsampling of data (used in [86, 261, 254, 209] and many others), mock catalogues [61, 35, 163] and analytic approximations [68, 35]. However, it was shown in many studies that the covariance matrix is substantially diagonal (see e.g. [35], [24], [50]) at small scales, i.e. uncertainties are dominated by Pois-

[1] http://www.autonlab.org

sonian shot noise. At large scales the covariance should be taken into account. The cosmic variance component of uncertainty arises due to the limited volume of the Universe sampled by a survey [28, 158], its contribution to the errors of clustering signal should to be taken into account when one intends to compare the clustering of the same population of sources across different fields [90]. In this case the best way is to use mock catalogues from $N$-body simulations or semi-analytical models of structure formation with a recipe for populating the dark matter distribution with galaxies. The mock catalogues have to reproduce the observed properties of sample galaxies (for example, color, redshift distribution, fluxes in a given band, luminosities). The ensemble errors are then computed from the scatter in the different realization of the catalogue. Unfortunately, in spite of the importance of error estimation in the clustering analysis, there is no standard prescriptions for calculating of uncertainties for the correlation function. In general, the procedure highly depends on the data sample [178, 209, 90, 24, 175]. For the samples of sources studied in this thesis, we estimate errors to the measured correlation functions using eq. 2.4.

*Power law fitting*

It is well recognized from observations (see e.g. [263], [201]) and cosmological $N$-body simulations (e.g. [127], [232]) that the correlation function has a complex form (Fig. 33) that might be interpreted through by taking into account the distribution of galaxies within non-linear structures forming as the perturbations in the Universe grow (see Chapter III, § 9.1). The features in the correlation function which reflect these effects are expected at small and large scales, while in the intermediate range of scales, it can usually be well approximated by a power law,

$$\xi(d) = (d/d_0)^{-\gamma}, \tag{2.5}$$

where $d_0$ is referred to as the correlation length (or the correlation strength), a scale at which the density of galaxies is greater than that of the background by a factor of two [140]. For galaxies in the local Universe the correlation strength is about $5\,h^{-1}$ Mpc and $\gamma \approx 1.8$. When one fits power law to the observational data, the resulting $d_0$ and $\gamma$ are often not independent and the uncertainties must include the covariance of these parameters.

*Integral Constraint*

Due to the finite size of any survey, the true mean density of objects is usually unknown and the correlation function measured from a sample is always underestimated especially at scales comparable with the size of the field [187, 184]. The resulting systematic offset in the estimated correlation function $\xi_{\rm LS}$ is corrected by a factor $\xi_\Omega$ [134]:

$$1 + \xi_{\rm LS} = \frac{1 + \xi}{1 + \xi_\Omega} \approx 1 + \xi - \xi_\Omega, \tag{2.6}$$

where $\xi$ is the model correlation function given by the Eq. 2.5, for instance, and $\xi_\Omega$ is the mean of $\xi$ over the sampling geometry. In literature, $\xi_\Omega$ is called the integral constraint and often denoted as IC (and we also use this notation hereafter). For a given sample, the integral constraint correction is constant over all scales. Its value increases with the clustering strength and decreases with the field size. For the angular correlation function (§ 4) the IC is significant and always considered, in spacial clustering (§ 3) the correction is negligible in comparison with detected clustering signal [188]. In practice, the integral constraint is calculated numerically using a method proposed by Rosher et al. [206]:

$$\mathrm{IC} = d_0^\gamma \frac{\sum_i RR(d_i)\, d_i^{1-\gamma}}{\sum_i RR(d_i)}. \tag{2.7}$$

As was mentioned above, the integral constraint depends on both the size of a survey and galaxy clustering. However, once the correlation slope $\gamma$ is fixed, $\mathrm{IC}/d_0$ only depends on the size and the shape of the survey area [136]. Therefore, one needs to make a robust estimate of $\gamma$ from the observational data or, as more often applied, the slope is fixed to a standard value 1.8 and then $\mathrm{IC}/d_0$ is estimated. We calculated the ICs for the angular correlation functions in Chapters IV and V using a grid of $\theta_0$ (which is the angular correlation length) and $\gamma$, and Equation 2.7.

## 3 Spatial correlation function

The correlation function in real space is expected to be isotropic, so $\xi(r)$ is a function of the 3D separation only. When the object redshifts are used to derive the distances, the correlation function is distorted in the line-of-sight direction.

This redshift-space distortion arises from two effects, operating respectively on small and large scales. On small scales, random peculiar velocities within galaxy groups or clusters cause radial elongations of observed structures (Fig. 1), known as "fingers of God". Another (Kaiser [112]) effect occurs at large scales due to the coherent motions of galaxies falling towards the potential well of assembled structures. It usually leads not to a stretching but to a flattening of structures along the line-of-sight. Thus, peculiar motions prevent galaxies from following pure Hubble flow, affect measured redshifts and, hence, the estimated line-of-sight distances, while the galaxy positions on the sky remain unaffected. Therefore, when the correlation function is computed from a redshift survey, it is usually measured as a function of the projected (transverse) separation, $r_{\rm p}$, and the line-of-sight (radial) separation, $\pi$. This function, $\xi(r_{\rm p}, \pi)$, is independent of direction in real space (see left panel in Fig. 37), but is strongly distorted in redshift space (Fig. 3), where the projected separations remain the true measures of distance, while the radial separations are altered by peculiar velocities. In order to estimate $\xi(r_{\rm p}, \pi)$, eq. [2.2] still can be used, but the pairs must be counted for each combination $(r_{\rm p}, \pi)$.

Given the angular separation between two objects, $\theta$, and redshifts, $z_1$ and $z_2$, the comoving separations $r_p$ and $\pi$ can be computed as follows. First, one computes the radial comoving distances, $D_{c,1}$ and $D_{c,2}$, corresponding to the object redshifts (see, e.g., [103]). Then, following [58] we have

$$\pi = |D_{\rm c,1} - D_{\rm c,2}|, \tag{2.8}$$

and

$$r_{\rm p} = \left[2 D_{\rm c,1} D_{\rm c,2} (1 - \cos\theta)\right]^{1/2}. \tag{2.9}$$

One can also define a formal 3D separation,

$$s = \left(r_{\rm p}^2 + \pi^2\right)^{1/2}, \tag{2.10}$$

but it should be kept in mind that $s$ is not equivalent to the true 3D separation, $r$, because of the redshift space distortions.

As only the line-of-sight separations, $\pi$, are affected by the object peculiar velocities, it is useful to consider the correlation function projected on the sky

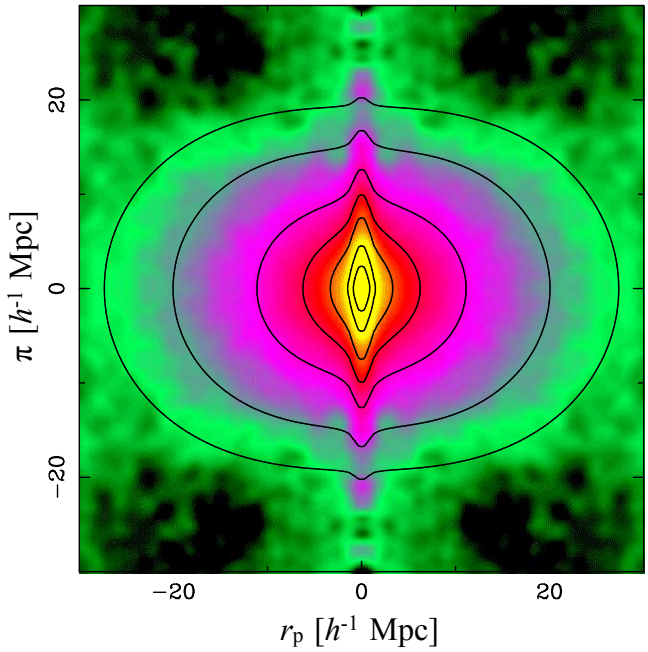



Figure 3. The two-dimensional correlation function $\xi(r_\mathrm{p}, \pi)$ for the 2dFGRS, plotted as a function of transverse ($r_\mathrm{p}$) and radial ($\pi$) pair separation. The function was estimated by counting pairs in boxes and then smoothing with a Gaussian. The results obtained for the first quadrant are repeated with reflection in both axes to show deviation from circular symmetry. Overplotted lines correspond to the function calculated for a given theoretical model. This diagram clearly demonstrates the two effects of redshift distortions: "fingers of God" elongations at small scales and the coherent Kaiser flattening at large radii. The figure is adopted from Peacock et al. [182].

plane,

$$w(r_\mathrm{p}) = \int_{-\infty}^{\infty} \xi\left(\sqrt{r_\mathrm{p}^2 + \pi^2}\right) d\pi, \tag{2.11}$$

because it is not modified by the redshift-space distortions [58]:

$$\int_{-\infty}^{\infty} \xi^{(\mathrm{true})}\left(\sqrt{r_\mathrm{p}^2 + \pi^2}\right) d\pi = \int_{-\infty}^{\infty} \xi^{(\mathrm{obs})}(r_\mathrm{p}, \pi)\, d\pi, \tag{2.12}$$

where $\xi^{(\mathrm{true})}(r_\mathrm{p}^2 + \pi^2)^{1/2} = \xi(r)$. Last expression assumes that, here, the radial separation $\pi$ is not distorted by peculiar motions. Note that $w(r_\mathrm{p})$ has dimensions of length. Since the correlation function converges rapidly to zero with increasing pair separation, the integration limits do not have to be $\pm\infty$, but they have to be large enough to include all correlated pairs. However, large separations do not contribute significantly to the integrals in equations 2.11 and 2.12 but add noise, therefore the integration is truncated at some line-of-sight separation. Equation

2.12 can be re-written as [184]

$$w(r_{\rm p}) = 2 \int\limits_{r_{\rm p}}^{\infty} \frac{r\xi(r)dr}{\sqrt{r^2 - r_{\rm p}^2}}\,. \tag{2.13}$$

Because of the insensitivity of $w(r_{\rm p})$ to the redshift-space distortions, most of the studies which involve detailed modeling of the shape of the galaxy two-point correlation function are based on fitting the $w(r_{\rm p})$ measurements (e.g., [262, 180]). Using $w(r_{\rm p})$ instead of $\xi(r)$ is particularly straightforward in those cases when $\xi(r)$ can be sufficiently accurately approximated by a power law,

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma}. \tag{2.14}$$

It follows from eq. 2.13 that in this case the relation between $\xi(r)$ and $w(r_{\rm p})$ is simply [184]

$$w(r_{\rm p}) = A(\gamma)\, r_{\rm p} \left(\frac{r_{\rm p}}{r_0}\right)^{-\gamma}, \tag{2.15}$$

where

$$A(\gamma) = \Gamma(1/2)\,\Gamma([\gamma - 1]/2)/\Gamma(\gamma/2). \tag{2.16}$$

Therefore, the correlation length, $r_0$, and the slope of the true spatial correlation function $\xi(r)$ can be obtained immediately from the power-law fit to $w(r_{\rm p})$.

In Chapter VI, we also consider a projection of the measured 3D correlation function on the line-of-sight direction,

$$w(\pi) = \int\limits_{-\infty}^{\infty} \xi^{\rm (obs)}(r_{\rm p}, \pi)\, dr_{\rm p}. \tag{2.17}$$

In the absence of peculiar motions, $w(\pi)$ should be equivalent to $w(r_{\rm p})$. In particular, for a power-law approximation, eq. 2.15 will be valid also for $w(\pi)$.

In practice, the integration in eq. 2.17 and 2.11 is truncated at some comoving separations $l_{\rm max}$ to minimize noise. Assuming that the true correlation function follows a power law with index $\gamma$ at large separations, we can calculate the effect of this truncation as

$$C_\gamma(r) = \frac{\int_0^{l_{\rm max}} (r^2 + l^2)^{-\gamma/2}\, dl}{\int_0^{\infty} (r^2 + l^2)^{-\gamma/2}\, dl}. \tag{2.18}$$

The correction coefficient, $C_\gamma(r)$, is close to 1 for $r \ll l_{\max}$ and gradually decreases at larger distances. If $l_{\max} = 40\,h^{-1}$ Mpc and $\gamma = 2$ (close to the best-fit value obtained in Chapter VI), $C_\gamma = 0.94$ at $r = 2\,h^{-1}$ Mpc, 0.86 at $r = 6\,h^{-1}$ Mpc ($\approx$ the correlation length), and 0.65 at $r = 20\,h^{-1}$ Mpc.

Using equations 2.15 and 2.18, we define the quantities

$$\xi(r_\mathrm{p}) \equiv \frac{1}{A(\gamma)\,C_\gamma(r_\mathrm{p})}\frac{w(r_\mathrm{p})}{r_\mathrm{p}} \tag{2.19}$$

and

$$\xi(\pi) \equiv \frac{1}{A(\gamma)\,C_\gamma(\pi)}\frac{w(\pi)}{\pi}, \tag{2.20}$$

where $\gamma$ is determined from the power-law fit to the measured $w(r_\mathrm{p})$ (eq. 2.15). Thus defined, $\xi(r_\mathrm{p})$ should be a close approximation to the true spatial correlation function $\xi(r)$. In the absence of redshift-space distortions, $\xi(\pi)$ also should approximate $\xi(r)$. However, the peculiar motions (and in particular, the "finger of God" effect) suppress $\xi(\pi)$ on the smallest separations and enhance it relative to $\xi(r)$ on the intermediate scales, that is clearly demonstrated in Fig. 3 (see, however, Chapter VI for details). Therefore, we used $\xi(r_\mathrm{p})$ to determine the correlation length of X-ray selected AGNs (§ 29). The ratio $\xi(\pi)/\xi(r_p) = w(\pi)/w(r_p)$ reflects the amplitude of the peculiar motions and hence we used it to constrain the fraction of our objects located in the satellite dark matter subhalos (§ 30.3).

### 4 Angular correlation function

If the real spatial correlation function, $\xi(r)$, is a power-law (eq. 2.14), its 2D projection, the angular correlation function $w(\theta)$, will also be a power-law with the same slope $\gamma$:

$$w(\theta) = \left(\frac{\theta}{\theta_0}\right)^{1-\gamma} = A_w\theta^{1-\gamma} \tag{2.21}$$

where $A_w = \theta_0^{\gamma-1}$ is angular clustering amplitude and $\theta_0$ is angular correlation length.

In Chapters IV and V, where only angular positions of IR-galaxies are available, we compute angular correlation functions and then derive spatial clustering strength applying Limber inversion, described in the next section § 5.

## 5 Limber equation

The integral equation relating the angular and the spatial correlation function was suggested by Limber [137] and first applied to get spatial clustering strength of galaxies from Lick survey. The equation is mainly based on an assumption that there is no dependance of galaxy luminosities on the density of environment [43] and on the interpretation of the angular correlation function $w(\theta)$ as the convolution of galaxy redshift distribution and the spatial correlation function. In approximation of small angles ($\theta \ll 1$ [rad]):

$$w(\theta) = \frac{2}{c} \frac{\int_0^\infty dz\, N(z)^2\, H(z) \int_0^{\pi_{\max}} d\pi\, \xi\left(\sqrt{[D_{\rm M}(z)\theta]^2 + \pi^2}\right)}{\left[\int_0^\infty dz\, N(z)\right]^2}, \tag{2.22}$$

where $D_{\rm M}(z)$ is the transverse comoving distance to redshift $z$ ($D_{\rm M} = D_{\rm c}$ in the flat universe, see e.g. [103]) and $N(z)$ is the redshift distribution of sample galaxies. $H(z) = H_0 \sqrt{\Omega_{\rm M}(1+z)^3 + \Omega_k(1+z)^2 + \Omega_\Lambda}$ is the Hubble parameter at redshift $z$ and $\xi\left(\sqrt{[D_{\rm M}(z)\theta]^2 + \pi^2}\right) = \xi(r)$ is the spatial correlation function in real space [2].

Assuming that $\xi(r)$ follows a power-law (Eq. 2.14), one can simplify the Equation 2.22 to derive a direct link between the angular and spatial correlation lengths:

$$\theta_0^{\gamma-1} = r_0^\gamma\, A(\gamma) \frac{2}{c} \frac{\int_0^\infty dz\, N(z)^2\, H(z) D_{\rm M}^{1-\gamma}}{\left[\int_0^\infty dz\, N(z)\right]^2}, \tag{2.23}$$

where all variables are defined above and $A(\gamma)$ is given by Eq. 2.16.

[2] Note that $r_{\rm p} = D_{\rm M}\theta$ in the approximation of small angles.

# III Models of dark matter halos

According to the current scenario of structure formation in the Universe, galaxies form and evolve inside dark matter halos. A natural prediction of this paradigm is that the observed properties of galaxies such as stellar mass, luminosity, color, and clustering are tightly coupled to the depth of the potential or mass of dark matter halos in which the galaxies reside. There are several approaches which can be applied to connect galaxies and their dark matter halos (see Section § 6 of this Chapter). The most relevant of them is based on using the results of dissipationless cosmological ΛCDM simulations (§ 7) together with a prescription for populating halos with galaxies (§ 9.3). This approach is taken in our work (Chapters IV, V and VI), therefore the Chapter also presents the state-of-the-art simulations (§ 7.1 and § 7.2) whose outputs we use to analyze the clustering measurements of the high-redshift sources.

## 6 Introduction

The commonly adopted view of the large-scale structures growth and formation of objects such as galaxies and galaxy groups is as follows. The dark matter (DM) consists of collisionless particles which interact very weakly with the rest of matter and with the radiation field, and thus the perturbations in this component start growing at early times, and that growth is driven only by the gravitational instability. When the amplitude of density perturbations becomes large, the self-gravity overcomes the local expansion of the Universe and the overdense regions begin to collapse and form bound objects. The resulting quasi-static configurations of virialized dark matter are called "dark matter halos" (DM halos). At the point of the reversal of the expansion, the perturbation

density contrast, $\delta\rho/\bar{\rho} = (\rho - \bar{\rho})/\bar{\rho}$, reaches values around unity[1]. Around this point, the analytic linear of the structure formation becomes inapplicable, and in most cases, the nonlinear evolution of the dark matter structures has to be followed by numerical simulations (such as those described in § 7). The current dissipationless $N$-body cold dark matter (CDM) simulations are sufficiently advanced to faithfully trace a full range of the hierarchical growth DM halos, as the first small bound systems cluster and merge to form larger DM halos. At early epochs, most DM halos were small, low-mass ($M_{\rm halo} \leq 10^6\,M_\odot$) objects, and massive halos were very rare. In the present-day Universe, the mass distribution is dominated by massive DM halos, while the number of low-mass halos is slowly decreasing [232].

The baryonic and dark matter are bound by gravity. The gas concentrates at the centers of halos, then cools and condenses to form galaxies [258, 71, 25]. In this scenario of the galaxy formation, galaxy physical properties, such as luminosity or stellar mass, are expected to be tightly coupled to the depth of the halo potential and, if the halos are self-similar, to the halo mass. Therefore, it is likely that different types of galaxies populate different DM halos. The large-scale galaxy surveys and $N$-body simulations have revealed a clear picture of the luminosity and mass distribution in the Universe. However, establishing a direct link between the galaxy and DM halo populations proved hard [246, 48, 96]. Several approaches have been developed to study the halo-galaxy relation. The first class of methods is purely observational, when one attempts to derive the mass of the host halos using, e.g. dynamical measurements of bound satellites [168, 169], gravitational lensing [155, 38, 5], or X-ray studies [138]. These techniques are observationally expensive and have been applied to either massive halos (hosting galaxy clusters) or to galaxies in the local Universe, with only a few studies done at higher redshifts, $z \sim 1$ [101, 45]. The second class of methods can be classified as purely numerical, when one tries to simulate the process of galaxy formation in its cosmological context. The $N$-body treatment of the evolution of the DM component is combined with a hydrodynamical (e.g. [41], [231], [225]) or a semi-analytic ([114], [113], [232], [203] among others)

[1]The value of the critical overdensity depends on the exact combination of the cosmological parameters and on the topology of the perturbation.

treatment of the baryonic evolution. The advantage of these methods is that they track galaxy and DM halo evolution across cosmic time in a self-consistent way, providing predictions for the positions, velocities, star formation histories and other physical properties of galaxies populating the DM halos of a given mass range. However, the physics of many processes involved in the galaxy formation and evolution (such as the star and black hole formation and energy feedback from these processes) are still poorly understood, and some observed quantities cannot be reliably reproduced by the simulations. The third class of methods includes the studies when one tries to bypass any complex galaxy formation by using simple statistical models with adjustable parameters to link galaxies and DM halos. By adjusting model assumptions and their associated parameters, this so-called Halo Occupation Distribution (HOD) approach (§ 9.1) is able to relate with high accuracy the observed statistics of the galaxy populations, such as the luminosity and correlation functions, to statistical indicators of the DM halo distribution (e.g. [12], [49],[260],[266]). A major disadvantage of the HOD model is that it does not allow one to embed the merging history of a system when populating it with galaxies, whereas such information may have significant influence on the properties of the galaxies (e.g. [85], [54]). A similar approach is to establish a relation between the galaxy properties and the host halo mass by matching the observed galaxy luminosity (or stellar mass) function to the mass function of the DM halos (the so-called "abundance matching" technique, see § 9.2 for more details). The connection is based on the assumption that there is a unique, monotonic relation between the galaxy luminosity and halo mass [127, 237, 246]. Many observational statistical properties, for example, galaxy clustering as a function of luminosity, have been well reproduced by this method [247, 48, 47, 121, 96].

In what follows, we describe the cosmological simulations used in this work, and the analytic and numeric HOD modeling together with the "abundance matching" approach.

## 7 Numerical simulations

While the early, linear-regime stages of growth of the density perturbations can be followed analytically, the collapse of the fluctuations and the subsequent hierarchical build-up of structures is a highly nonlinear process which is accessible only through direct numerical simulation [57, 232].

Numerical cosmological simulations, at present carried out on powerful parallel supercomputers, start from the initial density field at an early epoch, when the typical matter density fluctuations around the cosmic mean density are small ($\delta\rho/\overline{\rho} < 0.3$) at the scales of the mean interparticle separation. The initial conditions are a realization of the Gaussian density field with a power spectrum of fluctuations which can be exactly computed for the assumed cosmological model (e.g. [220]). The field is evolved to the chosen initial epoch using a model sufficiently accurate to describe the evolution of fluctuations in the linear and mildly non-linear regime, such as the Zeldovich approximation or higher-order perturbation theory (e.g. [264], [198], [52]). In recent years, a number of observational probes have provided highly precise measurements of the cosmological parameters. The global cosmological model is thus tightly constrained, and the initial conditions for the numerical simulations can be considered unique and well specified. The main challenge for the simulations is, therefore, to faithfully follow the co-evolution of the dark and baryonic matter driven by gravitational instability and gas-dynamical processes affecting the baryons. The resources of the modern computers are still insufficient for high-resolution gas-dynamical simulations of a large volume. Therefore, the biggest simulations available at present, such as those used in our work, follow only the evolution of the cold dark matter component and thus are referred to as the dissipationless CDM simulations.

The dynamics of CDM is described by the Poisson equation for the gravitational field and the collisionless Boltzmann equation (also known as the Vlasov equation). If the density field is represented by a finite number of gravitationally interacting particle, this equation are identical to the standard Newtonian equations of particle motion. The cold dark matter is assumed to be made of elementary particles which interact only gravitationally. Therefore, a rep-

resentative subset of the characteristic equations is solved by discretizing and sampling the initial phase space by $N$ particles ("bodies") and then integrating their equations of motion in the collective gravity field. Such substitution of fluid by particles is coarse but becomes ever more accurate as the number of particles in the simulations is increased (see e.g. [29] for details).

The crucial parameters defining the accuracy of a simulation are the mass and spatial resolution (also known as the softening length or force resolutions), which usually are related (see for details [119], [29]). The mass resolution is the mass of the smallest mass element (a mass of a particle). The spatial resolution is separation between two particles at which the Newton's gravity law, $f \propto r^{-2}$, is "softened" to make the computation manageable. It is sometimes said that *the art of N-body simulations lies chiefly in the computational algorithm used to obtain the gravitational force* [22]. The most direct scheme is based on computing the force among each pair of particles. Although this direct integration scheme (e.g. [1]) is in principle the most accurate, the required number of operations scales as the the total number of particles squared, making it prohibitively expensive for large cosmological simulations. For this reason, different optimization schemes have been developed to trade between the computational speed and numerical resolution or accuracy (for recent reviews, see [22], [122], [6], [29]).

The first simulations of galaxy clustering were carried out in the 1970s using $\sim 1000$ particles and applying the particle-particle method [185, 93]. In early the 1980s, the development of new sophisticated schemes (mentioned above) led to an increase in the of particles to $20000 - 60000$ [66, 162, 57]. Ever since, the ever-increasing power of supercomputers and continuous refinement of numerical methods allows one to use more and more particles while simultaneously resolving finer and finer structures (for examples of high-resolution cosmological simulations see [237], [232],[130], [31], [121]). It is worth keeping in mind that various simulations with different sets of parameters and implemented numerical schemes are not only important to achieve the optimal mass and spatial resolutions for the specific problem, but also to study systematic errors associated with the numerical techniques [119, 29, 121].

$N$-body simulations have been essential for determining the properties of dark matter halos. Dark matter halos are approximately spherical, mostly viri-

alized objects. According to the theory of spherical collapse [184], an overdensity stops collapsing and virializes forming a halo when its mean density is 178 times the background density[2]. This picture is highly simplified (even though it gives faithful estimates of the halo interior properties [8]) because it describes the formation of a halo as the monolithic collapse from a single overdensity. The numerical simulations show clearly that the structures form hierarchically — smaller halos collapse and merge to form larger halos, which then merge again. And yet this complicated process leads to a high degree of regularity (on average) in the properties of the dark matter halos. For example, it turns out that the density distribution inside a halo on average follows the "universal" profile whose parameters only weakly depend on the halo mass, the power spectrum of initial density fluctuations, and the cosmological parameters [174]. Further, it has been demonstrated through the simulations that the mass function of DM halos can be written as an almost universal formula [110] which in fact is very close to the predictions of the analytic calculations of the collapse of ellipsoidal halos [223]. The simulations were also instrumental for calibrating the expected trends of the DM halo concentration and shape with mass and redshift [37, 4, 145], and for for studying the halo bias [165, 195, 215].

In the analysis described in Chapters IV — VI we use clustering properties of dark matter halos derived in large-volume numerical simulations. Particularly, we employ the dependance of halo clustering strength on mass and redshift. Further we review main characteristics of these simulations.

Our choice of the numerical simulation was driven primarily by two considerations. First, some of the astrophysical sources we analyze (e.g., the IR-peakers, Chapter V) are sufficiently massive and rare objects; therefore, the simulation should have a large volume in order to contain a sufficient number of such objects. Second, our analysis of the IR-peakers requires correct modeling of the small-scale clustering (§ 23) properties; the simulation should thus have a sufficient mass and force resolution to properly model subhalos within larger halos which give a strong contribution to the correlation function at small scales (§ 9). One of the largest simulations performed thus far is the Millenium

[2] An overdensity of 178 is appropriate for the flat, critical density universe. For other sets of cosmological parameters, the virial overdensity can be slightly different, see [69].

Simulation (MS) [232]. This numerical run was performed in a $500\,h^{-1}$ Mpc box with $2160^3$ particles. Unfortunately, the Millenium's resolution is insufficient to properly resolve the 1-halo term in the correlation function (see, e.g., Fig. 4 in Guo et al. [96]). The problem with numerical resolution is solved in the *Bolshoi* simulation [121] (e.g., the force resolution is $1\,h^{-1}$ kpc in *Bolshoi* vs. $5\,h^{-1}$ kpc in MS), and that simulation was run in a $250\,h^{-1}$ Mpc box, large enough to contain a sufficient number of massive DM halos. Driven by these considerations, we use *Bolshoi* for the analysis of clustering for actively star forming galaxies (Chapter IV) and massive IR-selected galaxies at $z \approx 2$ (Chapter V). For the analysis of the clustering of X-ray selected AGNs (Chapter VI), we use (for historical reasons) a simulation which is similar to *Bolshoi* but was run in a smaller box, $120\,h^{-1}$ Mpc.

These two sets of $N$-body dissipationless cosmological $\Lambda$CDM simulations are described in detail in Tasitsiomi et al. [237] (§ 7.1) and Klypin et al. [121] (§ 7.2). Both simulations were run using the Adaptive Refinement Tree code (ART, [129], [126]) which is based on the Adaptive Mesh Refinement numerical method [22]. To identify dark matter halos, the simulation outputs were analyzed by the Bound-Density-Maxima (BDM) algorithm [120] (see [127] and [121], for details). The BDM halo finder locates density peaks at separations greater than $10\,h^{-1}$ kpc, constructs the density, circular velocity, and velocity dispersion profiles to test whether the peaks correspond to gravitationally bound clumps, and then removes unbound particles. The final profiles, using only bound particles, are used to calculate the halo characteristics such as the virial mass $M_{\rm vir}$ (or $M_{180}$, see below), virial radius $R_{\rm vir}$, maximum circular velocity $v_{\rm max}$, etc. The identified halos are classified into distinct (host, parent) halos whose centers are not located within any larger virialized systems, subhalos (satellites, substructure) which lie within the virial radius of a larger halo. In the real Universe, centers of host halos can be identified as locations of the groups' central galaxies, while subhalos correspond to satellite galaxies.

The mass of a halo is traditionally defined within a radius. In the simulations we describe here, virial radius $R_{\rm vir}$ and $r_{180}$ are used. $R_{\rm vir}$ is a radius encloses the virial overdensity $\Delta_{\rm vir}$ with respect to the mean matter density $\rho_{\rm m} = \Omega_{\rm m}\rho_{\rm crit}$ at a given redshift. $\Delta_{\rm vir}$ is predicted by the spherical collapse model [184] and

depends on $\Omega_{\rm m}(z)$, and hence, on redshift [36]. In the "concordance" flat, low-density $\Lambda$CDM cosmological model, the virial overdensity is 360 at $z = 0$ and asymptotically approaches 178 at higher redshifts. Therefore, $r_{180}$, the radius corresponding to a mean overdensity of 180, is a good approximation to $R_{\rm vir}$ at redshifts $z \gg 0$. Another choice of the overdensity threshold, often found in the literature, is 200 with respect to the *critical* density at the given redshift, $\Delta^{200} = 200/\Omega_{\rm m}(z)$. The halo mass is related to the overdensity threshold through $M_\Delta = 4\pi R_\Delta^3 \rho_{\rm m} \Delta/3$. The difference in $M$ caused by different choices of $\Delta$ is not high but should be kept in mind. For example, $M_{200}$ (corresponding to a critical overdensity of 200) is 20-30% smaller than the virial mass of halos at $z = 0$ [121].

The maximum circular velocity $v_{\rm max}$, rather than the viral mass, is considered to be more robust characteristic of the halos for the reasons outlined in [128], [172], and [48]. The definition of $v_{\rm max}$,

$$v_{\rm max} = V_{\rm circ}\Big|_{\rm max} = \sqrt{\frac{G\,M(<r)}{r}}\Big|_{\rm max}, \tag{3.1}$$

clearly reflects the depth of the halo potential well. It is a well-defined quantity even for a subhalo immersed in a larger gravitationally-bound halo. It is more closely related to the central properties of a (sub)halo, and is less affected by tidal stripping which alters the density profile and, consequently, subhalo mass at large radii. Therefore, one expects that the stellar content of the halo and all baryonic processes in the center, including the AGN activity, are better correlated with $v_{\rm max}$ than with $M_{\rm vir}$. In a Milky-Way type halo, for example, the stellar disk extends to 15–20 kpc, the radius of the maximum circular velocity is $\simeq 40$ kpc[3], while the virial radius is $\sim 300$ kpc. Nagai & Kravtsov [172] and Conroy et al. [48] extend the argument further and suggest that the best indicator for the stellar mass and luminosity of satellite galaxies is $v_{\rm max}$ which their parent subhalo had before accretion onto the host halo. These authors indicated that the choice of subhalo $v_{\rm max}$ before or after accretion affects the correlation function at small scales where the subhalo contribution dominates. They also show [48] that the difference between the two definitions becomes much smaller at higher redshifts. Because our analysis concentrates on the high-redshift objects, and

[3] And the circular velocity is nearly the same at 20 kpc.

only in one case we use the clustering data on small scales, we use the $v_{\rm max}$ at the redshift of the simulation output.[4]

Even though $v_{\rm max}$ is a preferred parameter for a halo, there is a tight correlation between $v_{\rm max}$ and the mass, so we can easily convert between the two parameters. From the self-similarity arguments, one predicts (e.g., [29])

$$M_{\rm vir} \propto v_{\rm max}^3 E^{-1}(z), \tag{3.2}$$

where $E(z) = H(z)/H_0 = \left[(1+z)^3\,\Omega_{\rm m} + (1+z)^2\,\Omega_{\rm k} + \Omega_\Lambda\right]^{1/2}$. The normalization of this relation reflects the concentration of the halos, and thus is expected to be slightly different for distinct halos and subhalos (on average, subhalos are more concentrated than distinct halos with the same mass [121]). The $v_{\rm max} - M_{\rm vir}$ is indeed observed for the DM halos in the numerical simulations. Figure 5 shows the mass vs. $v_{\rm max}$ relation for host halos (*left*) and subhalos (*right*) in the *Bolshoi* simulation (§ 7.2) at $z = 0$. Figure 4 presents the same relation but for all DM halos (without making a host halo / subhalo distinction) derived in $\Lambda$CDM$_{120}$ run (§ 7.1) at several output redshifts.

### *7.1 N-body simulations in the* $120\,h^{-1}$ *Mpc box*

The $\Lambda$CDM$_{120}$ run described in Tasitsiomi et al. [237] and Conroy et al. [48] is a high-resolution dissipationless simulation in a flat $\Lambda$CDM cosmology with parameters $\Omega_{\rm M} = 0.3$, $h = 0.7$, and $\sigma_8 = 0.9$. The simulation follows the evolution of dark matter in a $120\,h^{-1}$ Mpc box. The box contained $512^3$ dark matter particles with mass $m_p = 1.07 \times 10^9\,h^{-1}\,M_\odot$; the peak resolution reaches $1.8\,h^{-1}$ kpc. The locations and velocities of the dark matter particles in the simulations were recorded at $z = 0.09$, 0.5, 1.0, 2.0, 3.1, 4.0, and 5.0. The simulation outputs were analyzed to identify gravitationally-bound halos (see p. 23). The effective completeness for the halo detection is reached for $\sim 50$ particles. The corresponding $v_{\rm max}$ limit is $\sim 80$ km/s, and the associated mass limit is $\sim 5 \times 10^{10}\,h^{-1}\,M_\odot$.

[4] The choice is primarily programmatic. The $v_{\rm max}$ at the redshift of the simulation output is already provided. To establish $v_{\rm max}$ before accretion, on the other hand, we need to reconstruct the full merging tree for the given object.

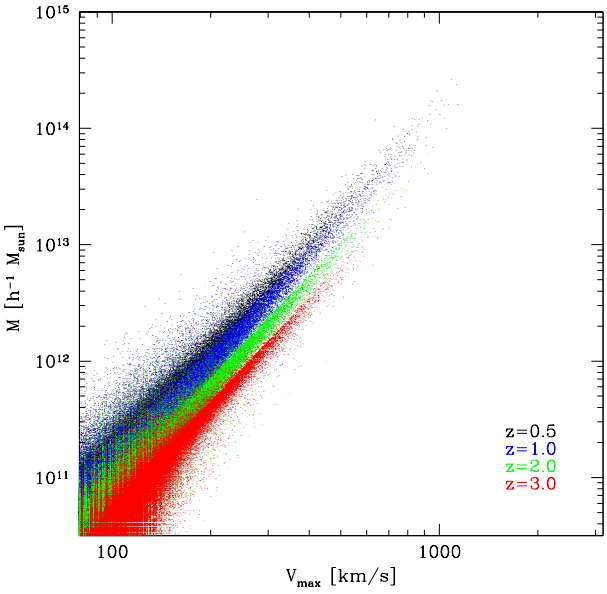



Figure 4. Mass vs. maximum circular velocity relation for the DM halos at several outputs from the $\Lambda$CDM$_{120}$ simulation.

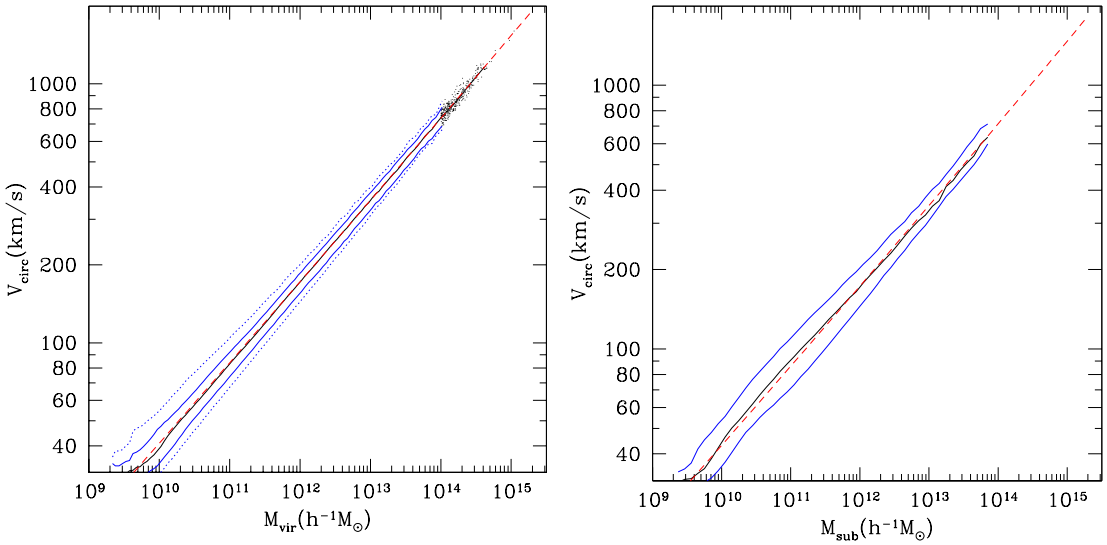



Figure 5. The $v_{\rm max} - M_{\rm vir}$ relations for distinct halos (*left panel*) and subhalos (*right panel*) in the *Bolshoi* simulation at $z = 0$ [121]. Thick lines on both panels show the median value of $v_{\rm max}$ ($V_{\rm circ}$ in Klypin et al. notations) as a function of mass; thin lines show the band containing 90% of the points. The thin dotted lines in the left panel are the 99% boundaries. The dots show individual halos with masses $> 10^{14}\,h^{-1}\,M_\odot$. For distinct halos and subhalos, the maximum circular velocity scales with virial mass as $v_{\rm max} \propto M_{\rm vir}^{1/3}$, with a small correction due to dependence halo concentration on mass. The scalings are shown as dashed lines.

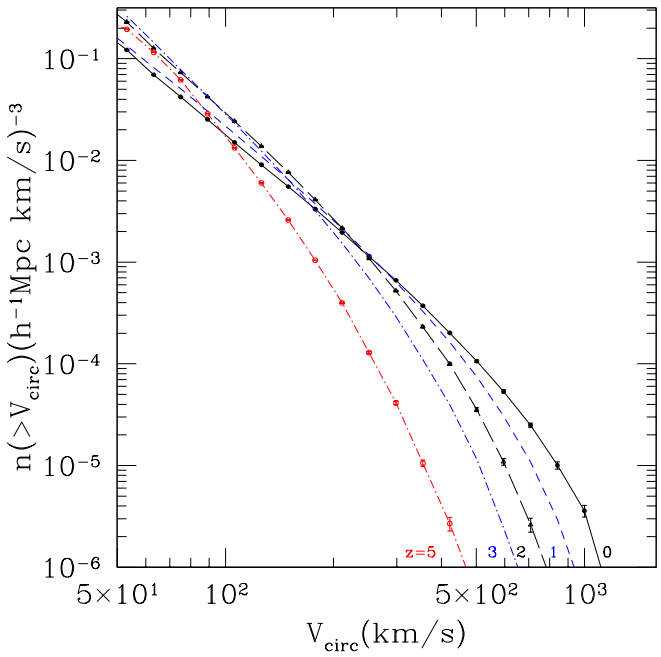



Figure 6. The cumulative velocity function of distinct halos at different redshifts in the *Bolshoi* simulation (adopted from [121]).

### 7.2 *Bolshoi simulation*

The *Bolshoi* simulation, described in [121], is a high-resolution and large-volume run performed with the WMAP5 and WMAP7 cosmological parameters [123, 124], $\Omega_{\rm M} = 0.27$, $h = 0.7$, and $\sigma_8 = 0.82$. The simulation contained $2048^3 \approx 8$ billion dark matter particles in a $250\,h^{-1}$ Mpc box. The corresponding mass and force resolutions are $m_p = 1.35 \times 10^8\,h^{-1}\,M_\odot$ (one particle mass) and $1.0\,h^{-1}$ kpc (the smallest cell size in physical coordinates), respectively. The simulation outputs were recored at 180 time steps. Both the main simulation code and the procedure for identification of the DM halos are similar to those used for the $\Lambda$CDM$_{120}$ run. The completeness limit for the halo catalogs derived from the *Bolshoi* outputs is $v_{\rm max} = 50$ km/s or $M_{\rm vir} \approx 1.5 \times 10^{10}\,h^{-1}\,M_\odot$.

The velocity function for distinct DM halos in the *Bolshoi* simulation is shown in Fig. 6 (adopted from [121]). The number density of halos is a very steep function of velocity. For small halos, $N \propto V^{-3}$, and there is an exponential cutoff at high velocities. Klypin et al. show that the velocity function shape is of this type at all redshifts, and it can be approximated as $n(>V) = AV^{-3}\exp(-[V/V_0]^\alpha)$, where the parameters $A$, $V_0$, $\alpha$ are the functions of the amplitude of perturbations

$\sigma_8(z)$. It is clear from Fig. 6 that the evolution of the halo number densities for $v_{\rm max} < 350$ km/s is weak at $z < 3$. We use this fact to simplify the analysis in Chapter IV. At all velocities probed in the simulation, the contribution of subhalos to the total number density of bound objects is small, $< 30\%$ (see Fig. 14 in [121]).[5]

## 8 Correlation function of distinct dark matter halos

Even though the formation of virialized DM halos is a highly non-linear process, the theoretical framework for computing the correlation function of distinct halos is fully developed and calibrated / verified by the numerical simulations. Here, we provide a brief summary.

The correlation function of distinct (those not contained in larger halos) objects with mass $M$ at a given redshift, $z$, can be written as

$$\xi(r, M) = b^2(M)\, \xi_{\rm DM}(r), \tag{3.3}$$

where $b(M)$ is the bias factor for halos with mass $M$, and $\xi_{\rm DM}(r)$ is the linear correlation function of the dark matter density field.

The $\xi_{\rm DM}(r)$ term can be computed from the linear theory of the cosmological density perturbations at any given redshift. Specifically, its shape is related to the power spectrum of density perturbations [184], and the amplitude is proportional to the square of the amplitude of linearly-evolved density perturbations,

$$\xi_{\rm DM}(r) \propto \sigma_8^2(z) = \sigma_{8,0}^2 D(z)^2 \tag{3.4}$$

where $D(z)$ is the linear perturbations growth factor for the given cosmology.

The second term, $b(M)$, is responsible for the mass-dependence of the correlation function DM halos. The bias factor can be derived only by tracking the non-linear collapse of the growing density perturbations either numerically or analytically under simplifying assumptions (for example, Sheth & Tormen [223] find a very accurate fitting formula by considering the collapse of ellipsoidal

[5] This does not contradict the fact that each halo has many subhalos because those subhalos typically have much lower masses/velocities than the host.

density perturbations). These studies show that $b(M)$ can be expressed as a function of the linear perturbation amplitude at the mass scale $M$, $b(M) = b(\sigma(M))$ [111, 165, 223], and $b$ as a function of $\sigma$ has an almost universal expression, i.e., a single fitting function applicable to a wide range of redshifts and cosmological parameters. For example, the Sheth & Tormen approximation [223] can be written as

$$b = 1 + \frac{a\,\delta_c^2/\sigma^2 - 1}{\delta_c} + \frac{2p}{\delta_c\left[1 + (a\,\delta_c^2/\sigma^2)^p\right]} \tag{3.5}$$

where the parameters are $a = 0.75$ and $p = 0.3$ (adopted from [108], see their page 704), and $\delta_c = 1.69$ is the threshold for spherical collapse in a matter-dominated universe.

For "rare" (massive) objects, for which $\sigma(M) \sim 1$ or less, this equation implies that the bias factor is high and strongly dependent on $\sigma(M)$. The astronomical sources considered in this thesis reside in halos with $M = 10^{12} - 10^{13}\,M_\odot$, which at $z \gtrsim 1$ makes them belong to the category of "rare" objects. In this mass range, $b(M)$ scales approximately as $M^{0.2-0.25}$ (see, e.g., Appendix A in [233]). It is because of this dependence one can use the correlation length derived for a given source population to estimate the typical mass scales for the DM halos associated with this population. This effect provides the basis for the analyses presented in Chapters IV–VI.

Even though the theory for the correlation function of distinct DM halos can be considered as fully developed, its application to *galaxies* requires additional steps, mainly because a significant fraction of galaxies reside in the satellite DM subhalos contained in even large halos. The current framework for connecting DM (sub)halos and galaxies is discussed in the next section.

## 9 Statistical models of the galaxy populations

It was shown by numerous observations that the correlation function of galaxies depends on their luminosity, morphology, color, and spectral type (e.g., [149], [236], [139], [238], among others). These trends are hard to explain from the first principles since the physics of galaxy formation is complicated and current models still do not properly include all relevant processes. Nevertheless, the

standard cold dark matter paradigm robustly predicts that galaxies form at the centers of virialized dark matter halos [258, 71, 25] and it was confirmed by many studies (see [13, 242, 237], among others) that the gravitational dynamics alone may explain the basic features of galaxy clustering. Therefore, to model the galaxy clustering in most cases we need simply to establish a link between the galaxy population and a specific set of DM halos.

A very useful statistical framework which serves this purpose and allows one to predict and interpret the clustering properties of galaxies is the so-called halo model of galaxy clustering. In this framework, a relation between galaxies and their host DM halos is specified through the halo occupation distribution (HOD) — the probability, $P(N|M)$, for a halo of virial mass $M$ to contain $N$ galaxies of a given type, along with additional prescriptions describing the distribution of galaxies within the halos. Another approach which is used to link the galaxies and DM halos is the so called "abundance matching" technique, useful when one can expect that there is a tight correlation between the galaxy properties and the mass of the host halo. Below, we describe in more detail the HOD-based modeling which can be implemented through analytical and numerical approaches, and then we briefly discuss the abundance matching technique.

### *9.1 Analytical HOD modeling*

Analytical methods of HOD modeling (see [218], [144], [183], [257], [216], [12], [13], among others) are based on the properties of DM halos that were extensively studied using $N$-body simulations, and on the parameterized functional form of $P(N|M)$ specifying a galaxy distribution within DM halos. The analytical approximations of the halo mass function, the linear bias of halos as a function of halo mass, the radial density profile of halos as a function of halo mass were calibrated against $N$-body simulations. The probability distribution $P(N|M)$ is defined by its first moment $\langle N\rangle_M = \sum_N NP(N|M)$ which gives the mean number of central or satellite galaxies per a halo.

For galaxy samples defined by a minimum luminosity threshold, the mean halo occupation $\langle N\rangle_M$ is usually assumed to be a power law at high halo masses with a cutoff at some mass, $M_{\rm min}$, which corresponds to minimum mass below which the DM halos can not host galaxies of this type. Such a parameterization

is supported by both theoretical models ([13] and references therein) and by fitting the observational data (e.g., [216], [197],[136], [152], [201], etc). At low halo masses, $\langle N \rangle_M$ is expected to reach a plateau $\langle N \rangle \sim 1$, where each halo contains on average only one galaxy, which is also cut off below a minimum mass threshold.

An alternative parameterization is based on the idea that there exist two separate galaxy populations — the central galaxies of which there can be zero or one per halo, and satellite galaxies within halos. These populations can then be parameterized separately [237, 265, 48, 35, 266, 240], improving the accuracy of the overall model. In particular, the probability for a halo of mass $M$ to host $N$ galaxies, $P(N|M)$, is split into the probability to host one central galaxy $P_c(M)$ and the probability of $N_s$ satellite galaxies $P_s(N_s|M) \equiv P(N_s+1|M)$. In the most simple case, the HOD of central galaxies can be modeled as a step function $\langle N_c \rangle_M = 1$ with $\langle N_c \rangle_M = 0$ for $M < M_{\rm min}$, while the mean number of satellite galaxies per a massive halo can be modeled as a power law, $\langle N_s \rangle_M = (M/M_1)^\alpha$, where $M_1$ is the limiting mass of halos harboring $\simeq 1$ satellite. The motivation for the separation of central and satellites galaxies comes partly from the analysis of hydrodynamic simulations [13], and partly from studies of central bright elliptical galaxies in groups and clusters, which are often considered as a separate population from the rest of galaxies [127].

Figure 7 shows the HOD for DM halos derived from a numerical simulation, illustrating the parameterizations discussed above. The premise of the HOD modeling is that the same parameterization can be applied to modeling the observations of galaxies, and by treating the quantities $M_{\rm min}$, $M_1$, $\alpha$ as free parameters and by deriving them from fitting the data we can learn about the DM halos associated with the galaxies of a given type. These parameters can be constrained by comparison of the observed and model luminosity, mass, or correlation functions of galaxies. Particularly, within the HOD formalism, the two-point correlation functions is considered to be a sum of two terms (Fig. 8):

$$\xi(r) = \xi_{\rm 1h}(r) + \xi_{\rm 2h}(r). \tag{3.6}$$

The one-halo term, $\xi_{\rm 1h}$, corresponds to pairs of galaxies residing within the same halo, i.e. the satellite-satellite and satellite-central galaxy pairs. This term

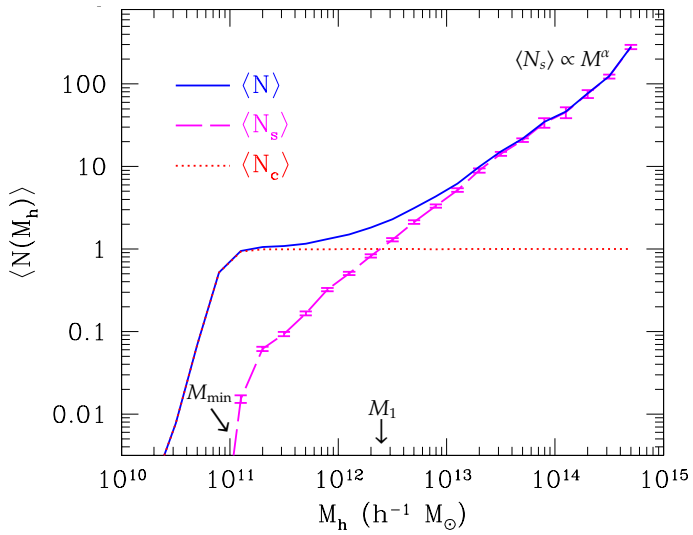



Figure 7. The first moment of the halo occupation distribution, as a function of host mass for the halo sample from ΛCDM simulation at $z = 0$ [127]. The solid line shows the mean total number of halos including the hosts, while the long-dashed line represents the mean number of satellite halos. The error bars show the uncertainty in the mean. The dotted line shows the step function corresponding to the mean number of "central" halos. The halo sample was selected using a threshold $V_{\rm min}$, while the HOD is plotted as a function of halo mass. Due to the intrinsic scatter between maximum circular velocity and halo mass, there is a gradual transition from $\langle N_c \rangle = 0$ to $\langle N_c \rangle = 1$. Note that by definition, the solid line is the sum of the dotted and long-dashed lines.

is dominant at small scales and its amplitude is proportional to the second moment of the $P(N|M)$, $\langle N(N-1)\rangle_M$ distribution. Therefore, the "1-halo" term is sensitive to the fraction of galaxies that are satellites in the more massive objects [136, 201, 50].

At scales larger than the virial diameter of the largest halos, all pairs consist of galaxies in separate halos and only the two-halo term, $\xi_{\rm 2h}$, contributes to the total correlation function (Fig. 8). In the analytical approximation, the "2-halo" terms depends on the average number of galaxies per halo of a given mass, $\langle N \rangle_M$.

Neither 1- nor 2-halo correlation function components are not power law functions of $r$. Therefore, it is expected that the correlation function of galaxies should show deviations from a power law when observed over a sufficiently broad range of scales. Indeed, the deviations of the correlation function from a power-law were recently detected by Zehavi et al. [263] in the correlation function of SDSS galaxies. The HOD modeling shows that the power-law break occurs at scale of transition between two- and one-halo terms.

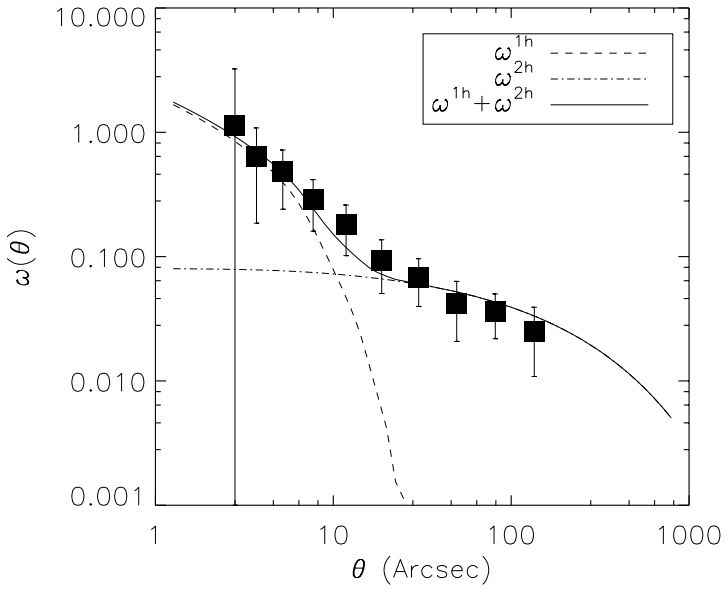



Figure 8. The example of two-component fit to the observed angular correlation function. The dashed line shows the one-halo term, and the dash-dotted line shows the two-halo term of the total $w(\theta)$. The solid line is the sum of the two contributions [136].

Several aspects of the HOD model have been studied using semianalytic galaxy formation models and cosmological hydrodynamical simulations (see references in [127]). Berlind et al. [13] presented a detailed comparison of the HOD in these two types of simulations. They find that, despite radically different treatments of the cooling, star formation, and stellar feedback in the two approaches to galaxy formation modeling, for galaxy samples of the same space density the predicted HODs are in almost perfect agreement. This result supports the idea that the HOD, and hence galaxy clustering, is driven primarily by gravitational dynamics rather than by processes such as cooling and star formation [127].

The analytical HOD formalism has been successfully applied to the clustering measurements in a number of surveys of low-redshift galaxies (e.g., [151], [188], [35], [266], [238], [208]) and high redshift sources (e.g., [97], [136], [152], [150], [201], [250], [50]). Such analyses are typically applied to the projected (§ 3) or angular correlation functions of galaxies (§ 4). The halo model was developed further by Tinker [239] to include to additional sources of information such as the redshift-space distortions (see also earlier works by Ballinger et al. [7] and Seljak [219]).

### *9.2 Numerical HOD modeling*

Numerical simulations in effect directly provide the HOD distribution for dark matter halos and subhalos, and one can use that for modeling a given galaxy sample. Significant pieces of the analytic HOD formalism are, in fact, inspired or calibrated by the results of numerical simulations. Typically, the more detailed the analysis, the stronger its reliance on the simulations. For example, in the analysis of the projected (or angular) correlation functions, one uses the numerically calibrated fitting functions for the density profiles of DM halos. If the analysis also includes modeling of the redshift-space distortions, one additionally includes the calibrations for the subhalo velocity bias and for segregation of massive subhalos towards the center of their host halo [239]. Furthermore, the properties of certain classes of sources may depend on the recent merging history of their host (sub)halos. Using numerical simulations, it is possible to follow the formation history of halos (see, e.g., [96] and references therein), but it is almost impossible to perform a similar analysis analytically.

Since the HOD modeling relies heavily on the calibrations provided by the numerical simulations, one can pose a question of whether it is worthwhile to bypass the analytic modeling entirely and simply "populate" the DM halos identified in the simulation outputs with galaxies. Such an analysis is certainly very easy from the technical point of view, if the algorithm for associating the DM (sub)halos with galaxies is specified. Early studies have shown (e.g., [48]) that very good results are achieved with an extremely simple algorithm when one associates the galaxies above some luminosity threshold with the DM halos above a certain mass or circular velocity threshold. This is the approach we use in our work, when we study the connection of the *Spitzer* and *Chandra* sources with their host dark matter halos using their clustering data.

### *9.3 Abundance matching*

An approach closely related in spirit to the HOD modeling has become popular in the recent years, especially in the analyses based on directly populating the numerical simulation outputs with galaxies (e.g. [247], [48], [96], [170]). This approach is based on the fact that in many cases, we have physical reasons to

expect that some galaxy parameters (e.g., stellar mass, or the red luminosity) have a tight, unique relation with the mass of the host halo [127], [237], [246]. Then, because the mass or velocity function of the DM halos is very steep, the mass scale of the parent DM halos can be estimated from the observed number density of objects.

In fact, this "abundance matching" technique allows one to reconstruct a relation between the luminosity or stellar mass, and the total mass or circular velocity of the halo. Specifically, they are tied through equations of the type:

$$n_g(> L_i) = n_h(> v_{\mathrm{max},i}), \quad \text{or} \tag{3.7}$$

$$n_g(> M^*) = n_h(> M_{\mathrm{halo}}), \tag{3.8}$$

(see Conroy et al. [48] and Guo et al. [96]). The $L_i$, $M^*$ in this equation are the luminosity and mass thresholds of a galaxy sample, and $v_{\mathrm{max},i} = V_{\mathrm{min}}$ and $M_{\mathrm{halo}}$ are the threshold characteristics of DM halos selected from simulations. It has been shown that despite simplicity and using only one observational input, the luminosity/stellar mass distributions, the abundance matching approach allows to reproduce a number of characteristics of the galaxy populations over a wide range of redshifts [127, 237, 246, 242, 247, 48, 156, 15, 47, 121, 240, 96, 170].

We apply abundance matching in the analyses presented in Chapters IV and V to put constraints on the connection of IR-selected galaxies and DM halos.

# IV Star-forming galaxies detected in mid-infrared

## 10 Summary

In this Chapter we discuss the clustering properties of galaxies with signs of ongoing star formation detected by the *Spitzer*/MIPS camera in the SWIRE Lockman Hole field. The sample of mid-IR selected galaxies includes $\sim 20000$ objects detected above a flux threshold of $S_\nu(24\,\mu\mathrm{m}) = 310\,\mu\mathrm{Jy}$. We adopt optical/near-IR color selection criteria to select from the sample lower-redshift ($z_{\mathrm{mean}} = 0.68$) and higher-redshift ($z_{\mathrm{mean}} = 1.72$) galaxy populations. We measure the angular correlation function on scales of $\theta = 0.01 - 1\,\mathrm{deg}$, from which using the Limber inversion we obtain comoving correlation lengths of $r_0 = 4.74 \pm 0.16\,h^{-1}\,\mathrm{Mpc}$ and $r_0 = 7.87 \pm 0.63\,h^{-1}\,\mathrm{Mpc}$ for the low-$z$ and high-$z$ subsamples, respectively. Comparing these measurements with the correlation functions of dark matter halos identified in the *Bolshoi* cosmological simulation [121], we find that the high-redshift objects reside in progressively more massive halos reaching $M_{\mathrm{tot}} \gtrsim 2 \times 10^{12}\,h^{-1}\,M_\odot$ for the high-$z$ population, compared to $M_{\mathrm{tot}} \gtrsim 5 \times 10^{11}\,h^{-1}\,M_\odot$ at the lower redshifts. Approximate estimates of the IR luminosities based on the catalogues of 24 $\mu$m sources in the GOODS fields [207] show that our high-$z$ subsample represents a population of "distant ultra-luminous infrared galaxies" (ULIRGs) with $L_{\mathrm{IR}} > 10^{12} L_\odot$, while the low-$z$ subsample mainly consists of "LIRGs", $L_{\mathrm{IR}} \sim 10^{11} L_\odot$. The comparison of number density of the 24 $\mu$m selected galaxies and of dark matter halos with derived minimum mass $M_{\mathrm{tot}}$ shows that only 20% of such halos may host star-forming galaxies.

## 11 Introduction

The discovery of the extragalactic infrared background [199, 99] revealed that half of the total extragalactic background energy averaged over cosmic time and wavelengths emits in the infrared (IR). This indicate that half of the star formation and active galactic nucleus (AGN) emission is obscured by dust. In the local Universe, the volume density of dusty, IR-luminous galaxies is insufficient to account for this background. Therefore, the density of such galaxies must evolve strongly with redshift [100, 133]. In fact, observations with the Infrared Space Observatory (ISO) and the Spitzer Space Telescope (hereafter *Spitzer*) revealed that more than 70% of the total IR luminosity density at redshifts $0.5 < z < 3$ is contributed by mid- and far-infrared bright galaxy populations [42, 135]. The mid-IR and far-IR spectral regimes are direct probes of hidden by dust activities in galaxies 9. Optical and UV radiant energy from young stars is absorbed by warm and cold dust and re-emitted at long wavelengths. But not only star formation contribute to the mid- and far-infrared galaxy emission. Particularly, at $z >\sim 1$, the *Spitzer* MIPS 24 $\mu$m band (§ 12) detects light (rest-frame $5 - 12$ $\mu$m) which can be decomposed into polycyclic aromatic hydrocarbon (PAH) and power- law (or warm blackbody) emissions (see for details and references [194]). The first is understood to be powered entirely by star formation, while the second predominantly comes from the accretion disks of AGNs [88]. The AGN and star-forming activities in galaxies are believed to be triggered by mergers or interactions with other galaxies, and majority of such events occur at redshift around 2 (e.g. [104], [147], [227]). Therefore, studying of the distant Universe in the infrared provides valuable information on the origin and assembly of present-day massive galaxies (e.g. [65], [212]).

In the work described in this chapter, we use observations of star-forming galaxies made by *Spitzer* at 24 $\mu$m. The *Spitzer* 24 $\mu$m surveys actually uncovered the population of "distant ULIRGs" — ultraluminous infrared galaxies, previously known in the local Universe. They are dusty star-forming galaxies with infrared luminosity $L_{IR} > 10^{12} L_{\odot}$[1] (e.g. [205],[259], [55]). While the average spec-

[1] $L_{IR} = \int_{8\,\mu\mathrm{m}}^{1000\,\mu\mathrm{m}} L_{\nu} d\nu$, [214]

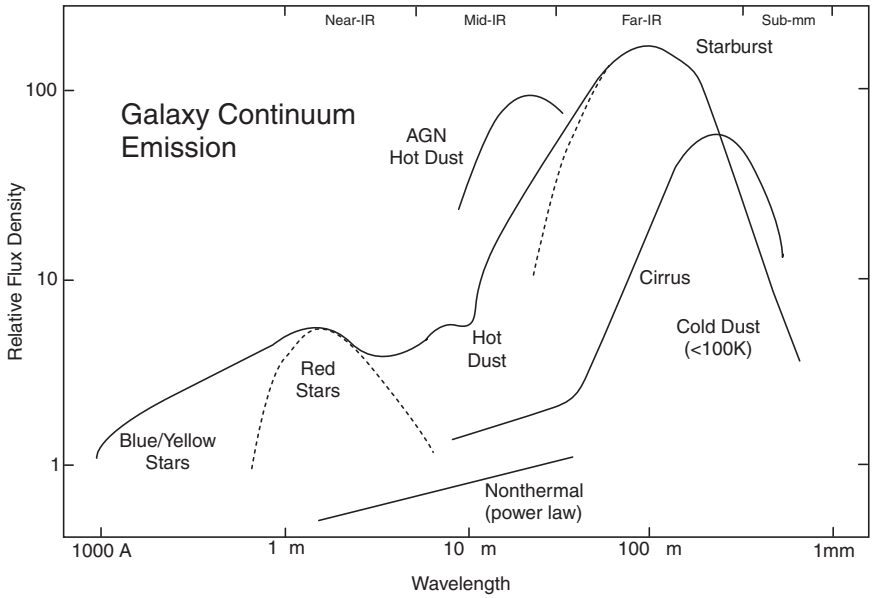



Figure 9. Spectral Energy Distribution of a starburst galaxy undergoing an intense episode of star formation. The spectrum of cirrus emission from interstellar heated dust is also presented together with two types of SED due to AGN activity: non-thermal emission (power-law) and emission due to the dust heated by the central engine. It is possible for a galaxy to show at the same time emission contributed by star formation, cirrus and AGN. (From M. Vaccari, PhD thesis, 2004.)

tral energy distributions of high-$z$ sources is consistent with that of present-day ULIRGs, the nature of the former must be still clarify[2](see [109] for details and references). Further, different photometric techniques are applied to identify high-redshift objects among thousands detected by wide-field *Spitzer* surveys, e.g. Yan et al. [259], Magliocchetti et al. [152], Farrah et al. [73], Lonsdale et al. [142], Fiolet et al. [76], Huang et al. [109], Dey et al. [64]. All these selected classes represent subpopulations of ULIRGs with observational characteristics overlapped with star-forming galaxies detected in optic and submillimetre (see recent papers by Huang et al. [109] and Fiolet et al. [76]). Theoretical models of galaxy formation and evolution are intended to connect these populations based on their physical properties, such as SEDs, SFRs, stellar and halo masses, etc (e.g. [94], [56], [173], [131]). A unique tool to estimate halo masses of galaxies at high redshifts is measurements of clustering amplitude (see Chapter III, § 8 for details). The goal of the following chapter is to present clustering analysis of

[2]Most local ULIRGs have disturbed morphologies, consistent with being merging systems.

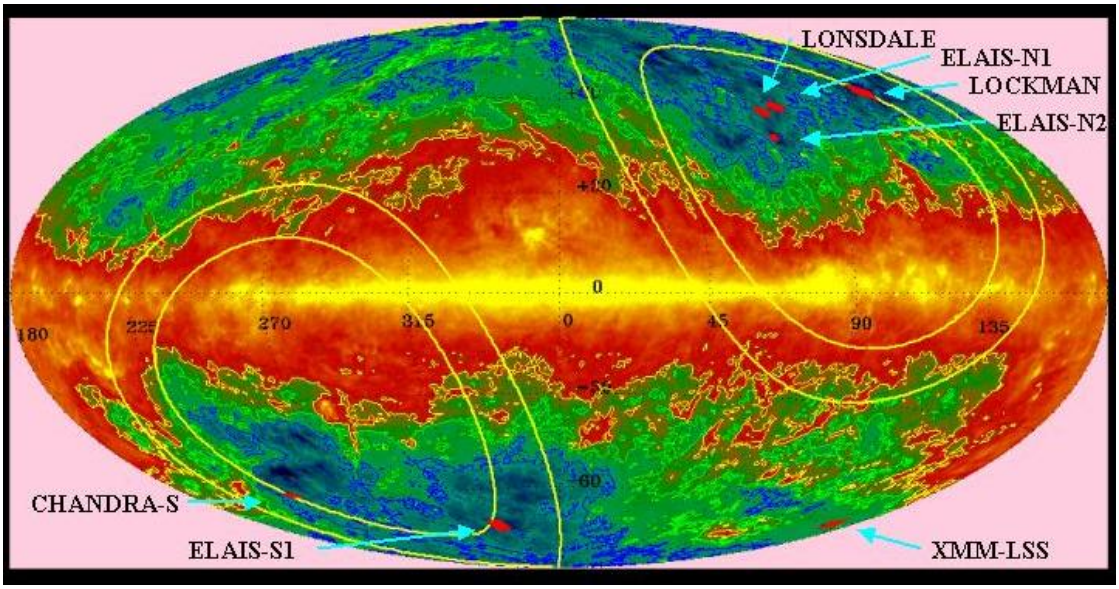



Figure 10. Hammer-Aitoff equal area projection in galactic coordinates of the 100 $\mu$m sky. SWIRE fields are shown in red. The contour levels in blue, green and yellow are 1, 2, 4 MJy/sr, respectively. The yellow ellipses mark ecliptic latitudes of 30° and 40°. (From SWIRE web-page: http://swire.ipac.caltech.edu/swire/swire.html.)

star-forming 24 $\mu$m-detected galaxies from one of the largest *Spitzer* extragalactic survey.

## 12 SWIRE survey

The *Spitzer* Wide-area InfraRed Extragalactic Survey (SWIRE, [143]) is the largest survey carried out with the Spitzer Space Telescope [143]. It covered 49 square degrees in six separate fields in the Northern and Southern sky 10 — the Lockman Hole (11.1 deg$^2$)[3], ELAIS-N1 (9.3 deg$^2$), ELAIS-N2 (4.2 deg$^2$), ELAIS-S1 (6.8 deg$^2$), XMM-LSS (9.1 deg$^2$), and Chandra-S (6.8 deg$^2$). The fields were selected so that they are minimally contaminated by the infrared background and bright objects, and covered by substantial datasets at other wavelength. Such a broad range of resources made the SWIRE a unique survey for studying properties and distribution of high-redshift star-forming and passively evolving IR galaxies, to carry out a galaxy population analyses requiring excellent statistics, and to search for rare objects [141].

[3]The quoted areas represent the coverage in all bands of the IRAC instrument

Each field was imaged in the seven near-to-far infrared bands: InfraRed Array Camera (IRAC) 3.6, 4.5, 5.8, 8.0 $\mu$m [74] and Multiband Imaging Photometer for Spitzer (MIPS) 24, 70, and 160 $\mu$m bands [204]. IRAC was designed to detect the 1.6 $\mu$m peak in the galaxy spectral energy distribution (SED) formed by emission of low mass or cold evolved stars due to the minimum of the $H^-$ opacity in their atmospheres [74]. This peak is a relatively sharp feature in the galaxy SED and thus can be used to determine the photometric redshifts in addition to the total amount of low mass and evolved stars which are the main contributors to the stellar mass of a galaxy. The emission in far infrared (FIR) is mostly attributed to the dust-obscured star formation or AGN activity in galaxies. Optical and UV light from stars or non-thermal emission from AGNs heat the dust which re-emits this energy in the FIR, making the MIPS flux one of the best tracers of on-going activity in galaxies.

The clustering analysis described in the thesis was carried out for IRAC and MIPS data in the two SWIRE fields — Lockman Hole (Chapter IV, § 14) and ELAIS S1 (Chapter V).

## 13 Catalogues of 24 $\mu$m sources

For reliable clustering measurements one needs a statistically complete, large, and homogeneous sample of sources selected over a large area of the sky to probe the correlation signal on a wide range of scales. The SWIRE survey is highly suitable for this purpose as was demonstrated in several papers [254, 61, 72]. In addition to the near- and mid-infrared observations, every SWIRE field has high-quality ancillary data making it possible to study clustering of various galaxy populations.

The goal of our work was to estimate the correlation function of star-forming galaxies detected in the MIPS 24 $\mu$m band. At first, we selected a sample of bright, $S_\nu(24) > 400\,\mu$Jy, sources from the SWIRE ELAIS-S1 catalogue (Vaccari et al., Data Fusion 2010), with the goal of comparison of clustering properties of active and passively evolving galaxies (the analysis for the later is in Chapter V below) observed in the same field. However, the angular correlation function, $w(\theta)$, for these sources unexpectedly showed a lack of clustering signal at scales

$\theta < 36''$. There were suggestions in the literature (e.g., Gilli et al. [90]) that because of the poor angular resolution of the MIPS instrument ($\sim 6''$ FWHM), there could be some difficulties in determining $w(\theta)$ for *faint* sources due to blending. However, the deficit of close pairs in the sample of *bright* sources was unexplained in the literature. To explore the origin of this peculiarity, we carried out a comparison of the angular correlation function of the 24 $\mu$m sources selected from the four largest SWIRE fields (Lockman Hole, ELAIS-N1, ELAIS-N2 and CDFS) using two releases of the SWIRE team catalogues (versions 2005 and 2010), and an additional source catalog based on the wavelet decomposition algorithm (see below). The results of this comparison are reported in the next sections before we turn to the discussion of the clustering properties of 24 $\mu$m sources using the best available catalog.

### *13.1 On comparison of clustering of 24 $\mu$m sources selected from different catalogs in four SWIRE fields*

As a first data set, we used publicly available catalogues from the SWIRE Data Release 2 (version 2005)[4]. These catalogues consist of the optical, IRAC, and MIPS 24 $\mu$m information merged into a single table for sources detected in the IRAC 3.6 and 4.5 $\mu$m bands above pre-defined SNR thresholds. The source detection in the MIPS data was carried out using *SExtractor* [21]. The estimated completeness threshold is $\sim 400\,\mu$Jy in all fields. For the clustering analysis, we selected all 24 $\mu$m sources above this flux threshold. To eliminate the Galactic stars (see below, p. 44), we cross-correlated this set of 24 $\mu$m sources with the objects in the 2MASS survey using a matching radius of 2.5''. Hereinafter, we refer to these source catalogs (with stars eliminated) as the "2005-catalogue" or "v.2005".

The second set of catalogues was produced from the SWIRE Final Data Release (2009) by M. Vaccari, L. Marchetti and E. Gonzales-Solares. These authors compiled the multi-band information available in the SWIRE fields (called "Data Fusion") for sources detected in the IRAC 3.6 $\mu$m and 4.5 $\mu$m bands [Vaccari et al., in preparation]. For the IRAC images, the source detection

[4] Available at `http://swire.ipac.caltech.edu/swire`.

was again done using *SExtractor*, while the *MOPEX* package [154] was used for MIPS data. The *MOPEX* package was specifically optimized for detection of point-like sources in crowded fields, and its application results in a significant improvement in the completeness limit for MIPS data, which can be as low as $\sim 150\,\mu$Jy (see below). The completeness of the IRAC detections was also improved compared to the previous data release. In order to avoid source confusion and false identification in the 24 $\mu$m band, Vaccari et al. matched 24 $\mu$m and IRAC sources within a radius of 1.5″. The data from other catalogues (e.g., the 2MASS PSC) were associated with the initial IRAC source list using a matching radius of 2.5″. For our analysis, we used all sources with the 24 $\mu$m detections, and this selected sample is referred to as the "2010-catalogues" or "v.2010".

Another significant difference between the 2005 and 2010 catalogues is in the methods of flux measurements for the MIPS sources. The 2005 data release used aperture photometry using a set of apertures $7.5'' - 15''$ radius, which contained $60 - 85\%$ of the total flux. The *MOPEX* package provides the total fluxes provided the PRF fitting. This is significant in our case because the aperture and PRF fitting photometry have different problems in dealing with the close source pairs, which can produce different results for the small-scale clustering.

As an independent test, an additional list of 24 $\mu$m sources was created [A. Vikhlinin, private communicaton] using the wavelet decomposition source detection algorithm [253] applied to the publicly available MIPS images. This method was designed to efficiently detect both point-like and slightly extended sources in the crowded fields. Originally, the wavelet decomposition program was intended for Poisson-noise limited X-ray images, where it generally outperforms its rivals [202], but it was found that with a suitable choice of parameters, it produces good results also for the 24 $\mu$m MIPS images. The main output of the wavelet decomposition algorithm is a list of source location detected above a predefined SNR threshold (which we set to $4.5\sigma$), and a map which allows one to split the original image into "empty" regions and those with significant emission "belonging" to a particular source. The source fluxes were then measured using aperture photometry. In choosing the aperture size, the tradeoff

is between our desire to include as much of the source flux as possible into the aperture size, and the fact that for wide apertures, the flux measurements are increasingly affected by the larger-scale background fluctuations and by the source confusion. Several tests have shown that the best results are achieved for an aperture size of 4″, encompassing approximately 50% of the PRF power, and corresponding to the bright core of the MIPS PRF. These aperture fluxes were then converted into total flux using the PRF model calibrated by bright stars. Using this method, the 24 $\mu$m sources were extracted in the MIPS observations of the Lockman Hole and ELAIS-S1 fields, and we cross-correlated them with the IRAC sources from the Vaccari catalogues [Vaccari et al., in preparation] using a matching radius radius of 3.2″. This third data set is referred to as the ″A1-catalogues″ below.

*Star-galaxy separation*

Since we are interested in the clustering properties of the extragalactic sources, the Galactic stars act as a contaminant in our analysis, and they should be removed from the catalogues. We followed the methods of Shupe et al. [224] and Waddington et al. [254] in which the foreground stars were identified using the 2MASS Point Source Catalogue (PSC) [230].[5]

All 24 $\mu$m-IRAC catalogues were cross correlated with the 2MASS survey using a matching radius of 2.5″. Figure 11 shows the color criteria established by Shupe et al. [224] to separate out the stars among the 24 $\mu$m-emitting sources. These authors proposed that most of the sources with $K_s - [24] < 2.0$ (Vega, mag) are Galactic stars. We applied this criterion to our samples and found that the stars contribute $\sim 2\%$ to the total number of sources detected in the 24 $\mu$m band SWIRE images.

*Stellar mask*

In addition to directly polluting the extragalactic samples, bright Galactic stars may affect our clustering measurements indirectly, by obscuring the background

[5]We note, however, that the star removal is not a crucial component of our analysis since the contamination of near- to mid-IR galaxy samples by foreground stars is a severe problem only at fluxes of brighter than several mJy. Moreover, our final 24 $\mu$m source sample will be limited in flux from above in both the 24 $\mu$m and IRAC (3.6, 4.5 $\mu$m) bands.

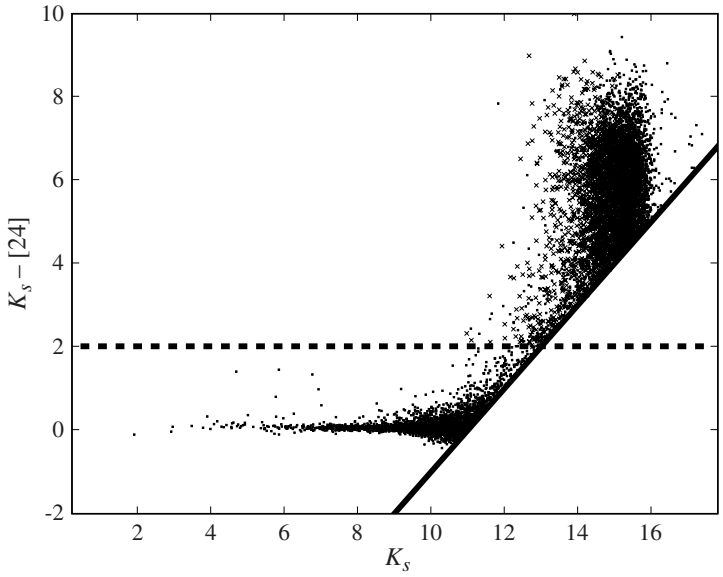



Figure 11. The color-magnitude diagram of 2MASS $K_s$ and [24] for all six SWIRE field, used by Shupe et al. [224] to separate out the stellar contribution to the 24 $\mu$m sources population. Most of the sources below the dashed line are the Galactic stars. The solid line corresponds to a 24 $\mu$m flux limit of 300 $\mu$Jy. Small crosses show the objects marked as extended in the 2MASS catalog.

galaxies or affecting the fluxes of the fainter galaxies near the same line of sight. Therefore, we need to completly exclude from the analysis the sky regions affected by the bright foreground stars. Following Waddington et al. [254] this was achived by masking out the circular regions around sources with $K_s < 12$ (Vega, mag) from the correlated 24 $\mu$m-IRAC-2MASS catalogues; the exclusion radius was determined as $\log(R'') = 3.1 - 0.16\,K_s$, which is the distance at which the stellar PSF merges into the background [254]. An example of the mask created with this algorithm is shown in Fig. 12).

A close examination of the 24 $\mu$m source catalogs shows that sometimes, there are spurious detections around very bright 24 $\mu$m sources (most of which correspond to Galactic stars or low-$z$ galaxies) [A. Vikhlinin, private communication]. Therefore, we decided to mask those regions as well. The exclusion radius was set to be $20'' - 80''$, depending on the source flux.

These region masks excluding the regions around bright Galactic stars and extremely bright 24 $\mu$m sources were used in the estimation of angular correlation function (§ 4 below).

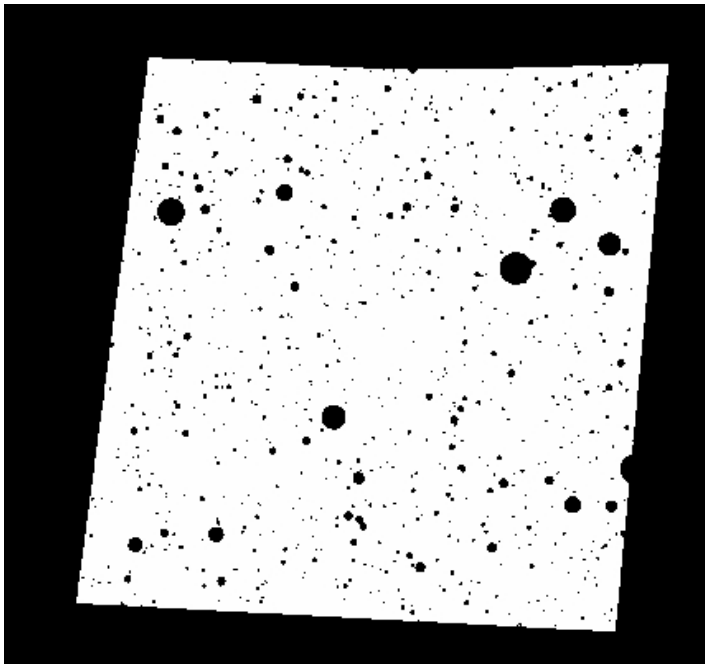


Figure 12. ELAIS-N1 mask: black circles, excluded from the analysis, mark regions around stars.

*Limiting fluxes for individual catalogues*

For a proper comparison of the angular correlation function between different versions of the source catalogues and different fields, we have to make sure that the sources are selected above a flux which exceeds completeness limits for each field/catalogue. Ideally, a completeness limits is a flux threshold above which (nearly) all real sources are detected and into which (almost) no fainter sources are migrated. The exact completeness limit for the MIPS/SWIRE data can be established only through exhausting Monte-Carlo simulations (e.g. [224]). However, we can apply a useful empirical criterion and identify the completeness limit with a point of maximum in the differential $\log N - \log S$ distribution observed for each field/catalogue.

In Fig. 13, we show the number of sources per square degree and the logarithmic flux bin contained in the 2010-catalogues for different SWIRE fields. The maxima in the differential $\log N - \log S$ distribution in all cases are achieved near a flux of $\sim 200\,\mu$Jy. However, there are clear differences in the number counts of faint sources up to a flux limit of $S_\nu(24) \sim 350\,\mu$Jy. This probably indicates a flux measurement uncertainty of $\sim 100\,\mu$Jy, which may explain also why the drop in the differential $\log N - \log S$ distribution below the point of maximum

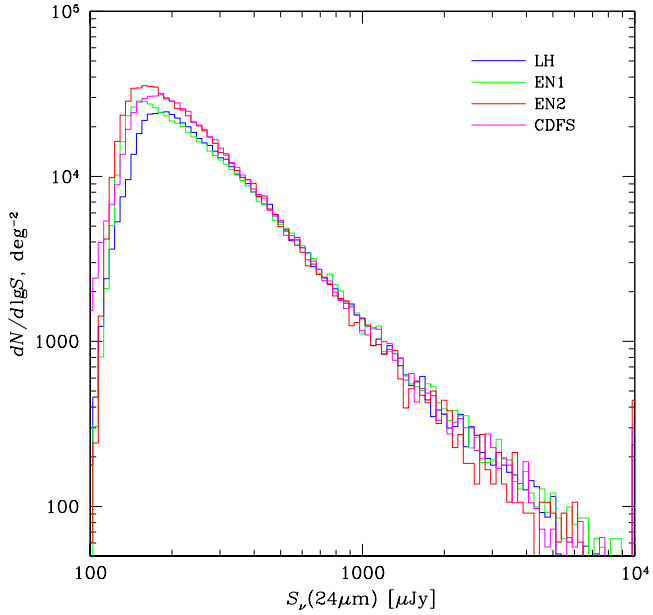



Figure 13. The number of 24 $\mu$m sources per sq. degree per log flux interval versus the logarithmic flux. The sources were selected from the 2010-catalogues [Vaccari et al., in preparation] in the four SWIRE fields — Lockman Hole (blue), ELAIS-N1 (green), ELAIS-N2 (red), and CDFS (magenta). Galactic stars were masked out and eliminated.

Table IV.1. The fluxes of completeness of selected samples in the SWIRE fields and areas of the fields used for clustering analysis

| Field | $S_{\rm comp}$ [$\mu$Jy] | Area [deg$^2$] |
|---|---|---|
| Lockman Hole | $S^{2010} = 180$<br>$S^{2005} = 400$<br>$S^{\rm A1} = 310$ | 8.7 |
| ELAIS-N1 | 160 | 7.1 |
| ELAIS-N2 | 170 | 3.3 |
| CDFS | 180 | 6.2 |
| ELAIS-S1 | $S^{\rm A1} = 400$ | 6.3 |

is not sharp but extends to $\sim 100\,\mu$Jy. Therefore, based on examination of the $\log N - \log S$ distributions, the correlation functions for the 2010-catalogue in different SWIRE fields should be compared for sources brighter than $350\,\mu$Jy.

In Fig. 14, we show the source counts for the three different catalogues in the Lockman Hole field. There is a striking difference in the completeness

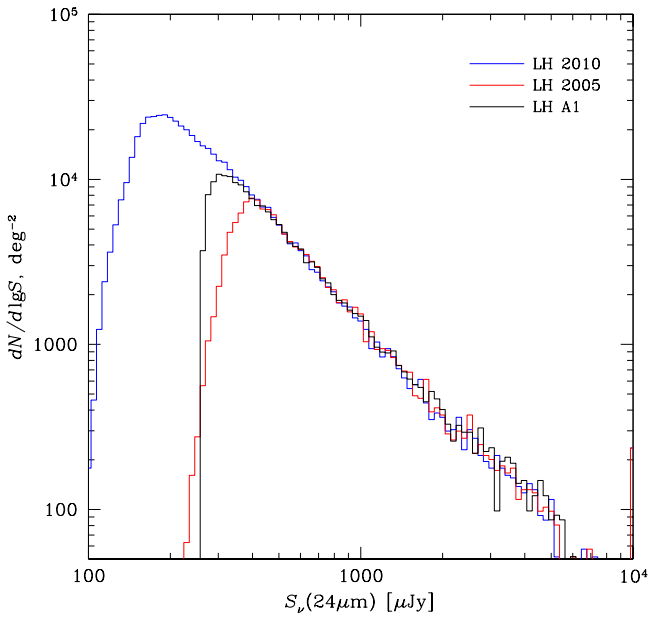



Figure 14. The number of sources per sq. degree per flux interval plotted versus the logarithmic 24 $\mu$mflux from three catalogues in the Lockman Hole. Blue, red and black lines are for 2010, 2005 and A1 catalogues, respectively.

levels between the 2005 and 2010 versions of the SWIRE team catalogues — the maxima in the differential $\log N - \log S$ distributions are at $S_\nu(24) = 400$ and 180 $\mu$Jy, respectively. The compleness level for the A1 catalogue is between these two values, at $\approx 310\,\mu$Jy. Note that the drop in number counts below the maximum is very sharp for the A1 catalogue, indicating a high level of reliability for the flux measurements. Even though the $\log N - \log S$ for the 2010-catalogue extends further down, the flux region $S_\nu(24) \lesssim 350\,\mu$Jy in this catalogue might be affected by the scatter in the source flux measurements, as we have just discussed.

The formal completeness limits (the points of maxima in the differential $\log N - \log S$ distribution) for different fields and catalogues are reported in Table IV.1 together with the field areas after applying the stellar mask (see discussion on p. 44 and Fig. 12). Below, we compare the angular correlation function computed for different fields/catalogues taking into account these completeness thresholds.

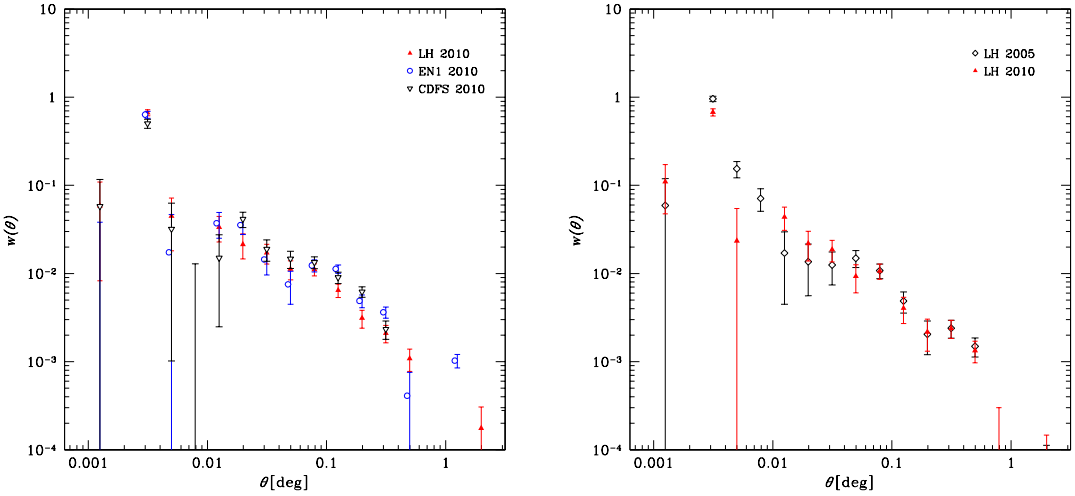



Figure 15. *Left:* Angular correlation function for sources from 2010-catalogues with fluxes brighter than $S_\nu(24) = 350\,\mu$Jy detected in the Lockman Hole (red filled triangles), ELAIS-N1 (blue open circles) and CDFS (black open triangles). *Right:* Comparison of the angular correlation functions in the Lockman Hole field using the sources from the 2010 and 2005 catalogues with $S_\nu(24) > 400\,\mu$Jy.

*Comparison of the angular correlation functions*

We start with comparison of the angular correlation functions, $w(\theta)$, computed (as described in § 4) for different SWIRE fields using the 2010-catalogues. As disscused above, we use a flux threshold of 350 μJy. This is the flux above which the $\log N - \log S$ distributions agree among different fields (Fig. 13), and it is higher than the formal completeness limit for the 2010-catalogues. The results are shown in Fig. 15a. Reassuringly, there is an excellent agreement between the results in different fields. At the largest separations, $\sim 1^\circ$ and above, the angular correlation function becomes consistent with zero, but one might expect distortions at such large scales because they are comparable to the size of the fields we are using. More relevant to our analysis are the obvious problems at small scales. There is a drop in the correlation signal at $0.003^\circ < \theta < 0.01^\circ$, and a strong positive signal located in a single bin at $\theta \sim 0.003^\circ$. As we discuss below, these distortions are probably related to blending of nearby sources due to a relatively large size of the MIPS PSF.

Next, we compare the correlation functions for the 2005- and 2010-catalogues above the completeness limit for v.2005 (400 μJy). The results for the Lockman Hole field are shown in Fig. 15b. There is a good agreement at large scales

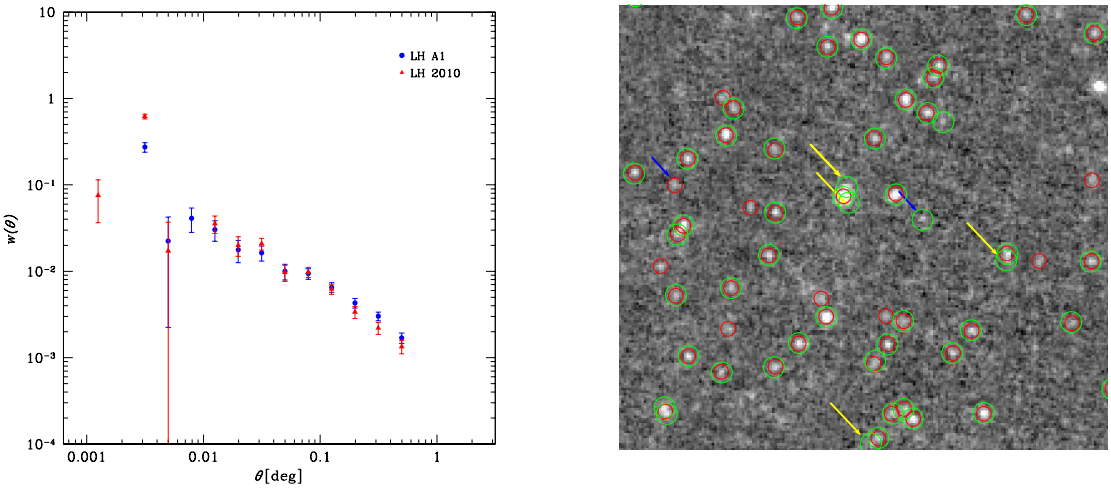



Figure 16. *Left:* Angular correlation function of 24 $\mu$m sources from the 2010 (red filled triangles) and A1 (blue filled circles) catalogues using a flux limit of $S_\nu(24) = 310\,\mu$Jy in both cases. *Right*: Comparison of the bright sources, $S_\nu(24) > 310\,\mu$Jy, in the 2010 and A1 catalogues (green and red circles, respectively) in a subsection of the Lockman Hole field. Blue arrows point to *real* detections which are not present simultaneously in two catalogues. Yellow arrows indicate spurious detections in the 2010-catalogue arising in proximity of bright/extended sources.

($\theta \gtrsim 0.02^\circ$) but a strong difference at small scales. While there is a drop in the correlation signal at $0.003^\circ < \theta < 0.01^\circ$ for the 2010 catalogue sources, there is a strong excess correlation in the same angular range for the v.2005 sources. The origin of the discrepancy is probably not because some real pairs at separations of $\sim 30''$ are missing from the 2010 catalogue — it is highly unlikely that this, more sensitive source list would miss *any* sources brighter that 400 $\mu$Jy. Rather, we think that some of these close pairs arise spuriously in the 2005 catalogue because high fluxes are erroneously assigned to some faint sources in the vicinity of bright ones (see also [235]).

Next, we compare the results for the Lockman Hole field using the sources from the 2010 and A1 catalogues above a flux threshold of $S_\nu(24) = 310\,\mu$Jy, the completeness limit of the A1 catalogue. The results are shown in Fig. 16a. The measurements are nearly identical at scales $\theta > 0.01^\circ$, but the A1 correlation function shows somewhat weaker small-scale distortions. This impression is confirmed by cross-examination of the source detections from both catalogues overlayed on the input MIPS image (Fig. 16b). Most of the sources are found in the both catalogues. There is a small number of real sources contained in one

catalogue but not the other (examples are marked by blue arrows) but this is not surprising because the source fluxes are derived using different methods and so we can expect some "migration" across the flux threshold. However, there are some cases (marked by yellow arrows) where obviously spurious sources are identified in the 2010 catalogue in the vicinity of bright or extended sources. We believe that these detections are responsible for stronger small-scale distortions seen in the v.2010 correlation function.

It is clear from the comparisons above that there is a good agreement in the correlation functions at larger scales, $\theta > 0.01^\circ$, when we compare the data for different fields and catalogues above a common completeness threshold. The differences are localized to small scales and are generally trackable to problems related to blending of sources in the MIPS images because of a relatively poor angular resolution of this instrument. These problems are not surprising. The MIPS PSF has a FWHM of $\approx 6''$ and so the sources become resolvable only when they are separated by $\sim 10'' \approx 0.003^\circ$. The MIPS PSF has wide wings — nearly 30% of the source flux is scattered outside the 8″ (radius) aperture. Therefore, there should be a substantial "cross-talk" in the flux measurements for sources separated by $\sim 15''$ (and up to 30″ depending on a source extracting algorithm in use). In any case, it appears that the angular correlation function measurements for the MIPS 24 mum sources are not reliable at $\theta < 0.01^\circ$, and it is best to restrict the analysis to larger scales. This is not a problem if our main goal is to measure the correlation length and the mass scale for the dark matter halos hosting the 24 mum sources, as these parameters are mainly constrained by the angular correlation observed near $\theta = 0.1^\circ$ (see § 15 and 16 below). However, it would be interesting to put constrains on the location of star-forming galaxies within their DM halos, which is determined by the shape of the correlation function at small scales (see § 9.1) and thus is not accessible for us.

Even though the A1 catalog appears to perform better for the smallest separations above its flux threshold, $S_\nu(24) = 310\,\mu$Jy, the difference is rather small. The 2010 catalogue, on the other hand, extends to significantly fainter fluxes, and so the question is, can we use these fainter sources to improve the statistics in the correlation function measurements? The comparison of the angular correlation function measurements in the Lockman Hole field for the A1 and

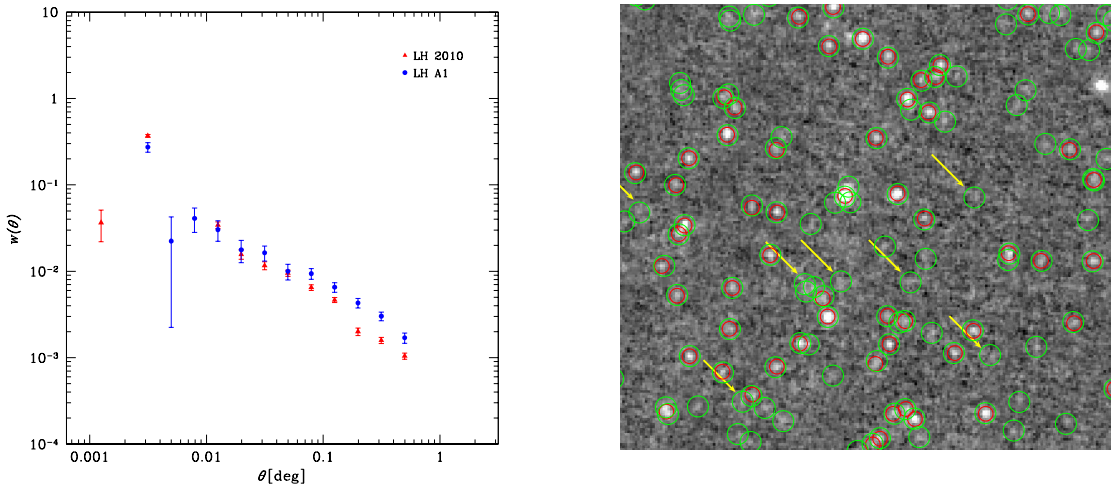



Figure 17. Same as Fig. 16 but the 2010-catalogue sources are selected above a flux limit of $S_\nu(24) = 180\,\mu$Jy. In the *right* panel, yellow arrows point to the faint sources in the 2010 catalogue for which the flux measurements are significantly affected by the large-scale background variations.

2010 catalogs above their respective flux limits of 310 and 180 $\mu$Ju is shown in Fig. 17a. Unfortunately, there are sistematic deviations for the 2010 sources at angular scales $0.2^\circ - 0.5^\circ$ (recall that the results for the two catalogues were an excellent agreement for a common flux threshold of 310 $\mu$Ju, see Fig. 16). The difference on these scales cannot be attributed to the edge effects — the size of the MIPS field in the Lockman Hole region is $\sim 4.6 \times 1.9$ deg. Rather, we believe that this difference can be traced to how the large-scale structures in the MIPS background affect the flux measurements for fainter sources in the 2010 catalogue. Examination of the MIPS image shows that, indeed, for a significant number of sources (some marked by yellow arrows in Fig. 17b), the flux above 180 $\mu$Jy is assigned spuriously, and many such sources appear on top of larger-scale background structures. These are likely *real* sources because by construction of the 2010-catalogue, they have IRAC counterparts. It is also possible that these sources are suitable for measurements of the luminosity function or similar studies because an approximately equal number of objects "migrate" below 180 $\mu$Jy in those regions with the negative residual background. However, for clustering studies, these sources cannot be used because they arise on top of spatially correlated structures and thus can distort the angular correlation function of intermediate scales.

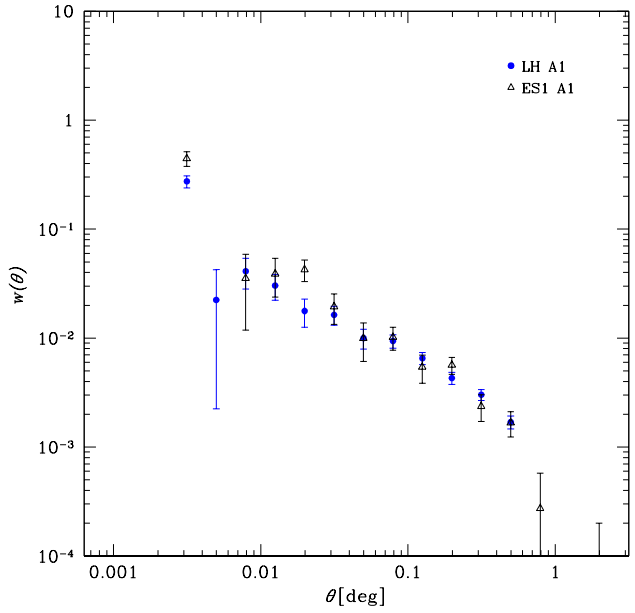



Figure 18. Angular correlation function of 24 $\mu$m sources from A1- catalogues in Lockman Hole ($S_\nu(24) > 310\,\mu$Jy) (*blue circles*) and ELAIS-S1 ($S_\nu(24) > 400\,\mu$Jy) (*black open triangles*).

As a final test, we compare the A1-based angular correlation functions for the Lockman Hole and ELAIS-S1 field (Fig. 18). The limiting flux for the A1-catalogue in the ELAIS-S1 field is $S_\nu(24) = 400\,\mu$Jy. At all angular scales, the correlation function computed for sources above this threshold in the ELAIS-S1 field is in excellent agreement with that for the Lockman Hole field and $S_\nu(24) > 310\,\mu$Jy.

On the basis of these considerations, we choose the A1 catalogue in the Lockman Hole to investigate clustering of 24 $\mu$m-selected galaxies. Lockman Hole is the largest of the SWIRE fields, which is covered by deep and uniform observations in many other bands. In addition to the MIPS and IRAC data in this field, we used the data from the 2MASS survey for the star-mask construction (see p. 44), and the optical observations carried out with INT-WFC and KPNO MOSAIC1 [Gonzalez-Solares et al., in preparation and Vaccari et al., in preparation] to photometrically separate the 24 $\mu$m-selected objects into the low- and high-redshift subsamples (§ 14).

Another option which we considered was using the publically available data for the COSMOS field, with deep MIPS observations ($S_\nu(24) > 150\,\mu$Jy)

and other numerous advantages such as the availability of the photometric and spectroscopic redshifts. However, the COSMOS field is rather small $\sim 1.5^\circ \times 1.5^\circ$, which is insufficient for reliable measurements of the clustering strength at large scales — and this is what is needed if we aim to derive the typical halo masses for the 24 mum sources using their clustering properties. Furthermore, it was shown by de la Torre et al. [60] that because of the limited size of the field, the COSMOS survey can be biased by a large-scale overdensity of galaxies at $z \simeq 1$, explaining the observed "flat-shape" two-point correlation function in this field [161, 163].

## 14 The sample of 24 $\mu$m-selected galaxies in the Lockman Hole field

We start with a list of 36529 sources (after removing Galactic stars) detected in the 24 $\mu$m MIPS image of the Lockman Hole field using the wavelet decomposition algorithm (the construction of this catalogue is discussed on p. 43). We then cross-matched this list with the catalogue of IRAC sources and applied the following flux cuts: $310 < S_\nu(24) < 2500\,\mu$Jy and $S_\nu(3.6) > 1000\,\mu$Jy, and $S_\nu(4.5) > 1000\,\mu$Jy, reducing number of sources to 30719. $S_\nu(24) = 310\,\mu$Jy is the flux at which catalogue is complete and the fluxes are measured reliably and accurately. The bright flux cuts are applied in order to conservatively discard obviously extended and/or saturated sources whose astrometry may be poor and whose flux estimates may be affected by saturation.

### *14.1 Splitting the sample into low- and high-redshift galaxy populations*

To derive spatial correlation length and investigate dependance of clustering on redshift, we need to know the redshift distribution of the sources. Unfortunately, the vast majority of the 24 $\mu$m sources selected in the Lockman Hole field have neither spectroscopic nor photometric redshifts. The SWIRE photometric redshift catalogue [211], available in this field, has a limited and heavily inhomogeneous coverage for our sample. Using such data for the clustering analysis is dangerous because an artificial signal can be introduced into the correlation function. The approach we are taking instead is to use simple photometric criteria to separate in the catalogue into the low and high-redshift subsamples, and

then use external data to derive the redshift distribution within each subsample.

To separate the sample into low and high-redshift sources, we applied the optical-to-NIR color selection criterion established by S. Berta (2008, private communication) in a similar SWIRE field (ELAIS-S1). Berta used the fluxes in the optical $I_{\rm VIMOS}$ (from ESIS-VIMOS survey) and SWIRE IRAC 4.5 $\mu$m bands to examine the dependence of $(I_{\rm VIMOS} - [4.5])_{\rm AB}$ color on the redshift for various galaxy spectral templates (Fig. 19a). It appears that for starburst galaxies (the M 82 template, shown by green line) the $(I_{\rm VIMOS} - [4.5])_{\rm AB} \sim 3$ color successfully separates low ($z < 1$) and high ($z > 1$) redshift galaxy populations, with only a small contaminations in both groups. Such a rapid color transition along the redshift may be explained by the passage of the Balmer break through or redward of the *I*-band (Fig. 20). Figure 19b shows a color-magnitude diagram of the 24 $\mu$m sources with $S_\nu(24) > 400\,\mu$Jy detected in the SWIRE + ESIS-VIMOS ELAIS-S1 field, along with the redshift tracks for the spectral templates of galaxies shown in Fig. 19a. The templates were normalized to a bolometric (or, more precisely, 0.1 $\mu$m – 1000 $\mu$m) luminosity of $10^{12} L_\odot$. This diagram provides an additional proof that a color cut at $(I_{\rm VIMOS} - [4.5])_{\rm AB} \sim 3$ can be used to divide the observed sources into low and high-$z$ galaxy populations.

To obtain an empirical confirmation of this color selection criterion, we applied it to the deep *Spitzer* observations of GOODS fields [207]. The GOODS-N and GOODS-S 24 $\mu$m catalogues include 889 and 614 sources, respectively, detected in a total area of $\sim 350\,\mathrm{arcmin}^2$. The catalogues are complete down to $S_\nu(24) = 80\,\mu$Jy. Observations in *i*-band were made by the ACS in both fields down to a magnitude limit a $i$=26.5 [95]. The redshift estimates are available for all these sources — 46% spectroscopic and 54% photometric, which are estimated with a high accuracy (the *rms* scatter in $z_{\rm phot} - z_{\rm spec}$ is 0.09 and 0.06 for the GOODS-N and GOODS-S samples, respectively [207]).

From the GOODS catalogues, we selected the sources with $S_\nu(24) > 310\,\mu$Jy and separated them into two redshift bins $z > 1.2$ and $z < 1.2$[6]. The color-magnitude diagram for these sources (Fig. 21a) shows that the low- and high-$z$ galaxies can be separated by a boundary value of $(i - 4.5) = 3$ (AB mag) (dashed

[6]The boundary was chosen near the minimum of the bimodal redshift distribution prodicted by the Franceschini et al. [81] model.

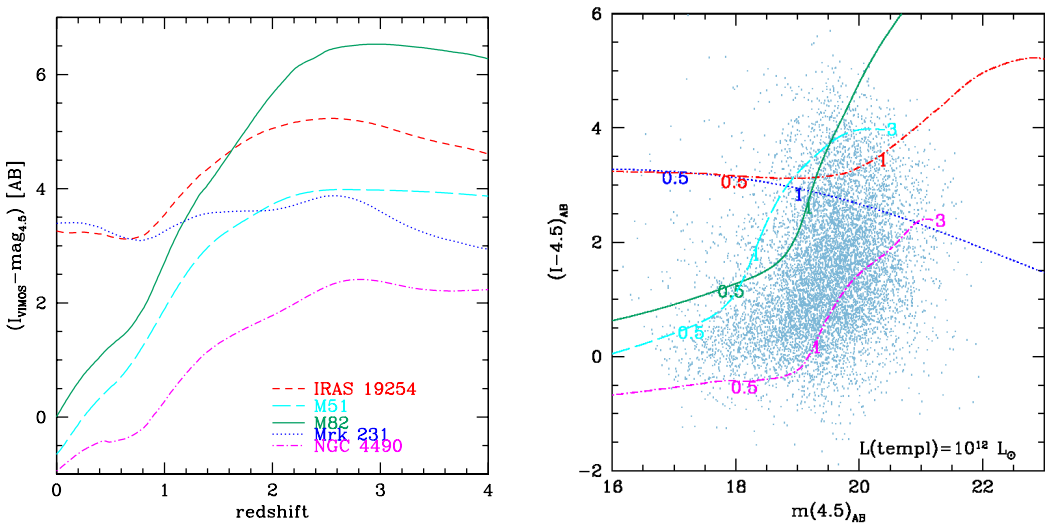



Figure 19. *Left:* Dependence of the $(I_{\rm VIMOS} - [4.5])_{\rm AB}$ color on redshift for several galaxy spectral templates. *Right*: The color-magnitude diagram where small dots represent real observed 24 $\mu$m-emitting galaxy populations in the SWIRE ELAIS-S1 field. Superimposed are the template tracks Mrk 231 (Sy-1), IRAS 19254 (Sy-2), a starburst (M 82), a spiral (M 51) and a blue spiral (NGC 4490) galaxies. On both panels, the template colors are the same. On the right panel, the templates were normalized to a bolometric luminosity of $10^{12}\,L_{\odot}$ and the redshift ranges are reported along each track. (adopted from Berta 2008, private communication.)

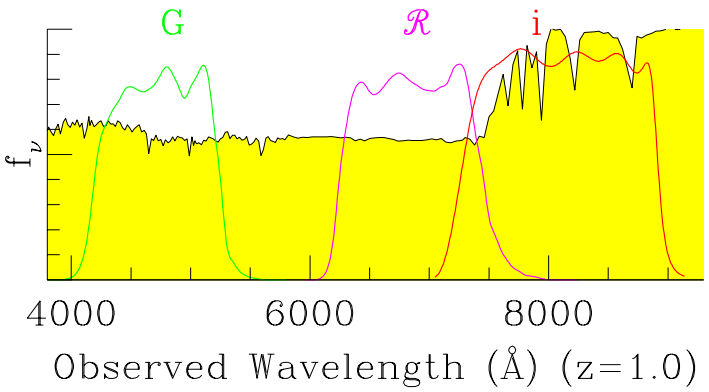



Figure 20. Example of the Balmer break, redshifted to $z = 1$, in SED of a model galaxy that has been forming stars at a constant rate for 1 Gyr [2].

line). The deepest optical data available in the Lockman Hole field are those from the INT/WFC which provides sufficiently uniform coverage to $i = 22.8$ (with the $5\sigma$ magnitude limit reaching $i = 23.6$ (AB) in the deepest sections of the survey). Therefore, a magnitude cut of $i = 22.8$ had to be incorporated in our selection. Figure 21b shows that the sources fainter than $i = 22.8$ and with the color $(i - 4.5) < 3$ (AB mag) in practice are very few and they only minimally contaminate the high-$z$ sample. Based on these considerations, we implemented the redshift separation as a combined color and magnitude criterion: the source is considered to belong to a high-redshift sample if it is undetectable in the INT/WFC $i$-band, or its measured $i$ magnitude is $> 22.8$, or the $(i - 4.5)$ (AB mag) color is $> 3$.

One of the main sources of concern for the color-magnitude based separation of 24 $\mu$m objects into low- and high-redshift subsamples using color-magnitude criteria is presence of AGNs in a sample. Figure 19 illustrates that AGNs of both types do not follow the color selection and thus can contaminate the high-redshift locus of color-magnitude diagram. Therefore, we checked the AGN contents in the GOODS sample of the 24 $\mu$m-selected sources. According to Rodighiero at al. [207], there are $\lesssim 10\%$ of type-1 AGNs in their GOODS sample. These authors classified the observed SEDs using the Polletta et al. [191] templates. This AGN fraction is consistent with that reported by Gilli et al. [90] and Treister et al. [244], who used very deep *Chandra* X-ray observations in the GOODS fields. Concerning the highly obscured (type-2) AGNs and the sources of composite spectral type (starburst+ANG), their contribution to the 24 $\mu$m emitting sources is hard to estimate. One of the reasons is that the AGN and star-formation activity often occur simultaneously, and both are revealed in the form of the 24 $\mu$m emission (see, e.g., [33], [207], [80] and references therein). Some studies suggest, on the basis of estimates by different methods, that the 24 $\mu$m-selected samples may contain $\sim 20\% - 30\%$ of AGNs of both types [213, 80]. However, we note that to estimate the redshift distribution within our color and $i$-magnitude selected subsamples, we used an empirical redshift distribution of identically selected GOODS sources (see below). As long as the GOODS redshifts are valid and the GOODS sample is a fair representation of our main Lockman Hole sample, the derived $dN/dz$ models for the low- and

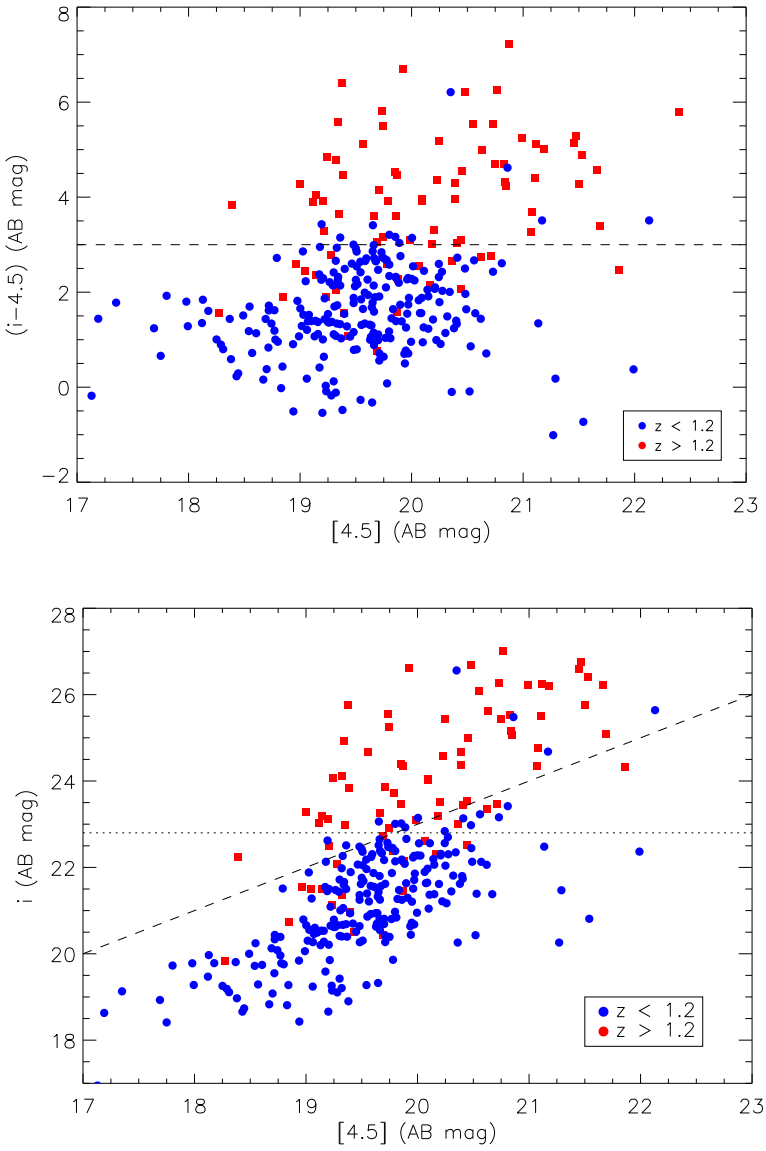



Figure 21. Color-magnitude *(a)* and magnitude $i$ vs magnitude [4.5] *(b)* diagrams for the GOODS-N and GOODS-S sources with $S_\nu(24) > 310\,\mu$Jy. Red squares and blue circles are galaxies at redshifts lower and higher than 1.2. On both figures, a dashed line represents $(i - 4.5) = 3$ (AB mag). A dotted line on the figure *(b)* corresponds to a magnitude $i = 22.8$ at which the WFC coverage is uniform in the Lockman Hole field.

high-redshift subsamples are correct, even though the high-$z$ subsample may be slightly contaminated by AGNs.

### *14.2 Empirical redshift distribution*

As was mentioned above, we need a model for the redshift distribution of the sources in order to use the Limber equation 2.23 to relate the angular and spatial correlation functions. We determined these redshift distributions empirically, using the GOODS sources selected identically to our main sample in the Lockman Hole field. All sources with $S_\nu(24) > 310\,\mu$Jy in GOODS-N and GOODS-S fields were divided into low and high-redshift subsamples by applying the color-magnitude selection criteria (§ 14.1 and Fig. 21b). The obtained redshift distributions within these photometrically-selected samples are shown in Fig. 22a and b. These empirical distributions can be well approximated by a Gaussian model:

$$dN/dz = \text{const} \times \exp(-(z - z_{\text{peak}})^2/2\sigma^2) \tag{4.1}$$

(blue and red lines in Fig. 22). The best-fit parameters for the low-$z$ subsample in the redshift range $0 < z < 2$ are const $= 50, \sigma^2 = 0.122, z_{\text{peak}} = 0.68$. For the high-$z$ subsample in the redshift range $0.5 < z < 3.5$, we find const $= 12, \sigma^2 = 0.396, z_{\text{peak}} = 1.72$.

This two-Gaussian model provides a good fit also to a redshift distribution of all GOODS sources with $S_\nu(24) > 310\,\mu$Jy (i.e., without the photometric separation into low and high-$z$ subsamples). The combined redshift distribution is shown in Figure 23, and the dashed line is the sum of two Gaussian models for the low and high-$z$ samples.

We also considered using the predictions of the Franceschini et al. [81] model for the redshift distribution of sources in our sample. This model reproduces the real redshift distribution of GOODS sources quite well [207], however it was very difficult to incorporate the color selection and optical magnitude limit which we employ to the real data. Therefore, we decided in favour of using the empirical redshift distribution obtained from GOODS sources for studying the clustering properties of the two photometrically selected subsamples of 24 $\mu$m galaxies with $z_{\text{mean}} = 0.68$ and $z_{\text{mean}} = 1.72$ (Fig. 22).

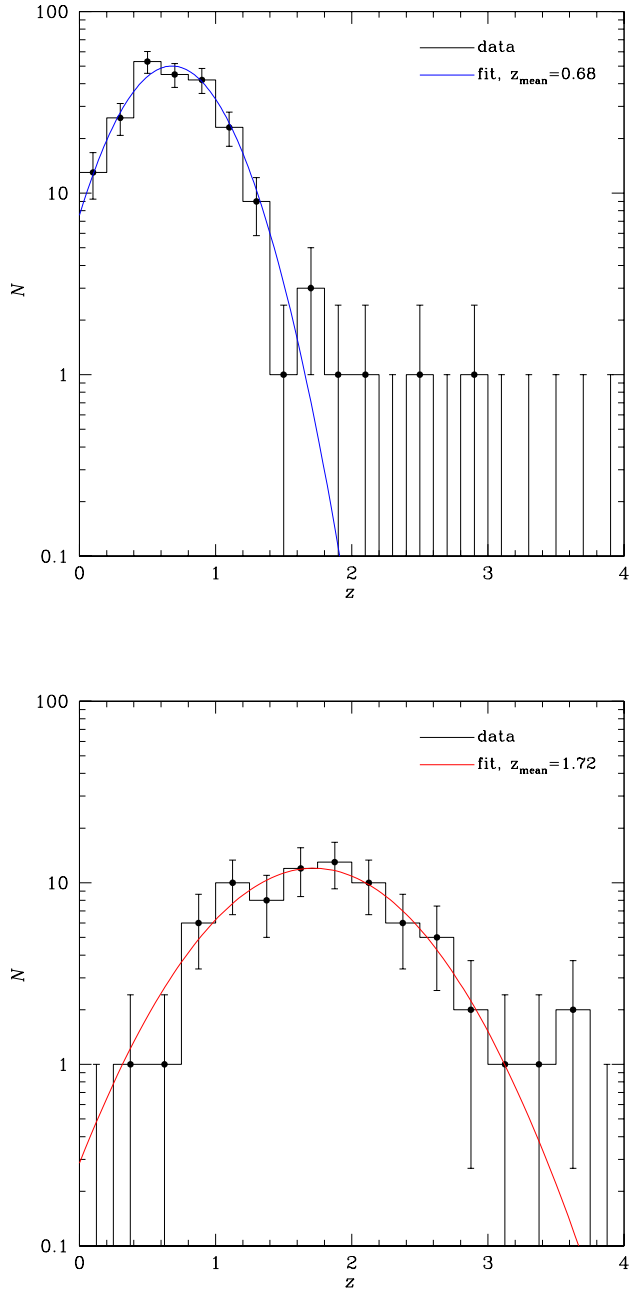



Figure 22. Redshift distribution of GOODS sources ($S_\nu(24) > 310\,\mu$Jy) incorporated into low-$z$ *(top)* and high-$z$ *(bottom)* subsamples based on their color ($i - [4.5]$) and $i$-band magnitude. Blue and red lines are Gaussian fits with $z_{\rm peak} = 0.68$ and $z_{\rm peak} = 1.72$, respectively.

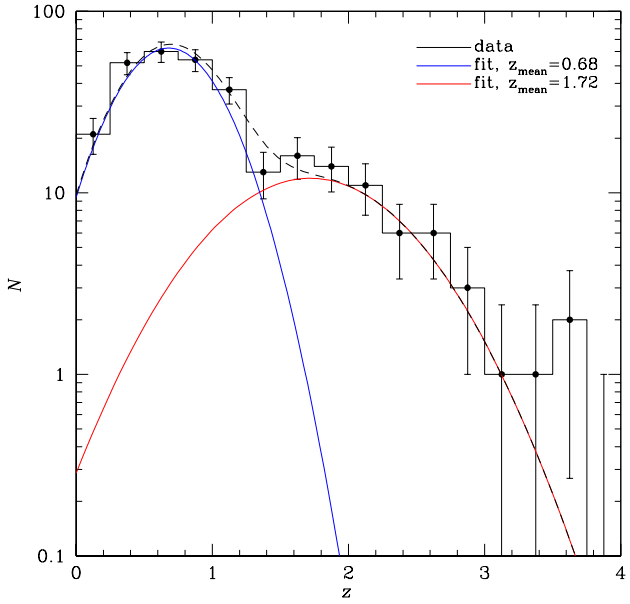



Figure 23. The redshift distribution of sources brighter than $S_\nu(24) = 310\,\mu$Jy from GOODS surveys. Blue and red lines are Gaussian fits to redshift distributions of sources undergone color-magnitude selection. The dashed line is a combined fit of two selected samples.

## 15 Estimation of angular and spatial clustering parameters

Before we begin the description of the clustering analysis of the 24 $\mu$m-selected galaxies in the Lockman Hole field, a short note on the region mask we used is in order. As was described above, we use the INT/WFC optical data to separate the sample photometrically into the low and high-$z$ subsamples (§ 14.1). Unfortunately, the INT/WFC observations are insufficiently deep in some subsections of the MIPS Lockman Hole image, and we had to mask out those regions. The resulting region mask is shown in Fig. 24. The circles in these mask mark the locations of bright stars and extremely bright 24 $\mu$m sources (see § 13.1 for details). To identify the regions of insufficient INT/WFC depth, we examined the distribution of optical counterparts for 3.6 $\mu$m IRAC sources at various $i$-band magnitude cuts. We found that the depth is held at least at $i = 22.8$ throughout the field except for the regions masked out as rectangles in Fig. 24. At fainter magnitudes, the WFC coverage becomes highly nonuniform. Therefore, we use a limit of $i = 22.8$ for the redshift separation (§ 14.1), and the region mask shown

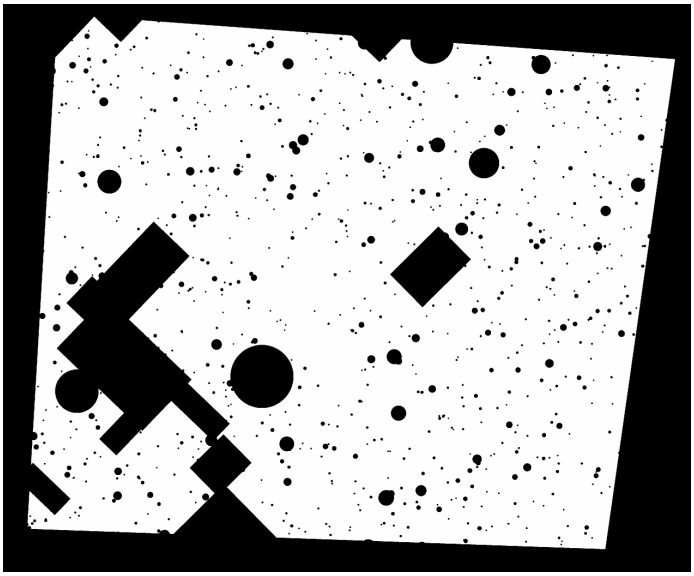


Figure 24. Final region mask for the clustering analysis in the Lockman Hole field. The circles mark the locations of stars and bright objects. The rectangles mask those regions where the completeness of INT/FWC images is not achieved for $i = 22.8$ (AB mag). All black patches were excluded from the subsequent analysis.

in Fig. 24 for estimation of the angular correlation function.

The total area of the remaining field (white regions in Fig. 24) is $\simeq 7.9$ deg$^2$. There are 21844 24 $\mu$m-emitting objects with fluxes greater than $S_\nu(24) = 310\,\mu$Jy within this area. Applying the color-magnitude selection criteria (§ 14) we obtained two subsamples of 14822 and 7022 sources with $z_{\rm mean} = 0.68$ and $z_{\rm mean} = 1.72$, respectively.

The angular correlation functions were estimated by the Landy & Szalay method (eq. 2.2) at angular scales $0.01 < \theta < 1.0$ degrees. The random points used in this estimator were homogeneously distributed in the field but avoiding the excluded regions of the mask shown in Fig. 24. In order to suppress the uncertainties related to a complex geometry of the field and to decrease the statistical errors, the number of simulated random points was factor of 100 greater than the number of data points in each sample. The correlation function was computed in the angular bins $\Delta \log \theta = 0.2$, and the Poisson errors were computed by eq. 2.4. In Fig. 25, we show the derived angular correlation functions for the whole sample (red), for the low-$z$ subsample with $z_{\rm mean} = 0.68$ (blue), and for high-$z$ subsample with $z_{\rm mean} = 1.72$ (black).

Because of the good statistics of the SWIRE sample and a large size of the

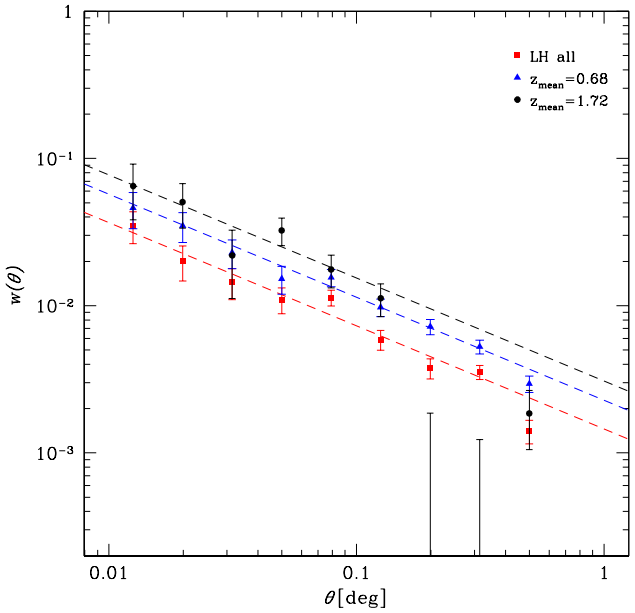



Figure 25. Two-point angular correlation function of SWIRE Lockman Hole sources brighter than $S_\nu(24) = 310\,\mu$Jy. The dashed lines are power-law fits. Red points represent clustering of the whole sample, blue and black points are for the low-$z$ and high-$z$ galaxies, respectively.

Lockman Hole field, we are able to measure the clustering signal at angular scales which correspond to fairly large spatial scales. Indeed, comoving sizes of $1 - 8\,h^{-1}$ Mpc at $z = 1.72$ correspond to an angular range of $0.017° - 0.13°$. A great advantage of the measurements done at such large scales is that we directly probe the clustering signal at angular separations which correspond to the expected range of three-dimensional correlation lengths, $r_0$, for the sources. This makes it possible to obtain robust estimates of $r_0$ from a standard power-law fit to the angular correlation function and application of the simplified Limber equation (see § 5, eq. 2.23). If the angular correlation function measurements at large scales are unavailable, a power law fit to the data at small angular/spatial scales may lead to incorrect estimates of the correlation lengths and incorrect conclusions about clustering properties of given galaxy populations (e.g., [127], [200], [201] and references there). Later in this thesis (Chapter V), we also discuss an example in which using a power law approximation leads to an underestimation of the total halo mass for very massive galaxies at $z \sim 2$.

The angular correlation functions shown in Fig. 25) were fitted over the

angular range $0.01^\circ < \theta < 0.4^\circ$ with a power law model, $w(\theta) = (\theta/\theta_0)^{1-\gamma} - \mathrm{IC}$ (eq. 2.21), where IC is the so-called Integral Constraint ( § 2). The best-fit parameters for the entire sample are $\theta_0 = 0.32'' \pm 0.03''$, and $\gamma = 1.7 \pm 0.2$. Splitting the whole sample into smaller subsamples obviously increases the statistical uncertainties. Therefore, we decided to fix the power law slope in the subsequent analysis at $\gamma = 1.7$. The best-fit amplitudes for the low-$z$ and high-$z$ data are then $\theta_0 = 0.60'' \pm 0.05''$ and $\theta_0 = 0.93'' \pm 0.18''$, respectively[7]. These best-fit models are shown in Fig. 25 with blue and black dashed lines.

The spatial correlation lengths $r_0$ were then obtained from the Limber inversion (§ 5, eq. 2.23) which used the fits to the empirical redshift distributions of GOODS survey sources, described in § 14.2. The derived correlation lengths are $r_0 = 4.74 \pm 0.16\,h^{-1}$ Mpc (comoving) for the low-$z$ ($z_{\mathrm{mean}} = 0.68$), and $r_0 = 7.87 \pm 0.63\,h^{-1}$ Mpc for the high-$z$ ($z_{\mathrm{mean}} = 1.72$) sample.

The uncertainties above include only statistical errors in the measurement of the angular correlation function. In principle, another source of uncertainty is the inaccuracies in the models for the redshift distribution. These are hard to estimate in our case since we use an empirical fit to the $dN/dz$ observed for the GOODS sources and any inaccuracies would be related to problems with the GOODS photometric redshifts[8]. A range of theoretical models for the redshift distribution of 24 $\mu$m sources is a poor guidance because these models sometimes provide contradictory results [63, 211, 81]. Qualitatively, if the real $dN/dz$ distribution for our sources is wider then what we assume, the correlation lengths should be corrected upwards.

We now turn to using the derived correlation lengths for the 24 $\mu$m-selected galaxies for estimating the mass range of their host dark matter halos through the comparison of our measurements with the clustering properties of DM halos from the *Bolshoi* cosmological simulation.

[7]For the high-$z$ sample, the fit was over the angular range $0.01^\circ < \theta < 0.13^\circ$ because at larger angles, we expect that the correlation function can start to deviate from the power law (cf. Fig. 27 below).

[8]We are unaware of such problems, and in any case, their discussion is beyond the scope of our work.

## 16 Properties of dark matter halos hosting 24 $\mu$m-selected galaxies

Several methods can be used to connect a population of galaxies with a population of their host DM halos (see, e.g., [96] and references therein, and discussion in Chapter III, § 6). Here, we use the clustering properties assuming that the mass scale of the dark matter halos hosting the galaxies can be established by requiring that the observed correlation function of galaxies selected above a luminosity threshold matches the correlation function of DM halos selected above some mass limit. This is an approach similar to that used in the "halo matching" technique [246, 48], which assumes that there is a unique and monotonic relation between the galaxy luminosity and mass or maximum circular velocity of its host DM halo.

To compute the correlation function of the dark matter halos, we used the outputs of the *Bolshoi* cosmological simulation (described in Chapter III, § 9) at redshifts range from 0.5 to 2.5 with a step of $\Delta z = 0.5$.

Each output of the *Bolshoi* simulation provides a catalog of gravitationally bound objects (unique halos and subhalos contained in more massive halos). The maximum circular velocity, $v_{\rm max}$, is provided for each (sub)halo. As was discussed in Chapter III, $v_{\rm max}$ of a dark matter halo is more closely matched than its total mass with the properties of a galaxy residing in this halo. Therefore, we "populated" all halos and subhalos selected above different $v_{\rm max}$ thresholds (these threshold values of $v_{\rm max}$ are referred to as $V_{\rm min}$ hereafter) with "galaxies" and computed the spatial correlation functions. The considered range of $V_{\rm min}$ is $130 < V_{\rm min} < 385\,{\rm km\,s^{-1}}$. The lower velocity limit is chosen so that the correlation length for such DM halos is below the $r_0$ derived for our low-$z$ subsample of 24 $\mu$m galaxies. The high velocity limit is chosen to ensure that the statistics of DM halos is sufficiently good at all output redshifts of the *Bolshoi* simulation. We estimated the correlation lengths for the model galaxy populations by fitting their spatial correlation functions with a power-law at scales $1 < r < 25\,h^{-1}$ Mpc.

Figure 26 shows the derived model correlation lengths for DM halos as a function of $V_{\rm min}$ and redshift. Clearly, the $r_0$ significantly increases with $V_{\rm min}$ (or mass) of the halos and also changes with redshift. These correlation lengths can be matched to the observed $r_0$ for our samples of 24 $\mu$m-selected galaxies. The

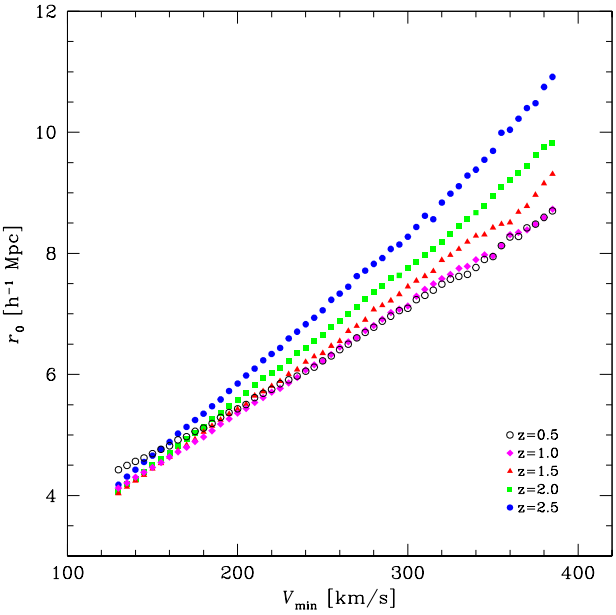



Figure 26. The spatial correlation length of dark matter halos as a function of the maximum rotational velocity threshold and redshift.

redshifts of the simulation outputs do not match exactly the mean redshifts of our galaxy samples, $z_{\rm mean} = 0.68$ and $z_{\rm mean} = 1.72$. However, the trend of the model $r_0$ with $z$ for a given $V_{\rm min}$ is weak[9], and so we can linearly interpolate between the results for the outputs branching the mean redshifts in the data.

Using these data, each observed value of $r_0$ can be matched to the corresponding $V_{\rm min}$. The uncertainty intervals for our low and high-$z$ subsamples, $r_0 = 4.74 \pm 0.16\,h^{-1}$ Mpc and $r_0 = 7.87 \pm 0.63\,h^{-1}$ Mpc, respectively, correspond to $V_{\rm min}$ intervals of $V_{\rm min} = 158 \pm 10$ km s$^{-1}$ for low-$z$ 24 $\mu$m galaxies and $V_{\rm min} = 313 \pm 31$ km s$^{-1}$ for the high-$z$ subsample with $z_{\rm mean} = 1.72$.

These velocity thresholds can be easily converted to the corresponding virial mass limits, $M_{\rm vir}$, using a tight scaling, which approximately goes as $v_{\rm max} \propto M_{\rm vir}^{1/3}$ (e.g., [121]; see also § 7.2, Fig. 5). This relation is valid for both distinct halos and subhalos at different redshift. Fitting the $v_{\rm max} - M_{\rm vir}$ relation for all halos and subhalos above $v_{\rm max} > 130$ km s$^{-1}$ in the *Bolshoi* outputs, we obtain the following

[9]Note that $r_0$ as a function of *mass* does evolve with redshift, as expected. However, this evolution appears to be canceled by the evolution in the $M - v_{\rm max}$ relation and the trend of $r_0$ with $M$ at a given redshift.

power-law scalings:

$$\log M_{\rm vir} = 4.60 + 3.25 \log v_{\rm max}, \quad \text{for } z = 0.5, \tag{4.2}$$

$$\log M_{\rm vir} = 4.69 + 3.13 \log v_{\rm max}, \quad \text{for } z = 1.5, \tag{4.3}$$

where $M_{\rm vir}$ is in units of $h^{-1}\,M_\odot$. These results can be scaled to the mean redshifts of the sample using the expected reshift evolution of the $v_{\rm max} - M$ relation, which goes as $M \propto E(z)^{-1}$ for a fixed $v_{\rm max}$ [29], where $E(z) = H(z)/H_0$. Using these scalings, we find that the limiting total mass for the 24 $\mu$m-emitting galaxies with $z_{\rm mean} = 0.68$ is $M_{\rm tot} = (0.5 \pm 0.1) \times 10^{12}\,h^{-1}\,M_\odot$[10] and $M_{\rm tot} = (2.9 \pm 0.9) \times 10^{12}\,h^{-1}\,M_\odot$ for our high-$z$ sample at $z_{\rm mean} = 1.72$.

In the modeling presented above, it was implicitly assumed (following the halo matching technique) that every halo (distinct or satellite) with mass above some threshold, harbors a galaxy. One can reasonably expect that this assumption is well-justified if the sample is selected by the emission from the main-sequence stellar populations/late-type giants and thus the selection is closely related to the total stellar mass of a halo. However, we might expect that only some fraction of massive halos contain galaxies which show the emission related to the star-formation or AGN activities. Indeed, we find that the observed number density of 24 $\mu$m-selected galaxies is only $\sim 20\%$ of the number density of dark matter halos with the masses derived above from the clustering properties.

The comoving number density of galaxies near the mean redshift of the sample can be estimated as:

$$n_{\rm gal} = \frac{dN/dz}{dV/dz} = 1.1 \times 10^{-3}\,h^3\,{\rm Mpc}^{-3}, \quad \text{for } z_{\rm mean} = 0.68, \tag{4.4}$$

$$n_{\rm gal} = 0.12 \times 10^{-3}\,h^3\,{\rm Mpc}^{-3}, \quad \text{for } z_{\rm mean} = 1.72, \tag{4.5}$$

where $dV/dz$ is the comoving volume within the survey area [103]. These values can be compared with the number density of halos in the *Bolshoi* outputs above the derived $V_{\rm min}$ thresholds. For $z = 0.5$, $v_{\rm max} > 158\,{\rm km\ s^{-1}}$, we find $n_{\rm halo} = 6.3 \times 10^{-3}\,h^3\,{\rm Mpc}^{-3}$, or $n_{\rm halo} \approx 6\,n_{\rm gal}$. For $z = 1.5$, $v_{\rm max} > 313\,{\rm km\ s^{-1}}$, the corresponding

[10] For reference, the Milky Way dark matter halo is estimated to have $v_{\rm max} = 201\,{\rm km\ s^{-1}}$ and $M_{\rm tot} \sim 1.4 \times 10^{12}\,h^{-1}\,M_\odot$ (e.g., [96]).

number densities are $n_{\rm halo} = 0.53 \times 10^{-3}\, h^3\,{\rm Mpc}^{-3}$ or $n_{\rm halo} \approx 5\, n_{\rm gal}$.[11] Therefore, we find that similar fractions of DM halos contain galaxies with the 24 $\mu$m flux $> 310\,\mu$Jy at both low- and high-redshift. This may be simply a coincidence since the mass and 24 $\mu$m luminosity scales for the two samples are quite different and so we can not separate the luminosity and redshift dependences. Such an analysis will be extremely interesting when larger samples of the mid-IR selected galaxies become available.

Also, we note that even though our results clearly imply that the 24 $\mu$m-selected galaxies reside only in a subsample of DM halos, the clustering modeling would still be valid if these objects sample the DM halos *randomly*. This is probably not the case since one expects that the star forming galaxies should be associated with the objects which recently experienced a major merger. The studies which make a distinction between halos with different formation histories will also be possible in the future, if the data on the small-scale clustering of the 24 $\mu$m-selected galaxies become available.

Finally, we test that our analysis based on the power law approximation of the observed angular correlation functions provides unbiased answers even though the correlation function of DM halos shows clear deviations from the power law (§ 9.1). For this, we compute a full projection of the two-point spatial correlation function of DM halos for $v_{\rm max} > 158\,{\rm km\ s}^{-1}$ at $z = 0.5$ and $v_{\rm max} > 313\,{\rm km\ s}^{-1}$ at $z = 1.5$[12] The 3D correlation functions for the halos were calculated at scales $0 < r < 50\,h^{-1}$Mpc in narrow bins, $\Delta r = 0.1\,h^{-1}$Mpc, bins, and then were used in the full Limber transformation (eq. 2.22, § 5). This equation computes the expected angular correlation function without assuming a power-law approximation. The results are shown in Fig. 27. The blue and black data points show the observed angular correlation functions for the low-$z$ and high-$z$ samples (same as those in Fig. 25), and the dashed lines are the respective

[11] Note that we derived the halo number density at the closest output redshifts of the *Bolshoi* simulation and compared them with the observed number densities at slightly different redshifts. This comparison is valid due to the weak dependence of halo velocity function on redshift for moderate $v_{\rm max}$ and $z$ (Fig. 6, [121]).

[12] Note that in calculating the projected models, we neglected the redshift evolution of the DM halo correlation function within the redshift intervals covered by the data. As is clear from Fig. 26, the change in the clustering length at our $V_{\rm min}$ thresholds is comparable to the statistical uncertainties for the $r_0$ measurements, so this assumption is justified.

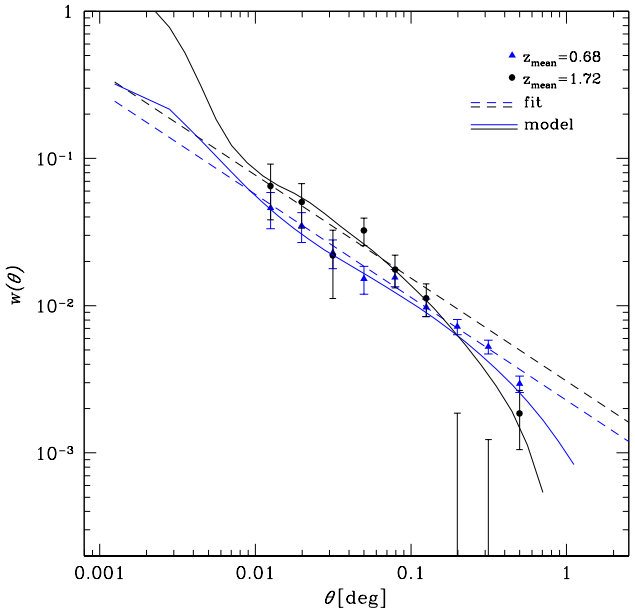



Figure 27. The observed two-point angular correlation function for low-$z$ (blue) and high-$z$ (black) samples of the 24 $\mu$m-selected galaxies. The dashed lines are the power law approximations discussed on p. 64. Solid lines are the angular correlation function models derived from the full Limber inversion of spatial correlation functions of DM halos with maximum circular velocities greater than $V_{\rm min}$ = 158 km s$^{-1}$ (blue) and $V_{\rm min}$ = 313 km s$^{-1}$ (black).

power law approximations. The solid lines shows the full projections of the halo correlation functions.

Clearly, the full models fit the data points very well, confirming that the power-law approximation to the observed $w(\theta)$ yields accurate estimates of the spatial correlation lengths, $r_0$, and thus accurate mass scales for the DM halos hosting the 24 $\mu$m-selected galaxies. Furthermore, at $\theta > 0.2$ deg we observed a decline of the observed correlation functions relative to the power-law approximations, and this could be related to the large-scale drop in the correlation function of DM halos, which is related to the transition from nonlinear to linear regimes of the structure formation theory [51]. At the opposite end, $\theta < 0.01$ deg, models show enhancements in the clustering signal relative to the power-law fit. These enhancements correspond to the correlation function of galaxies sharing a single parent halo (the so-called "1-halo" term, see Chapter III, § 9 for details); they can be used to constrain the fraction of galaxies which reside in subhalos. Unfortunately, a broad PSF of the MIPS instrument does not allow

one to make reliable measurements of the clustering of 24 $\mu$m sources at such small scales (see discussion in § 13.1 above).

## 17 Comparisons with previous studies

It is interesting and important to compare our measurements with the previous studies of the clustering properties of the high-redshift galaxy populations. In doing this, we should keep in mind that direct comparisons with other studies, even those based on *Spitzer* data, are difficult to make because of a wide variety of criteria applied to select high-redshift galaxies. However, even in this case a comparison is useful because oftentime, the galaxy samples selected by different techniques might be intrinsically connected (e.g. [173]).

First, in order to better understand what types of galaxies we probe in this work, we estimated typical infrared (IR) luminosities (8 $\mu$m$-$1000 $\mu$m) for galaxies brighter than $S_\nu(24) = 310\,\mu$Jy using GOODS catalogues [207]. We found that for the low redshift subsample, $z < 1.2$, the mean luminosity is $L_{\rm IR} \sim 3\times10^{11} L_\odot$[13], so these selected objects probably belong to the class of luminous infrared galaxies ("LIRGs", $10^{11}L_\odot < L_{\rm IR} < 10^{12}L_\odot$, [214]). The high redshift galaxies, $z > 1.2$, have an order of magnitude higher mean luminosity, $L_{\rm IR} \sim 3\times10^{12}L_\odot$ which places them into the category of ultraluminous infrared galaxies ("distant ULIRGs", $L_{\rm IR} > 10^{12}L_\odot$, [214]). Barring an unexpectedly high level of cosmic variance, these mean luminosities estimated from the GOODS data should properly indicate the mean luminosities for our 24 $\mu$m sources selected in the SWIRE Lockman Hole.

At $z < 1$, the correlation strength and environment of star forming galaxies were estimated in different surveys. For example, the blue galaxies at $z < 0.25$ selected from the SDSS have $r_0 \sim 4\,h^{-1}$ Mpc [238]. The UV-detected galaxies at $z < 0.4$ selected from the *GALEX* data show a similar correlation length of about $4\,h^{-1}$ Mpc [164, 139], which at the same time is close to that derived for the "Lyman Break Galaxies" (LBGs) at $z \simeq 3$ [196, 3]. The correlation lengths measured for these optically and UV-selected star-forming galaxies indicate a

[13] $L_\odot = 3.83\times10^{33}$ erg s$^{-1}$ is the bolometric luminosity of the Sun.

mass scale of their parent DM halos of $\gtrsim 4 \times 10^{11}\, h^{-1}\, M_{\odot}$. It is worth noting the measurements of clustering for Ultraviolet Luminous Galaxies ("UVLGs") by Basu-Zych et al. [10]. These authors analyzed the low-to-intermediate redshift galaxies ($0.4 < z < 1.2$) with a high FUV luminosity, indicating a high star formation rate. These objects are considered to be direct counterparts to local LIRGs and ULIRGs. Indeed, the correlation length $r_0 \sim 4.8\, h^{-1}$ Mpc estimated for the UVLGs (unfortunately, with high statistical uncertainties) is fully consistent with the clustering strength in the low-$z$ sample studied in our work. Gilli et al. [90] presented the correlation function measurements of the $S_\nu(24\,\mu\mathrm{m}) > 20\,\mu$Jy galaxies with the mean $z \sim 0.8$, detected in the GOODS fields. They found that the correlation length increases with the infrared luminosity, reaching for LIRGS a level of $r_0 \sim 5.1\, h^{-1}$ Mpc, corresponding to a mass of host halos $> 1 \times 10^{12} h^{-1}$ Mpc. This total mass is twice high than what we report for our LIRGs. Another study of bright 24 $\mu$m emitting galaxies was performed by Magliocchetti et al. [150]. The galaxies brighter than $S_\nu(24\,\mu\mathrm{m}) = 400\,\mu$Jy detected in the SWIRE XMM-LSS field were divided into low-redshift ($z_{\mathrm{mean}} = 0.79$) and high-redshift ($z_{\mathrm{mean}} = 2.02$) subsamples based on photometric redshifts. The samples are thus comparable to those selected in our work. The derived correlation length were $5.9^{+1.1}_{-1.3}\, h^{-1}$Mpc and $r_0 \sim 11.1^{+2.0}_{-2.4}\, h^{-1}$Mpc for the low and high-$z$ subsamples, respectively. The corresponding masses of DM halos were inferred to be $> 4 \times 10^{11}\, h^{-1}\, M_{\odot}$ and $> 4 \times 10^{12}\, h^{-1}\, M_{\odot}$. Within uncertainties, these results are in a fairly good agreement with our measurements. However, our sample contains a much larger number of sources and covers a wider area, so we were able to measure the angular correlation function at larger scales (probing directly the "2-halo" term, see Chapter III for details) and significantly reduce the statistical uncertainties.

Several additional studies were focused on ULIRGs at $z \sim 2$ but they featured additional selection criteria [152, 34, 72] so we should compare their and our results with caution. For example, Farrah et al. [72] used a sample of ULIRGs with $S_\nu(24\,\mu\mathrm{m}) > 400\,\mu$Jy which in addition had a spectral peak in the 4.5 $\mu$m and 5.8 $\mu$m IRAC bands, corresponding to the redshifted stellar 1.6 $\mu$m peak. The 4.5 $\mu$m-peak were estimated to be at $1.5 < z < 2.0$; their derived correlation length was $r_0 \sim 9\, h^{-1}$ Mpc. The 5.8 $\mu$m-peak sources are at $2 < z < 3$ and

their angular clustering corresponded to the correlation length $r_0 \sim 14\,h^{-1}$ Mpc. These very high values of the correlation lengths would imply that the selected galaxies are located in extremely massive DM halos, $M \sim 4 \times 10^{13}\,h^{-1}\,M_\odot$. It appears, however, that the clustering results are dominated by the detection of the angular clustering at very small scales, which can bias the results if one assumes a fixed slope for the angular correlation function fit. In another work, a sample of Dust Obscured Galaxies ("DOGs", [64]) was selected. DOGs are mid-IR luminous ($S_\nu(24\,\mu\text{m}) > 300\,\mu$Jy) and optically faint ($R - [24] > 14$) galaxies estimated to be at $z \sim 2$. Their measured correlation lenght is $7.4\,h^{-1}$ Mpc, corresponding to the dark matter halos with masses $M_{\rm tot} = 10^{12}\,h^{-1}$ Mpc [34]. Our results are very similar to these measurements which is not surprising because DOGs are a subpopulation of bright 24 $\mu$m-selected galaxies.

Models of galaxy formation suggest that DOGs and submillimetre galaxies ("SMGs") form by mergers of massive ($M_{\rm tot} \sim 10^{12-13}\,h^{-1}M_\odot$) galaxies (see a recent paper by Narayanan et al. [173] and references there) and may represent different phases in the evolution of a merging system. It would be interesting to compare the clustering of SMGs and other classes of ULIRGs, but, unfortunately, the present estimations of the SMGs correlation length and halo mass are too uncertain [23, 217, 256, 148, 50]. Viero et al. [250] reported a clustering strength of $r_0 = 4.9 \pm 0.7\,h^{-1}$Mpc for the *BLAST* $1.3 < z < 2.2$ sources detected at 250 $\mu$m, corresponding to the minimum host halo mass $M_{\rm tot} \sim 2 \times 10^{11}\,h^{-1}M_\odot$. This result supports a scenario in which the SMGs are lower mass sources undergoing burst of star formation with a top heavy IMF [11, 131]. The relation with our 24 $\mu$m selected galaxies is doubted because of a vary different level of the inferred correlation length.

## 18 Conclusions

We presented the clustering analysis of 24 $\mu$m emitting ($S_\nu(24\,\mu\text{m}) > 310\,\mu$Jy) galaxies detected in the Lockman Hole — one of the largest fields in the Spitzer/SWIRE survey. The large number of sources ($\sim 20000$) and the size of the field allowed us to detect the clustering signal with high level of significance and probe large angular scales. Due to the lack of direct redshift measurements for

the objects in the Lockman Hole sample, we used the optical and near-IR data to separate the sample into high-redshift and low-redshift galaxies. The selection criteria as well as the redshift distributions for color-separated subsamples were empirically obtained using the catalogues of GOODS 24 $\mu$m sources [207] whose redshifts were measured spectroscopically or estimated from multi-band photometry. Using a power-law approximation to the correlation function, we derived the spatial correlation length, $r_0$, through the Limber inversion. We found $r_0 = 4.74 \pm 0.16\,h^{-1}$ Mpc and $r_0 = 7.87 \pm 0.63\,h^{-1}$ Mpc for $z_{\rm mean} = 0.68$ and $z_{\rm mean} = 1.72$ populations, respectively. From the comparison with the clustering of dark matter halos in the *Bolshoi* cosmological simulation, we conclude that the 24 $\mu$m-selected starburst galaxies are found in 20% of halos with masses $M_{\rm tot} \gtrsim 5 \times 10^{11}\,h^{-1}\,M_\odot$ and $M_{\rm tot} \gtrsim 2 \times 10^{12}\,h^{-1}\,M_\odot$ at low and high redshifts. The estimated infrared luminosities showed that our 24 $\mu$m-selected galaxies may belong to populations of distant ULIRGs and local LIRGs. The comparison with previous analyses of star-forming galaxies revealed that our low-$z$ subsample has similar clustering properties to those of blue and UV-selected galaxies in the local Universe. The high-redshift subsample is similar in properties to the dust-obscured $z \sim 2$ galaxies but shows a higher correlation length, correspondingly implying more massive dark matter halos, than those for the currently published samples of Lyman Break Galaxies and submillimetre galaxies. Based on the clustering analysis, we can conclude with a high level of confidence that our low-$z$ and high-$z$ 24 $\mu$m-selected galaxies represent different populations of objects found in differently-sized dark matter halos. Their high level of mid-IR luminosities may be caused by similar physical processes (e.g., triggered by mergers or interactions), but occurring in different environments. Further information can be obtained by studying in detail the dependence of clustering properties on the IR-luminosity at each redshift. This will be the subject of a future work.

In the next chapter, we investigate the clustering properties of a galaxy population in many respects opposite to the actively star-forming galaxies. These objects are massive, passively evolving galaxies which already formed most of their stellar mass $M^* \sim 10^{11} M_\odot$ before a redshift of $\sim 2$.

# V Massive high-redshift galaxies detected in near-infrared

## 19 Summary

We present the clustering analysis of 982 massive, high-redshift ($z_{\rm mean} = 2.2$) galaxies selected by Berta et al. [20] in the SWIRE ELAIS-S1 field. These objects, called "IR-peakers", likely represent the most massive galaxies at $z \sim 2$, with the estimated stellar mass $M^* \approx 1.6 \times 10^{11}\, M_\odot$ [18]. Our modeling of the angular correlation function is based on the correlation function of dark matter halos in the *Bolshoi* cosmological simulation [121]. We find that the key observable characteristics of IR-peakers — their stellar mass, number density, and clustering, are best explained if IR-peakers reside in those $10 - 25\%$ of dark matter halos with $M_{\rm tot} \approx [0.5 - 1] \times 10^{13} M_\odot$, in which the star conversion efficiency reaches $\sim 10 - 20\%$. We also show that IR-peakers can not be located only in the central galaxies of parent dark matter halos, or only in the subhalos.

## 20 Introduction

Massive galaxies at $z > 2$ are fascinating objects. They accumulated large quantities (comparable to the stellar mass of typical low-redshift galaxies, e.g. the mass of the Milky Way is $\sim 6 \times 10^{10}\, M_\odot$) of stellar material over a relatively short period of time (the age of the Universe at $z = 2$ is 3.3 Gyr and this redshift is near the epoch of the peak star formation). Because they are such "peculiar" objects, the data on massive high-$z$ galaxies are extremely useful for developing models of galaxy formation, and for obtaining additional clues to the long-standing problems in the modern astrophysics such as does "nature or nurture" play a

more important role in the formation of low-$z$ elliptical galaxies. The properties of high-redshift galaxies probably depend on a large number of factors, from the overall assembly history of massive dark matter halos to the importance of energy feedback from early SMBH and supernovae. The progress in studies of these objects will help to put observational constraints on these processes as well.

In recent years, there has been an explosion in developing different methods for selecting high-redshift galaxies. In the optical, distant galaxies can be selected photometrically by red colors in the I, J, and K filters [82, 248], or by detecting features consistent with the redshifted Lyman-break in the spectrum (this technique works for $2.5 < z < 5$, see [89] for a review) or the Ly-$\alpha$ emission line [107, 177]. Another popular method selects massive star-forming and passive galaxies at $z \sim 2$ using BzK colors [55].

More relevant to our work is the selection of high-redshift objects in the infrared. As we discussed in Chapter IV, a significant fraction of objects simply detected in the 24 $\mu$m Spitzer band are at $z = 1.5-3$, and simple optical selection criteria applied in addition to the 24 $\mu$m detection results in an efficient selection of high-$z$ objects. Other works which used similar technique (together with a near-IR condition in some cases) include Magliocchetti et al. [152], Farrah et al. ("distant ULIRGs", [73]), Brodwin et al. ("DOGs", [34]) and many others (see Chapter IV for details). The 24 $\mu$m emission is mostly related to the star formation or AGN activity in the host galaxies. Therefore, the far-IR selection works great for selecting various types of active galaxies, but not specifically tuned for selecting the objects by their stellar mass.

The Spitzer Infrared Array Camera (IRAC, Fazio et al. [74]) samples the rest frame near-IR ($1-2\,\mu$m) light of distant galaxies ($z \sim 1.5-3$). A near-IR selection not only directly probes the low-mass stars dominating the baryonic mass of a galaxy, but also is minimally affected by dust extinction. Therefore Spitzer is best suited to the study of the stellar content of galaxies up to $z = 3$ [18]. At rest frame near-IR, spectral energy distribution (SED) of galaxies has a prominent feature — a broad peak — caused by a minimum in the $H^-$ opacity in atmospheres of late types stars. This 1.6 $\mu$m peak, redshifted to the IRAC bands, is a good indicator of stellar mass assembled in the distant Universe. In addition, the selection based on the rest frame near-IR colors is less affected by

dust extinction, stellar ages and metallicities than most other techniques [109].

In a recent work, Berta et al. [18] selected a sample of objects, the so-called IR-peakers, by existence of a flux maximum in the Spitzer-IRAC 5.8 $\mu$m band. The IR-peakers are unique objects, because these are the galaxies with the highest stellar mass (see [18] and § 21.3 below) which exist at $z \sim 2.2$, not very far from the epoch of active star formation (see, e.g. Fig.3 in Madau [146]). Moreover, not only IR-peakers are very massive, but many of them also show signs of old stellar population at $z = 2$. They seem to be very special objects, and it is interesting to obtain additional insights on their nature using the information from their clustering properties. This is the goal of the work presented below.

## 21 Properties of galaxies in the selected sample

### *21.1 Selection criteria*

Using a careful photometric analysis of the infrared sources detected in the SWIRE ELAIS-S1 field, Berta et al. [20] selected a sample of galaxies in which the 1.6 $\mu$m (rest frame) "stellar" peak in spectral energy distribution is shifted into the Spitzer/IRAC passband. The main principle of selecting IR-peakers is illustrated in Fig. 28. This Figure shows an example of SED for a galaxy at $z = 2.455$ (photometric redshift). Overplotted on the observed fluxes, is the best fit solution of the mixed stellar population spectrophotometric synthesis [18]: the dashed (blue) line corresponds to the young or intermediate-age (< 1 Gyr) stellar populations, and the dotted (red) line shows the contribution from old ($\geq$ 1 Gyr) stars. The solid line provides the total emitted light in the optical and near-IR wavelength range. Longward of 5 $\mu$m (restframe), the long-dashed line (light blue) is the starburst template used to model the IR emission from dust, heated by young stars. This figure illustrates that at $z \sim 2$, the 1.6 $\mu$m peak is redshifted into the IRAC's 4.5 $\mu$m and 5.8 $\mu$m bands, and thus massive galaxies around those redshift can be efficiently selected by their near-IR colors.

IR-peakers were selected in the 4.0 deg$^2$ section of the SWIRE ELAIS-S1 field, in which the optical data from ESIS VIMOS survey were available [20][1]. All

[1]Berta et al. have used the optical optical data for removal of low-$z$ interlopers and for refined

selected sources have fluxes above the $3\sigma$ threshold for the 5.8 $\mu$m IRAC channel $S_\nu(5.8) > 25.8\,\mu$Jy, and satisfy the following condition:

$$S_\nu(3.6) < S_\nu(4.5) < S_\nu(5.8) > S_\nu(8.0). \tag{5.1}$$

The derived sample includes some sources not detected in the 8.0 $\mu$m band but their 8.0 $\mu$m $3\sigma$ upper limit (28.8 $\mu$Jy) is consistent with this selection criteria. Figure 28b shows location of IR-peakers in the IRAC color-color diagram [132, 234]. Red triangles represent the IR-peakers with the flux maximum in the 5.8 $\mu$m band. Blue circles (filled and open) show the objects with maximum flux in the 4.5 $\mu$m band (i.e., those with $S_\nu(3.6) < S_\nu(4.5) > S_\nu(5.8)$). These "4.5 $\mu$m peakers" are not considered in our work, because they can be contaminated by low-$z$ interloppers (see below). The small dots represent the general galaxy population in ELAIS-S1 SWIRE-ESIS field: light blue having a four band detection, and orange having a 8.0 $\mu$m upper limit only. Redshift tracks for a variety of templates are also shown as a function of redshift: a Seyfert-1 (dotted lines, Mrk 231, [83]), a Seyfert-2 (dot-dashed, IRAS 19254-7245, [17]), a starburst galaxy (short-dashed, M82,[226]), a spiral (long-dashed, M51,[226]) galaxy. The tracks are limited to the range $z = 1 - 3$, apart for the starburst one, which is extended down to $z \simeq 0$.

It is worth noting the selection based on the presence of the IR-peak in the observed SED automatically avoids contamination by bright AGNs which are characterized by a power-law-like (monotonic) SED in the near IR, even though the objects with the sub-dominant AGN component in the spectrum still can be present [19]. A stronger concern [19], is that the IR-peak-like colors could in principle be produced not only by the redshifted stellar 1.6 $\mu$m feature, but also by a strong 3.3 $\mu$m Polycyclic Aromatic Hydrocarbon (PAH) emission presented in SED of starburst galaxies at redshifts $z = 0.3 - 0.5$. The real 1.6 $\mu$m peak lies shortward of the IRAC channels, resulting in a blue NIR-IRAC observed color (Fig.28b, where blue open circles represent low-redshift incomers of 4.5 $\mu$m-peak sources and dashed straight line is optical-IRAC condition separating the two galaxy populations with different nature of IRAC-peak). Therefore, the constrain

modeling of the SEDs for individual objects

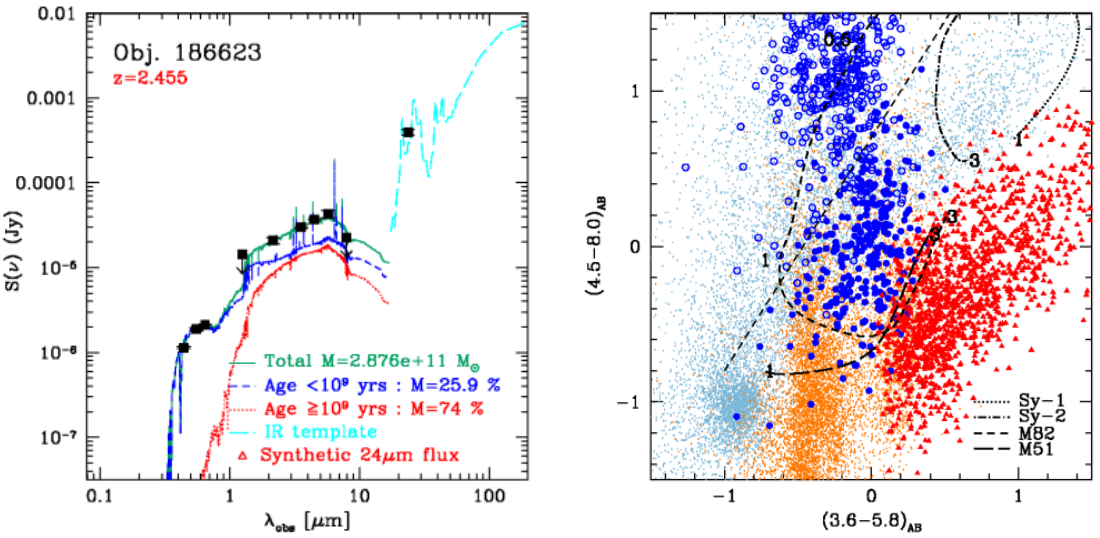



Figure 28. *(a)* Example of SED fits of IR-peak sources. The dashed line represents the contribution to the best fit model by young and intermediate-age (< 1 Gyr) stars, while the dotted line is the light emitted by older stars (age ≥ 1 Gyr). The solid line is the total best fit emission up to 5 $\mu$m (restframe). The long-dashed line longward of 5 $\mu$m (restframe) is the adopted starburst template re-processing the UV-optical light absorbed by dust to the IR. *(b)* Position of IR-peak galaxies in the IRAC color space [132, 234]. Filled triangles and circles represent 5.8 $\mu$m-peakers and 4.5 $\mu$m-peakers respectively. Open circles are 4.5 $\mu$m-peakers identified as low-redshift interlopers on the basis of $(I-3.6)_{AB}$. Small dots trace the general SWIRE-ESIS population; orange dots are general sources with no 8.0 $\mu$m detection (upper limits are used). Redshift tracks for different templates are plotted: M82 (dusty starburst) and M51 (spiral) belong to the *GRASIL* library [226]; the Sy1 and Sy2 templates represent Mrk231 and IRAS19254-7245 [83, 17]. Redshift values are reported along tracks. The diagonal line marks the color conditions to identify low-redshift interlopers (see for details [20]).

applied for selecting objects in the present work, $S_\nu(3.6) < S_\nu(4.5) < S_\nu(5.8)$, is likely to exclude any low-redshift interlopers.

### *21.2 Redshift distribution*

Since the spectroscopic redshifts are unavailable for most of the IR-peakers[2], Berta et al. [18] estimated their redshifts photometrically. They used the *Hyper-z code* [26] with the template library based on *GRASIL* [226] models of spiral and elliptical galaxies, in which the M82 and Arp220 templates were upgraded with PAH mid-IR features [78] and a ULIRG template [17]. The SED fits used the optical $(B, V, R)$, $J, K_s$ and IRAC $(3.6-8\,\mu\text{m})$ data. The choice of the spectral templates in the library and the allowed range of extinction were tested using

[2]The spectroscopic surveys covered only a central 0.6 deg$^2$ region of ELAIS-S1 field and targeted a small number of sources, mainly AGNs and low-redshift ($z \leq 1$) galaxies [213].

available spectroscopic redshifts of galaxies in the ELAIS-S1 field. This analysis also provided a test of photometric redshift uncertainties at least to $z \simeq 1$. A typical uncertainty of $z_{\rm ph}$ was found to be $\approx 10\%$, and 50% of the data are within $(1 + z_{\rm ph})/(1 + z_{\rm sp}) = 1 \pm 0.076$ of the scatter area.

While the *Hyper-z* performance was optimized for work with the IR-peak-like SEDs, the expected accuracy of photometric redshifts is poor for the SEDs of the power law type (AGNs). This class is particularly delicate to fit with a template-based procedure and photometric redshifts are typically mis-interpreted in 50% of the cases [19], because of the lack of sharp features in AGN SEDs. Therefore, all sources a type-1 AGN spectroscopic classification and all those showing a power-law like SED in the IRAC band were excluded from the analysis to minimize the contamination by AGNs. Note that many AGNs were excluded automatically because the initial selection criterion based on existence of the IR-peak in the spectrum selects against power law spectra (see above).

The redshift distribution of IR-peakers is shown in Fig.29a. As expected from the object selection criteria, it peaks around $z = 2.2$ (and this is also the mean redshift for the sample). Further, the distribution is rather narrow ($\Delta z \approx 0.5$ at the FWHM), and thanks to that we are able to measure the angular correlation function even with this small sample (see § 22 below).

To make conversion of the angular correlation function into the spatial correlation function and vice versa, we need an analytic fit to the observed redshift distribution (§ 5). We find that a good fit is provided by a fifth-order polynomial in the $\log N - z$ coordinated (solid red line in Fig. 29a). Using this fit, we also can find the number density of IR-peakers near the mean redshift of the sample,

$$n_{\rm pks} = \frac{dN/dz}{dV/dz} = 0.087 \times 10^{-3}\, h^3\, {\rm Mpc}^{-3}, \tag{5.2}$$

where $dV/dz$ is the comoving volume within the survey area [103]. This number density will be used for comparison with the number density of dark matter halos in the cosmological simulation.

We note also that IR-peakers represent a population distinct from another class of high-$z$ IR-selected objects, the so-called ultraluminous infrared galaxies (ULIRGs) selected by Farrah et al. [72] in other SWIRE survey fields. We checked that only 15% of IR-peakers have a flux in the 24$\mu$m *Spitzer*/MIPS channel in

excess of 400 $\mu$Jy, the selection criterion used by Farrah et al. Furthermore, only 28% of IR-peakers are detectable in the 24$\mu$m band at a 3$\sigma$ level.

### *21.3 Stellar masses*

Berta et al. [18] also estimated stellar masses for all IR-peakers from the available photometry using the *Sim-Phot-Spec* code [16, 190]. This code performs mixed stellar population spectrophotometric synthesis. The galaxy SEDs are modeled by a combination of single/simple stellar populations (SSPs) of solar metallicity and different ages assuming Salpeter initial mass function between 0.15 and 120$M_\odot$. Every SSP has its own age-dependent value of star formation rate and correction for the dust extinction. The number of SSPs involved in a fit depends on the redshift of a given source. As Berta et al. [18] show, the emission of 30% of the sources is dominated by stars older than 1 Gyr and in the majority of remaining cases the old stellar populations considerably contribute to the observed SEDs.

The output of the *Sim-Phot-Spec* modeling is the assembled stellar mass in individual galaxies. Typical uncertainties, — mainly related to degeneracies in the star formation history space, — of the estimated stellar masses are in the range 0.1–0.3 dex, depending on multiwavelength coverage. The distribution of stellar masses of IR-peakers is shown in Fig. 29b. Due to the adopted shallow flux cut in the 5.8 $\mu$m band, the SWIRE IR-peaker sample automatically contains only very massive galaxies. Most of the selected sources have stellar masses in excess of $10^{11}M_\odot$; the peak of the distribution is centered at $M_* \sim 1.6 \times 10^{11}M_\odot$.[3] As Berta et al. [18] have shown, IR-peakers occupy the very highest end of the galaxy stellar mass function at $z \sim 2$ (Fig. 30).

[3] All masses in this Chapter are reported for a Hubble constant of $h = 0.71$. This is unlike the rest of this thesis where the masses are quoted for $h = 1$ and the explicit $h$-dependence is given, reflecting the accepted practice in the literature. A different approach is taken here because we discuss ratios of the observed stellar masses and inferred total mass, which have a different $h$-dependence, $h^{-2}$ and $h^{-1}$ for $M_*$ and $M_{\rm tot}$, respectively. In this case, it is better to use the actual value of $h$ now sufficiently well constrained by the body of the cosmological data (e.g., [124]).

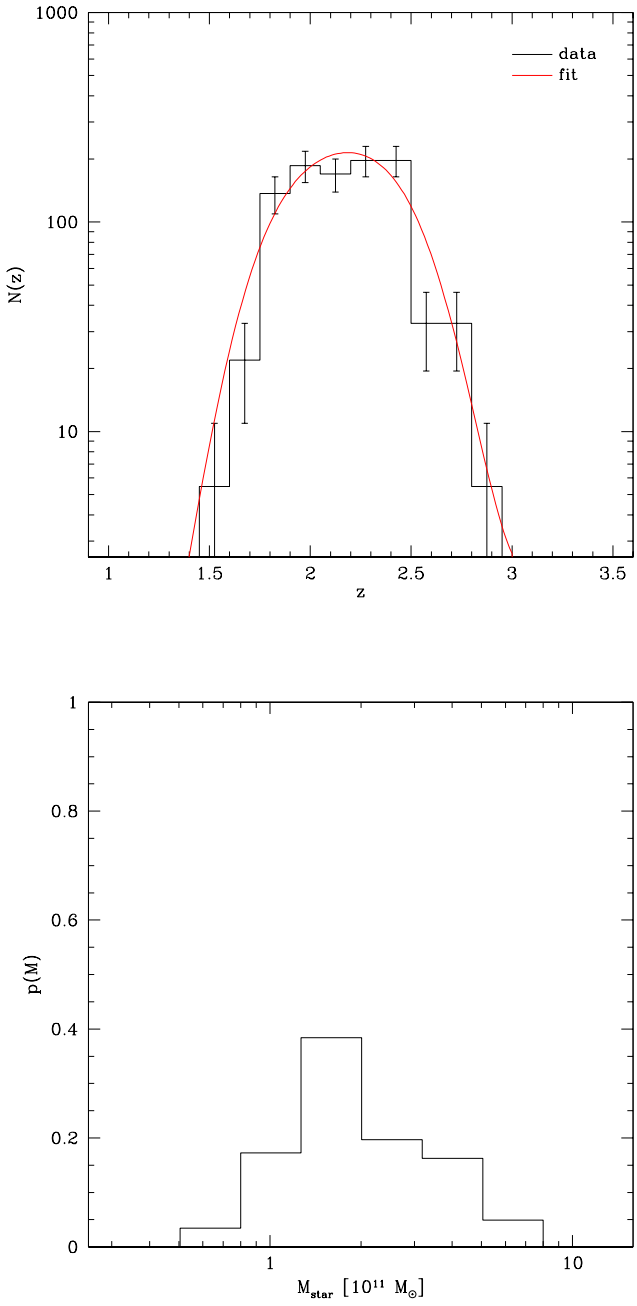



Figure 29. Photometric redshift *(a)* and stellar mass *(b)* distribution of IR-peakers.

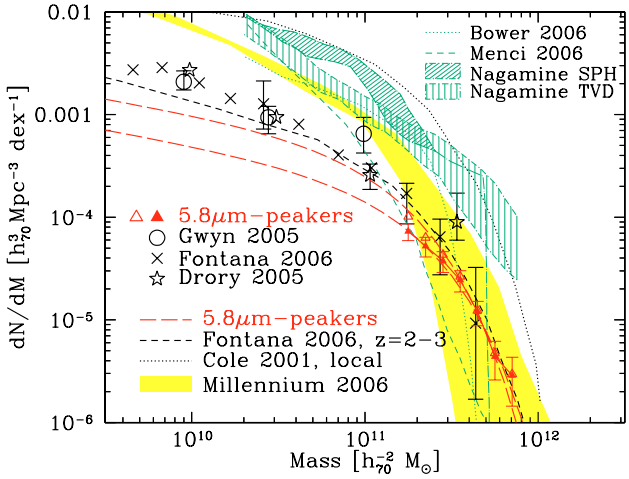



Figure 30. Comparison of stellar mass function of IR-peakers with the literature measurements at $z = 2 - 3$ (reproduced from [18]). The IR-peakers (red data points) populate the highest $M_*$ tail of the high-$z$ galaxy population.

## 22 Clustering measurements

Even though the photometric redshift estimates are available for all IR-peakers, their accuracy is insufficient for a direct measurement of the spatial two-point correlation function or even of its projection on the sky plane (see § 3). Therefore, as in the case of mid-IR selected star forming galaxies (Chapter IV), we resort to using the measurements of the *angular* correlation function and its analysis via the Limber inversion (§ 5).

We use 982 IR-peakers selected from 3.6 deg$^2$ subsection of the SWIRE-ESIS ELAIS-S1 field. The overall properties, such as the distribution of redshifts and estimated stellar masses, for this sample are described above. The distribution of IR-peakers on the sky is shown in Fig. 31.

The two-point angular correlation function of IR-peakers was calculated using the Landy & Szalay estimator (eq. 2.2). The random catalog was generated taking into account the region mask (Fig. 31), and contained a factor of 100 more points than there were galaxies in our sample. The data and random pairs were counted in the angular bins of equal log-width, $\Delta \log \theta = 0.2$, on scales $< 1°$. To estimate statistical uncertaintes for the correlation function in each bin, we used the Poisson expression which includes data clustering (eq. 2.4). The derived angular correlation function is shown in Fig. 32. We confidently

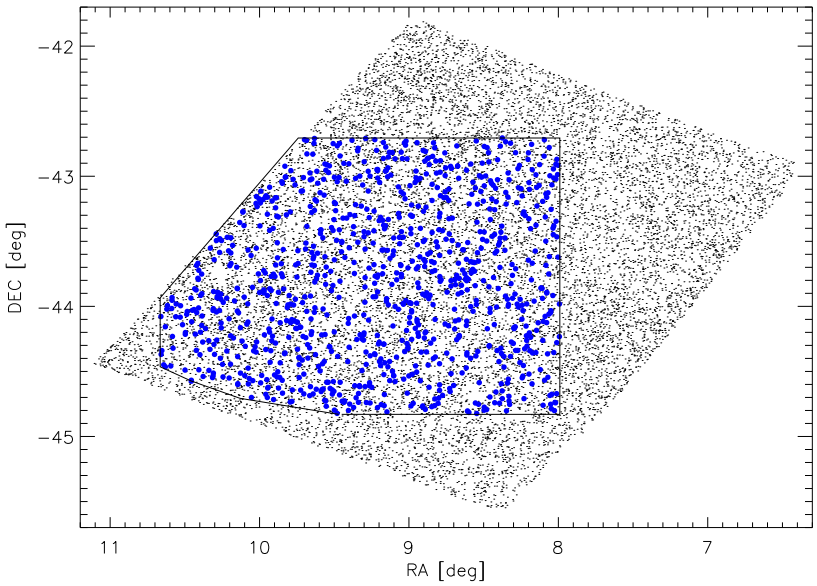



Figure 31. Sky distribution of 982 IR-peakers used for the clustering analysis (blue filled circles within boundaries). For comparison, the distribution of the 24 $\mu$m sources with $S_\nu(24) > 400\,\mu$Jy in the SWIRE ELAIS-S1 field is shown by black points.

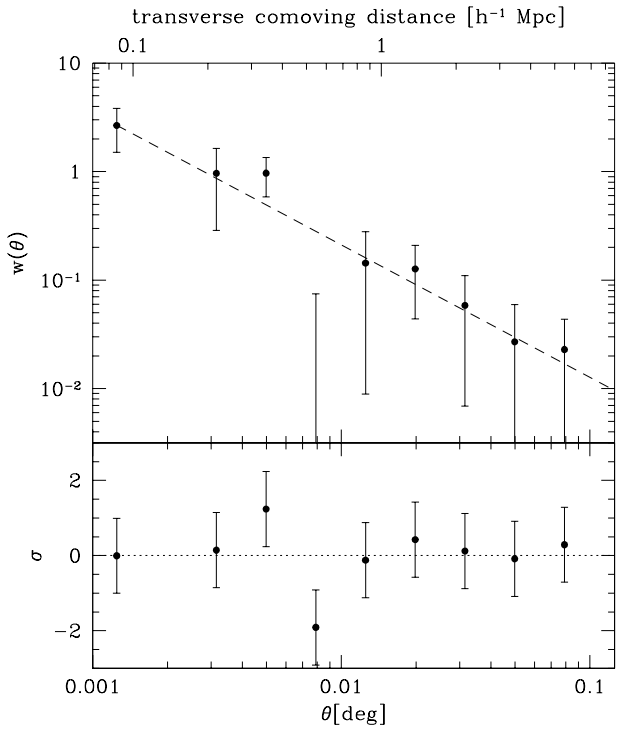



Figure 32. Two-point angular correlation function of IR-peakers. The dashed line is the best fit of a power law model. Corresponding to angular scales the transverse comoving distances at $z = 2.2$ are given in the upper axis. The bottom panel represents the residuals to the power law fit.

detect positive clustering of the IR-peakers, primarily at the small scales, with an overall statistical significance of $\gtrsim 4\sigma$. The detection of positive clustering at the smallest separations, $0'' < \theta < 9''$, is robust because it is based on the existence of 10 pairs in which the source separations exceed the IRAC angular resolution ($\sim$ 1.66″ FWHM in the 3.6 $\mu$m band). At $\theta > 360''$, the angular correlation function is consistent with zero within the uncertainties.

Within the statistical uncertainties, a power law fit,

$$w(\theta) = (\theta/\theta_0)^{1-\gamma} - \mathrm{IC}, \tag{5.3}$$

where IC is the so-called "integral constraint" term (§ 2), provides an acceptable model to the data (dashed line in Fig. 32). The best fit parameters are $\theta_0 = 10.1'' \pm 2.5''$ and $\gamma = 2.2 \pm 0.2$.

### *22.1 Limber inversion of the power law fit to $w(\theta)$ for IR-peakers*

Under a common assumption that the spatial correlation function is described by the power law, $\xi(r) = (r/r_0)^{-\gamma}$, the clustering length, $r_0$ can be obtained from the power-law fit to the angular correlation function via the Limber equation (see § 5, eq. 2.23). We show below that this approach is too simplistic for modeling of our measurements of the IR-peakers clustering, but we provide the results for completeness.

Using the best power law fit to the angular correlation function and the redshift distribution model shown in Fig. 29a, the Limber inversion gives $r_0 = 5.63 \pm 0.65\,h^{-1}$Mpc (comoving), where the uncertainties correspond to the statistical uncertainties in the $w(\theta)$ measurements, assuming the slope fixed to its best-fit value $\gamma = 2.2$.

An important source of uncertainty in estimating of $r_0$ by the Limber equation is that related to the source redshift distribution. To test how the photometric redshift uncertainties could affect the derivation of $r_0$, we computed the Limber inversion using the same $w(\theta)$ model but varying the width and location of the $N(z)$ model. This analysis showed that shifting the peak of the redshift distribution (this corresponds to a systematic bias in the photometric redshifts) within a reasonable range has little effect on the derived $r_0$. Changing the width of the distribution has a stronger effect. For example, if the width is reduced

by 30% (15%), the correlation length is reduced by 14% (7%). This level of variations in the $N(z)$ distribution is conservative given estimated uncertainties of the photometric redshifts (see above), and it is reassuring that the associated changes in the correlation length are still within the purely statistical errors associated with the $w(\theta)$ uncertainties.

### 22.2 *Inadequacy of the power law model for description of the clustering of IR-peakers*

According to the theory of the cosmic structure formation, galaxies are formed and then follow the distribution of dark matter halos. The dark matter distribution is faithfully reproduced by the current cosmological simulations (e.g., [232]), and the outputs of such simulations can be compared with the galaxy distribution data. In fact, a very simple paradigm, the so-called "abundance matching technique", in which one associates galaxies above a given luminosity threshold with dark matter halos above some total mass threshold[4], has been very successful in simultaneously describing a relation between the number density and the correlation length of galaxies (see, e.g., [48]). An attempt to apply this technique to the power law model of the observed clustering of IR-peakers results in a failure. Using the analysis of the cosmological simulations described below, we find that a correlation length of $5.6\,h^{-1}$ Mpc, derived from the power law fit to $w(\theta)$, corresponds to a mass threshold of only $M_{\rm tot} = 2.0 \times 10^{12}\,M_\odot$. This total mass scale is too low for the characteristic *stellar* mass of IR-peakers (§ 21). A convenient way of comparing the total and stellar mass for a given object is through the so-called star formation efficiency coefficient, $f_* = M_*/M_{\rm tot}\,(\Omega_b/\Omega_M)^{-1}$ where $\Omega_b/\Omega_M = 0.17$ is the cosmic average baryon fraction now precisely determined by the WMAP observations [124]. With the above value of $M_{\rm tot}$ for IR-peakers, $f_*$ should be in excess of 50%, far higher than both observational and theoretical estimates for present-day galaxies of similar mass (5%, [96], see their Fig. 3). Furthermore, the number density of such low-mass halos is significantly higher than that of the IR-peakers, $n_{\rm halo} \approx 50\,n_{\rm pks}$. Therefore, the observed angular clustering appears to be in a disagreement with the high stellar mass and low

[4]The relation between galaxy luminosity and total mass is assumed to be monotonic but does not have to be, e.g., a simple proportionality.

number density of IR-peakers. In what follows, we show that these problems can be alleviated by applying a more realistic model to the angular correlation function.

## 23 Modeling of the angular correlation function based on on results of cosmological simulations

With the immense statistical power of the APM and SDSS surveys, it became clear [179, 263] that the observed correlation functions of galaxies do not follow a pure power law over a wide range of scales. A straightforward interpretation of these deviations is provided by the clustering properties of dark matter halos identified in the cosmological simulations. The correlation function of dark matter halos shows a two-component shape (Fig. 33): the first — the so-called "1-halo" term — is prominent at small separations and corresponds to objects within a single gravitationally-bound halo, while a larger scale component — a "2-halo" term — is contributed by objects in separate distinct halos (see chapter III, § 9.1). Depending on the mass scale of the dark matter halos associated with galaxies of a given type, and of the galaxies environmental properties, the 1- and 2-halo terms are present in different proportion in the galaxy correlation function, and that correlation function shows stronger or weaker deviations from the power law. In the Figure 33, the real spatial correlation function is presented for three samples of dark matter halos, drawn from *Bolshoi* simulations, with limited maximum circular velocities, $V_{\rm min} = 320, 250$, and $180\,{\rm km/s}$. *Bolshoi* is a very high resolution cosmological simulations, presented in a Chapter III.

Since it is expected that the correlation function deviates from the power law — and such deviations are indeed observed — the correlation length, $r_0$ derived from a power law fit to the noisy data can be biased depending on the range of scales in which the correlation function measurement uncertainties are the lowest. For example, if the 1-halo term is prominent and the correlation function measurements are sufficiently accurate only on small scales ($\lesssim 1\,h^{-1}\,{\rm Mpc}$), then a power-law fit with a free $\gamma$ would result in a steep slope and underestimation of the correlation length, $r_0$ [127]. A fit with a slope fixed at $\gamma = 1.8$, a value typically observed at scales around $6\,h^{-1}\,{\rm Mpc}$, on the contrary, will provide an

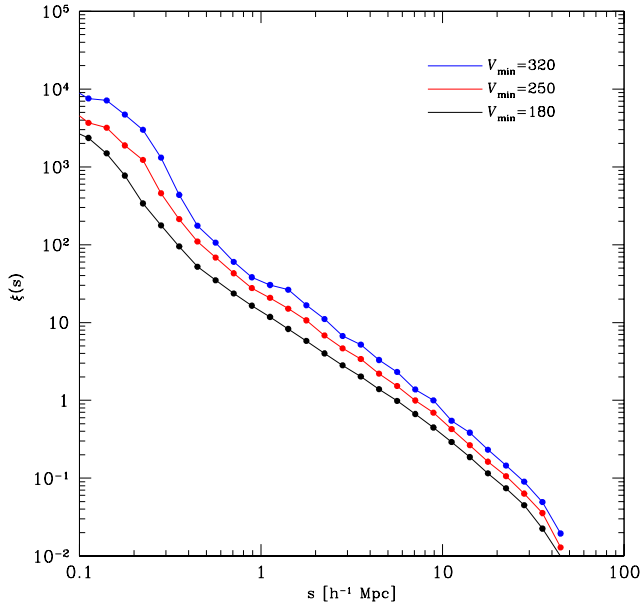



Figure 33. Spatial correlation function of the dark matter subhalos derived from *Bolshoi* simulation (§ 7.2) at $z = 2.0271$. The subhalo catalogs were selected for three thresholds of the maximum circular velocity, $V_{\rm min} = 320$, 250, and 180 km/s (top to bottom).

overestimated value of $r_0$ [200].

It is very likely that this effect biases the correlation length derived from a power law fit to the angular correlation function in our case. Indeed, the middle point of our angular range, 0.01°, corresponds to a $0.6\,h^{-1}$ Mpc comoving separation at $z = 2$ (Fig. 32). This is just the separation within which a 1-halo term is expected to dominate the correlation function (Fig. 33). A typical correlation length of galaxies, $6\,h^{-1}$ Mpc corresponds to angles of $\sim 0.1°$, which is at the boundary of our angular range and where the $w(\theta)$ measurement uncertainties are very large. Therefore, instead of using the power-law fits, we compute the angular correlation function models for our range of $\theta$ using numerical simulation outputs.

### *23.1 Derivation of the $w(\theta)$ model*

The peak in the redshift distribution of IR-peakers is at $z = 2.2$ (Fig. 29a); to compute the $w(\theta)$ models at $z = 2.2$, we use the nearest *Bolshoi* simulation snapshot at $z = 2.0271$, ignoring possible evolutionary effects within the IR-

peakers $N(z)$ distribution. The validity of such an approximation is based on two considerations. First, from Figure 26, we see that that the evolution of the clustering strength of halos for a given $V_{\rm min}$ is mild within the redshift range of IR-peakers and certainly within statistical uncertainties on the derived $V_{\rm min}$ (see below). This means that when we select the halos by $V_{\rm min}$, we can compare clustering of DM halos at a fixed redshift near 2.0 with the clustering of the IR-peakers throughout their redshift distribution (due to its narrowness). The second consideration is a weak evolution of the DM halo number density for a fixed $V_{\rm min}$. This is illustrated in Fig. 6 (see 7.2, [121]) which shows the dependance of the cumulative number density of distinct halos on the maximum circular velocity at different redshifts. For $v_{\rm max} > 268$ km/s ($V_{\rm circ}$ in the notation used in [121]), the number density of halos at redshifts $z = 2$ and $z = 3$ is almost the same. Therefore, not only the clustering properties but also the number density of DM halos evolves very mildly if we select them by the maximum rotational velocity, as we do in this work.

Using the correlation function of the DM halos in the Bolshoi output, we can compute the corresponding model for the angular correlation function appying the full Limber transformation [196, 48] (see § 5):

$$w(\theta) = \frac{2}{c}\,\frac{\int\limits_0^\infty dz\, N(z)^2\, H(z) \int\limits_0^{x_{\rm max}} dx\, \xi\left(\sqrt{[D_{\rm M}(z)\theta]^2 + x^2}\right)}{\left[\int\limits_0^\infty dz\, N(z)\right]^2}, \tag{5.4}$$

where $D_{\rm M}(z)$ is the transverse comoving distance to redshift $z$ ($D_{\rm M} = D_{\rm c}$ in the flat universe, see, e.g., [103]), $H(z)$ is the Hubble parameter at that redshift, and $N(z)$ is the redshift distribution of the objects. We assume that $N(z)$ corresponds to the observed redshift distribution of IR-peakers shown in Fig. 29a and described in § 21.2. Analysis of the Limber inversion for the power law fit (§ 22.1) shows that the uncertainties associated with our knowledge of $N(z)$ are small compared to those associated with statistical uncertainties in $w(\theta)$; therefore we will ignore them in the analysis described below.

The observational selection of IR-peakers is very closely related to the total stellar mass in these galaxies. It has been argued [128, 172, 48] that of the commonly used parameters of dark matter halos, the maximum circular velocity,

$v_{\rm max}$, should be the one most tightly related to the stellar mass in the halo (see § 7 for more details). Therefore, we derived models for the IR-peakers correlation function using the dark matter halos selected above some threshold value of $v_{\rm max}$. This minimum value of $v_{\rm max}$ above which a dark matter halo or subhalo can contain an IR-peaker, is referred to as $V_{\rm min}$ hereafter.

The 3D correlation function models were derived for a grid of $V_{\rm min}$ threshold in the range $160 < V_{\rm min} < 460$ km/s. The choice of the lower limit is motivated by the consideration that for $v_{\rm max} = 160$ km/s, the total halo mass is $M_{\rm tot} \approx 4.0 \times 10^{11}\, M_{\odot}$ (Fig. 5), only a factor of 2.5 higher then the mean estimated stellar mass of IR-peakers. It is highly unlikely that IR-peakers are associated with dark matter halos of a lower mass. For $V_{\rm min} = 460$ km/s, the *Bolshoi* output at $z = 2$ contains only 975 objects in the $(250\,h^{-1}\,{\rm Mpc})^3$ box; for higher velocity thresholds, the number of halos is insufficient for derivation of the accurate models of the 3D correlation function.

For each $V_{\rm min}$ threshold, the 3D correlation function was tabulated in a range of scales $0 < r < 50\,h^{-1}$ Mpc, and then used in evaluation of the inner integral in the full Limber inversion by eq. (5.4). This integral was evaluated numerically, which $\xi(r)$ interpolated between the tabulated values in the $\log r - \log \xi$ space. The integration limit of the integration was set to be $x_{\rm max} = 50\,h^{-1}$ Mpc.

The procedure described above provides the angular correlation function models as a function of the circular velocity threshold, $V_{\rm min}$. The models were derived for $\theta < 0.8^{\circ}$ and compared with the data through the $\chi^2$ computed as

$$\chi^2(V_{\rm min}) = \sum \frac{(w(\theta) - w_{\rm mod}(\theta, V_{\rm min}))^2}{(\delta w(\theta))^2}. \qquad (5.5)$$

Minimizing the $\chi^2$-function, we estimate the circular velocity of the dark matter halos associated with IR-peakers and determine their total mass and comoving number density. Then these parameters are compared with the stellar mass and the observed number density of IR-peakers. The obtained results are discussed in the next section.

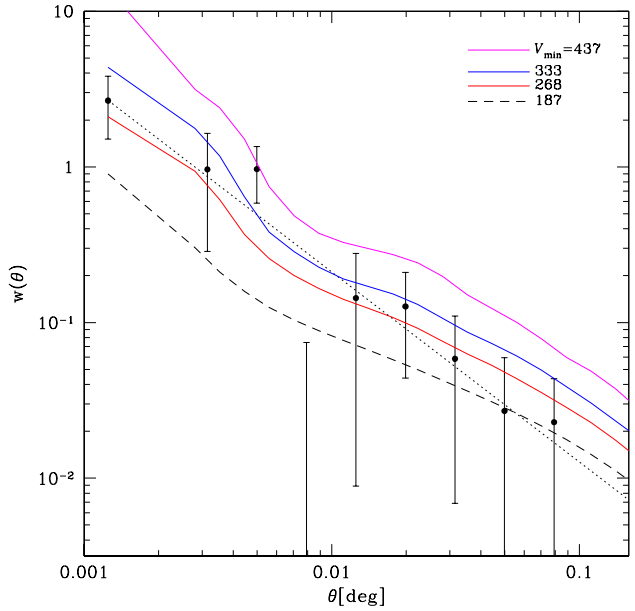



Figure 34. Angular correlation function models for dark matter halos with different threshold of the maximum circular velocity, computed using equation (5.4). $V_{\rm min}$ = 268 and 333 km/s corresponds to the best fit and a 90% upper limit, respectively. The model with $V_{\rm min}$ = 187 km/s has the correlation length $r_0 = 5.6\,h^{-1}$Mpc implied by simple power law Limber inversion. The $V_{\rm min}$ = 437 km/s threshold yields the number density of halos equal to that of IR-peakers, $n_{\rm halo} = n_{\rm pks}$.

## 24 Results

### *24.1 Circular velocity of dark matter halos associated with IR-peakers*

The angular correlation function models described above were fitted to the observed $w(\theta)$ for IR-peakers in the angular range $9'' < \theta < 360''$[5] The derived circular velocity threshold is $V_{\rm min} = 268 \pm 46\,{\rm km\ s^{-1}}$ (the $w(\theta)$ model for this best-fit value of $V_{\rm min}$ is represented by the red line in Fig. 34). The total mass corresponding to this velocity threshold is $2.1 \times 10^{12}\,M_\odot$. Note that for comparison with the *mean* stellar mass of IR-peakers (§ 21), a more appropriate quantity is the *mean* total mass of dark matter halos for the corresponding circular velocity threshold. Intergrating over the mass function of halos in the simulation output,

[5]The first data point in Fig. 34 was excluded because at $z \approx 2$, it corresponds to separations $< 100h^{-1}$ kpc; modeling of a pair of massive halos at such small separations is strongly sensitive to the numerical resolution of the simulations.

we find that DM halos with the minimum circular velocity $V_{\rm min} = 268 \pm 46\,{\rm km\,s^{-1}}$ have a mean total mass of $(4.8 \pm 2.0) \times 10^{12}\,M_\odot$. The number density of such halos is $n_{\rm halo} = 9.5 \times 10^{-4}\,h^3\,{\rm Mpc^{-3}}$, a factor of 11 higher than the observed number density of IR-peakers.

The 90% CL upper limit on $V_{\rm min}$ is 333 km s$^{-1}$ (the corresponding $w(\theta)$ model is shown by blue line in Fig. 34). We can consider this velocity threshold as upper boundary of $V_{\rm min}$ values consistent with the observed angular correlation function of IR-peakers. The set of parameters retrieved for DM halos with $v_{\rm max} > 333\,{\rm km\,s^{-1}}$ is a mean total mass of $8.2 \times 10^{12}\,M_\odot$, and a halo number density of $n_{\rm halo} = 3.5 \times 10^{-4}\,h^3\,{\rm Mpc^{-3}}$, a factor of 4 higher than that of IR-peakers.

A tension between the number densities of IR-peakers and DM halos with matching clustering properties can not be resolved by increasing the $V_{\rm min}$ threshold still further. This is illustrated by the magenta line in Fig. 34, which represents the $w(\theta)$ model computed for $V_{\rm min} = 437\,{\rm km\,s^{-1}}$, the threshold value for which the number density of DM halos is $n_{\rm halo} = n_{\rm pks} = 8.7 \times 10^{-5}\,h^3\,{\rm Mpc^{-3}}$. Clearly, this model predicts the angular correlation function significantly in excess of that observed.

*Notes on the correlation length of IR-peakers*

It is instructive to compare the correlation lengths derived from the simulation based $w(\theta)$ models with that given by the Limber inversion of the power law fit. We can define $r_0$ as a point at which the three-dimensional correlation function for DM halos with the given velocity threshold equals unity. The best fit and uncertainty interval of $V_{\rm min}$, the parameter defining the simulation based $w(\theta)$ models, corresponds to the central value and uncertainties for $r_0$. For $V_{\rm min} = 268 \pm 46\,{\rm km\,s^{-1}}$ (see above), we find $r_0 = 7.7 \pm 1.1\,h^{-1}\,{\rm Mpc}$. Recall that the power law fit to the same data yields $r_0 = 5.63 \pm 0.65\,{\rm Mpc}$ (§ 22.1). Therefore, the power law fit indeed yields biased results from fitting the correlation function in a narrow range of small scales, as suggested by [127].

### *24.2 Star formation efficiency and incidence rate for IR-peakers*

The clustering analysis of IR-peakers presented above allows us to estimate the total mass scale for their parent dark matter halos. As we just discussed,

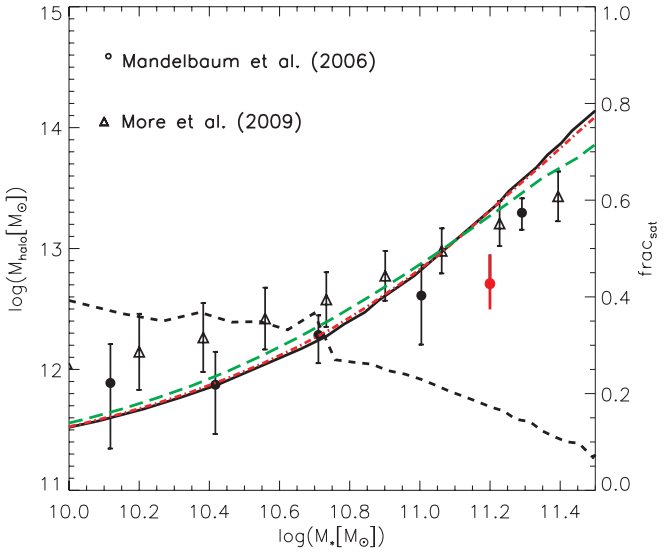



Figure 35. Dark matter halo mass as a function of stellar mass for low-redshift objects (reproduced from [96]). Black circles with error bars show weak lensing mass estimates of the mean halo mass of central galaxies as a function of their stellar mass [155]. Triangles represent the measurements based on the stacked kinematics of satellite galaxies [168]. The thick black, and red and green dashed lines show the prediction of $M_{\rm tot}$ from abundance matching assuming dispersions of 0, 0.1, and 0.2, respectively, in $\log M^*$ at the given halo mass. The red circle with error bars shows our results for IR-peakers at $z \approx 2$; the error bar shows the mean mass interval corresponding to the uncertainty interval on the minimum rotational velocity, $V_{\rm min} = 268 \pm 46$ km s$^{-1}$ (§ 24).

such an analysis based on the power-law fit to the observed angular correlation function and its Limber inversion leads to the biased low estimate of the correlation length, and hence a low total mass which is in tension with the observed stellar mass and comoving number density of IR-peakers (§ 22.1). A more complete analysis based on $w(\theta)$ models derived from the *Bolshoi* simulation outputs singficantly eases these tensions. It is, therefore, interesting to discuss further implications of these results for the star formation efficiency in IR-peaker galaxies.

The star formation efficiency, $f_*$, can be defined as the ratio of the stellar mass within a halo and the total baryon mass expected within this halo,

$$f_* = \frac{M_*}{M_{\rm tot}} \frac{\Omega_M}{\Omega_b}, \tag{5.6}$$

where $\Omega_b/\Omega_M = 0.17$ is the mean cosmic baryon fraction precisely measured by WMAP [124].

Using the mean total masses corresponding to the uncertainty interval of

$V_{\rm min}$, $268 \pm 46$ km s$^{-1}$, and a mean stellar mass of IR-peakers of $1.6 \times 10^{11}\,M_\odot$, we compute the star formation efficiency coefficient $f_* = 0.19^{+0.09}_{-0.06}$. For the 90% CL upper limit of $V_{\rm min}$, we find $f_* = 0.11$.

To put these results in context, we compare them with the star formation coefficient determinations for low redshift galaxies. Figure 35 shows a compilation of measurements and theoretical models from [96]. The red data point in this figure shows our results for IR-peakers at $z \approx 2$. Clearly, the derived $f_*$ for IR-peakers is comparable to, but slightly lower than, that for low-$z$ galaxies of a comparable stellar mass. The statistical significance of this difference is not great. Furthermore, the $f_*$ measurements at such different redshifts are not necessarily directly comparable.

However, the number density of IR-peakers is also lower than that of the dark matter halos with the same clustering properties (§ 24.1). Therefore it is interesting to consider a possibility that the IR-peakers represent a subpopulation (10–25%) of massive objects in which the star formation efficiency has been somewhat above a typical value in this mass range. In this regard, we note that we observe the IR-peakers near the epoch of the active star formation in the Universe ($z \approx 2 - 2.5$, [147]), we can expect that the stellar mass within the halo is sensitive to its exact formation history. As a result, our sample selection prefers DM halos with a higher than average stellar mass (e.g., those which were mostly assembled by $z = 3-4$), and misses DM halos with a lower stellar content (e.g., those which just formed from small units near $z \approx 2$). This situations is exacerbated further by the fact that given their high mass, the parent DM halos of IR-peakers must still be actively merging at $z = 2$. Thus we indeed can expect the existence of DM halos with very different assembly history. Only a subset of halos with sufficiently massive progenitors already in place by $z = 3 - 4$ is likely to host IR-peakers with extremely high stellar mass.

Mathematically, the effect described above is equivalent to existence of a substantial scatter in the stellar mass for the fixed $M_{\rm tot}$. In this case, when we select objects by stellar mass from a population with a steep mass function, a large fraction of objects are in fact lower-mass halos with a higher than average $M_*$. The clustering properties of the selected population will correspond to this low mass scale. Thus if we infer $M_{\rm tot}$ from clustering, we will overestimate $f_*$

somewhat, and also find a mismatch between the inferred $M_{\rm tot}$ and the comoving number density of objects. This effect is the stronger the steeper is the mass function, and so it should be particularly relevant for IR-peakers (cf. Fig. 30). A similar effect was considered by Guo et al. [96]. They find that at the highest mass end of the galaxy stellar mass function, a scatter of 0.2 dex in $M_*$ for a fixed $M_{\rm tot}$ changes the $M_*$ inferred from abundance matching by a factor of $1.5-2$ (see solid black and dashed green lines in Fig. 35), approximately the offset between $f_*$ we infer from the clustering analysis of IR-peakers and those from dynamical measurements of low-$z$ galaxies.

### *24.3 Constraints on location of IR-peakers within the dark-matter halos*

Finally, as the best-fit in Fig. 34 suggests, the 1-halo term in the correlation function at the smallest separations is marginally detectable in our $w(\theta)$ measurements, and we can use it to put constraints on the locations of IR-peakers within their parent dark matter halos. This would be interesting for a number of reasons. First, we argued above that the IR-peakers reside only in a subpopulation of massive DM halos at $z \approx 2$, and so any constraints on their location would help to better understand their physical origin. Second, a recent study by Tinker et al. [240] led to an interesting conclusion on the location for another class of high-redshift galaxies, Distant Red Galaxies (DRGs) — the objects with a factor of $\sim 2$ lower stellar mass and a factor of $\sim 8$ higher number density than those for IR-peakers [249]. Tinker et al. used the clustering properties of DRGs at $2 < z < 3$ reported by Quadri et al. [201] to apply a halo occupation model (§ 9) in which the incidence rate of DRGs in the main DM halos and subhalos (red satellite fraction in Tinker et al.) varies with the halo mass. The results of Tinker et al. imply that nearly all massive DM *subhalos* at $z = 2-3$ host a DRG while for central galaxies, the DRG fraction is a function of the halo mass.

The detailed modeling of the DRG correlation function was possible because the observed angular correlation function of DRGs has sufficiently small observational errors to reveal an obvious two-component form (Fig. 1 in [201]). The quality of our data does not allow us to perform a similarly detailed modeling for the IR-peakers. However, we can still try two limiting cases. The first is in which all IR-peakers are the central galaxies of parent DM halos, and the

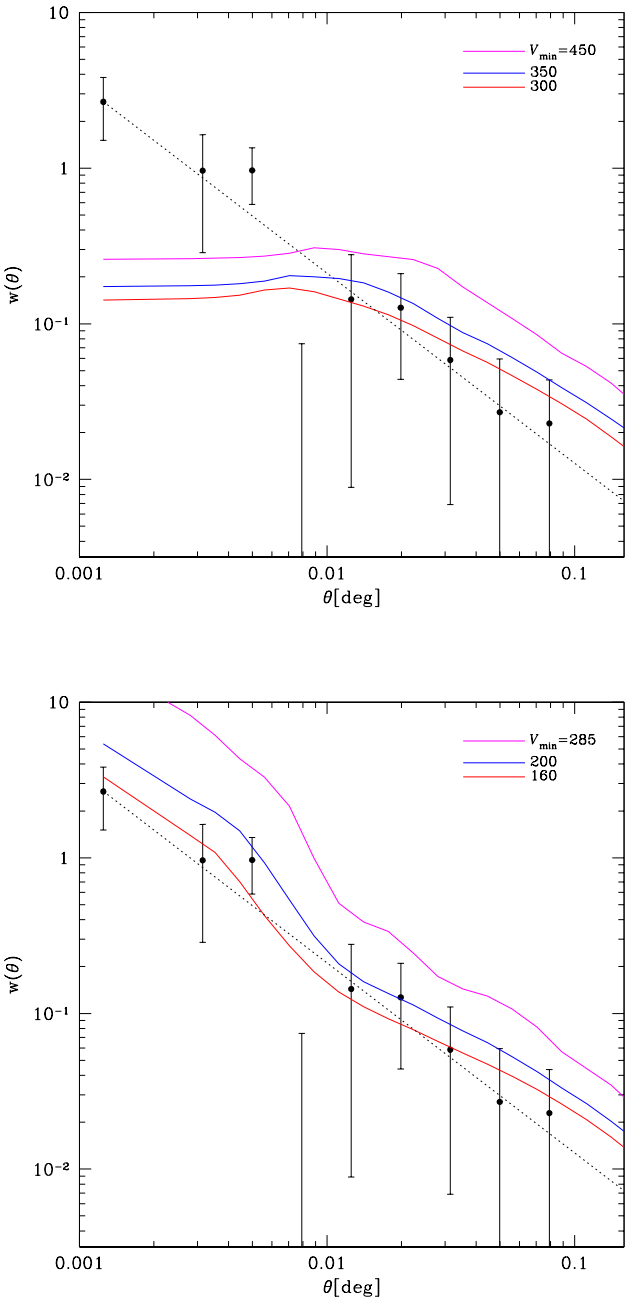



Figure 36. Angular correlation function models for for the cases when IR-peakers are forced to reside at the centers of distinct DM halos *(a)* and subhalos within more massive halos *(b)*. The models showed by magenta lines correspond to the $V_{\rm min}$ values for which the number density of DM (sub)halos equals that of IR-peakers.

second is a case in which they are forced to be located in the subhalos. Both possibilities can be straightforwardly realized using the outputs from the *Bolshoi* simulation, in which all DM halos are already classified into the distinct objects and subhalos within a bigger halo. After we populate the appropriate set of (sub)halos with IR-peakers, the $w(\theta)$ model can be computed identically to the procedure described in § 23 above.

If the IR-peakers are located only in the *central* galaxies, there should be no object pairs with $3D$ separations less than $\sim$ the virial radius of the corresponding DM halos. This leads to the $w(\theta)$ models being flat at small angular separations, $\theta \lesssim 0.01^\circ$ where we in fact detect a strong clustering signal (Fig. 36a). For any limiting $V_{\rm min}$, there are residuals between the model and the data at a $\sim 2.5 - 3\,\sigma$ level. Therefore, the scenario in which the IR-peakers are associated only with the central galaxies is marginally excluded by the observed small-scale clustering.

In the model, in which the IR-peakers reside only in the *satellite* galaxies, we can obtain a good fit to the observed angular correlation function (Fig. 36b). However, the parameters are unphysical. The best fit $V_{\rm min}$ for the subhalos is 160 km s$^{-1}$, and the 90% CL upper limit is 200 km s$^{-1}$ (red and blue lines in Fig. 36b). This corresponds to a mean total mass of 1.2 and $2.0 \times 10^{12}\,M_\odot$, respectively, — too low for the observed stellar mass of IR-peakers (e.g., it implies a star formation efficiency of $\sim$ 80% for the best fit and 47% for the 90% CL upper limit of $V_{\rm min}$). Furthermore, for such low $V_{\rm min}$, the mismatch between the expected number density of DM subhalos and that of the IR-peakers is a factor of $\sim$ 10.

## 25 Conclusions

We detect angular clustering of IR-peaker galaxies — the highest stellar mass galaxies at $z \approx 2$ — in the SWIRE ELAIS-S1 field with a total statistical significance of $\gtrsim 4\,\sigma$. Because of the insufficient statistics and a relatively small field size, the angular correlation function measurements are confined to small scales ($\sim 0.01^\circ$). An attempt to fit these data with a simple power law and then derive the total mass scale for IR-peakers from the spatial correlation length obtained

from the Limber inversion leads to unphysical results. The estimated total mass is too low for the observed stellar mass of IR-peakers and the number density of DM halos with such a low mass is a factor of 50 higher than that of IR-peakers.

We argue that this unphysical results arise because the spatial correlation function of IR-peakers at such small scales strongly deviates from the power law because of the presence of the "1-halo" term. A more complete modeling, where the spatial correlation function model is obtained using the catalog of DM halos in a large-box cosmological simulation leads to much more physical results (§ 23). Based on this analysis, we argue that the IR-peakers reside in a subpopulation (10–20%) of massive DM halos (with maximum circular velocities $v_{\rm max} \gtrsim 300\,{\rm km\,s^{-1}}$), and a slightly enhanced efficiency of star formation ($f_* \simeq 10 - 20\%$).

# VI Clustering of *Chandra* X-ray AGNs in the Boötes field

## 26 Summary

We present the measurement of the spatial clustering of X-ray selected AGNs using a sample derived from the *Chandra* survey in the Bootes field. The real-space two-point correlation function over a redshift interval from $z = 0.17$ to $z \sim 3$ is well described by the power law, $\xi(r) = (r/r_0)^{-\gamma}$, for comoving separations $r \lesssim 20\,h^{-1}\,\mathrm{Mpc}$. We find $\gamma = 1.97 \pm 0.09$ and $r_0$ consistent with no redshift trend within the sample (varying between $r_0 = 5.4 \pm 0.5\ h^{-1}\,\mathrm{Mpc}$ for $\langle z \rangle = 0.37$ and $r_0 = 7.0 \pm 0.8\ h^{-1}\,\mathrm{Mpc}$ for $\langle z \rangle = 1.28$. Further, we are able to measure the projections of the two-point correlation function both on the sky plane and in the line of sight. We use these measurements to show that the *Chandra*/Boötes AGNs are predominantly located at the centers of dark matter halos with the circular velocity $v_{\mathrm{max}} > 310\,\mathrm{km\ s^{-1}}$ or $M_{180} > 3.7 \times 10^{12}\,h^{-1}\,M_{\odot}$, and tend to *avoid* satellite galaxies in halos of this or higher mass. The halo occupation properties inferred from the clustering properties of *Chandra*/Boötes AGNs — the mass scale of the parent dark matter halos, the lack of significant redshift evolution of the clustering length, and the low satellite fraction — are broadly consistent with the [105] scenario of quasar activity triggered by mergers of similarly-sized galaxies.

## 27 Introduction

Direct observations of host galaxies of the high-redshift active galactic nuclei are hard or impossible with the current instrumentation, except for highly obscured

or low-luminosity objects. Studies of the AGN clustering properties are thus a unique source of information on the AGN hosts and their environment. At low redshifts, the supermassive black holes exist at the centers of most low-redshift galaxies, and there is a tight correlation between the SMBH mass and the properties of the bulges of host galaxies [153, 75, 87]. This suggests that most galaxies hosted an AGN at some point in their evolution, and that AGNs at each redshift are stochastic "markers" of a population of galaxies in which the conditions are favorable for accretion of matter onto the central SMBH. Through matching the clustering properties of AGNs to those of dark matter halos, or with those of different types of galaxies, or with those of AGNs of different types, we can determine the typical mass scale of the AGN hosts, their morphological type, and determine whether different types of AGNs are hosted in the same type of objects.

Clustering of optical quasars indeed is very similar to that of galaxies. The two-point correlation function in the separation range $1 - 20\,h^{-1}$ Mpc is well described by a power law, $\xi(r) = (r/r_0)^{-\gamma}$ with a slope of $\gamma = 1.9$ and a correlation length of $r_0 = 5.5\,h^{-1}$ Mpc [210, 222]. Further, using SDSS and 2QZ quasar sample, [53] and [210] track the evolution of the optical quasar clustering amplitude over the redshift range $z = 0.5 - 2.5$ and find only a mild evolution. The most recent direct measurements of the X-ray AGN clustering were presented in [91, 92]. [102] studied clustering properties of different types of AGNs at $z \approx 0.5$ through cross-correlation with the galaxy catalog. Generally, the two-point correlation function of the X-ray AGNs was found to be similar to that of the optical quasars; however, the previous X-ray studies could not constrain the evolution of the correlation function over a sufficiently wide redshift range.

The main conclusion from previous clustering analyses is that AGNs are located in galaxy group-sized dark matter halos ($M \sim 2 \times 10^{12}\,h^{-1}\,M_\odot$, see [210]), with the mass scale fairly independent of the object redshift or observed luminosity. [102] also find some differences in the clustering properties and color of host galaxies for the X-ray, radio, and infrared-selected AGNs at $z = 0.5$. The difference in the AGN clustering properties and colors of their host galaxies lead [102] to conclude that these different techniques select distinct source populations and not simply different stages of rapidly changing AGN properties.

In addition to determining the mass scale of the the AGN host dark matter halos, it is interesting to establish *where* within the halos the active galaxies are located. This question can be addressed by direct observations of individual objects only in rich galaxy clusters. Indeed, some studies [40, 166, 159] indicate an excess of X-ray AGNs in the cluster outskirts (see, however, [125]). However, the majority of quasars and AGNs are located within galaxy group-sized objects, where it is hard or impossible to independently localize the centroid of the system, especially at high redshifts. For such systems, it is possible to determine the fraction of objects in the satellite galaxies through analysis of the two-point correlation function at small separations. An example of such an approach can be found in [180]. Based on cross-correlation of the SDSS quasar and Large Red Galaxies samples, these authors argue that a high fraction, > 25%, of the optical quasars must be located in non-central galaxies.

One of the main goals of presented below work is to constrain the location of X-ray emitting AGNs within their host dark matter halos by a more direct method. The effect we are using is a strong dependence of a galaxy's peculiar velocity on its location within the host halo. The central galaxies are predicted (and observed, at least in massive clusters, see [176]) to be nearly at rest with respect to the host halo, and their random motions correspond to motions of the halos as a whole. The satellite galaxies move at approximately the virial velocities *within* the halo. As a result, the satellite galaxies have much faster peculiar motions and form "finger of God" structures in the radial velocity space. The effect of high peculiar velocities of the satellite galaxies can be detected though comparison of the objects' clustering properties as a function of projected separation, $r_p$, and the line-of-sight separation, $\pi$. Peculiar motions do not affect $r_p$ but can strongly distort $\pi$ because it is derived from the measured redshift.

Using peculiar motions in the previous studies of the AGN clustering was hard because of the small number of objects and insufficient accuracy of the redshift measurements[1]. Fortunately, we now can use an excellent sample for

[1]For example, the redshift uncertainties corresponded to peculiar velocities of $\approx$ 420 – 500 km s$^{-1}$ in the 2QZ sample [53]. Because such a high level of uncertainties, [53] could not extract useful information from the velocity-space distortions although they were included

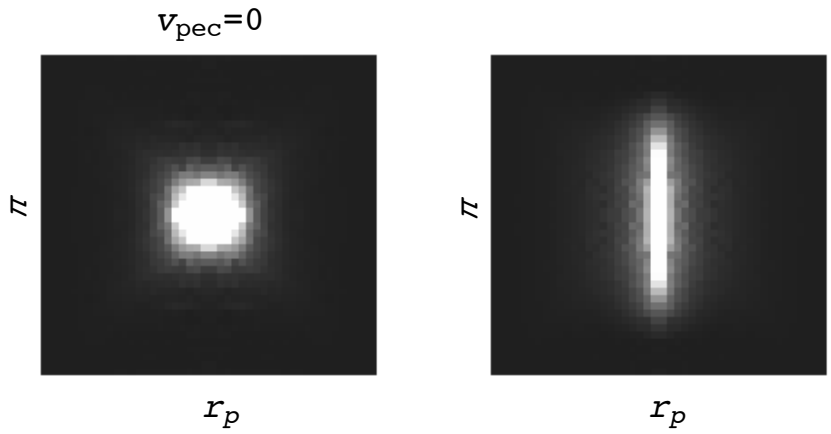



Figure 37. The two-dimensional correlation function $\xi(r_p, \pi)$ for objects located within dark matter halos with the maximum circular velocity $V_{max} > 250$ km s$^{-1}$ at $z = 1$, plotted as a function of transverse ($r_p$) and line-of-sight ($\pi$) pair separation. 70% of objects are forced to be at the centers of such halos, while 30% are put in one of the satellite subhalos. In the left panel, the peculiar velocities were set to 0.

such studies. The *Chandra* survey of 9.3 deg$^2$ in the Boötes region [171] provides a uniformly selected sample of > 3000 X-ray selected AGNs, for a uniform subsample of which (∼ 1900 sources) high-quality spectra were measured with MMT/Hectospec (Kochanek et al., in preparation).

Full information on the object clustering and peculiar velocities is contained in the two-dimensional correlation function, $\xi(r_p, \pi)$. An example derived from the numerical simulations we use here is shown in Fig. 37. See, e.g., [261] for an example of the detailed modeling of the $\xi(r_p, \pi)$ function measured for the SDSS galaxies. Unfortunately, such a detailed modeling is impossible for the present high-$z$ AGN samples due to limited statistics. However, we show that useful information can still be obtained using projections of $\xi(r_p, \pi)$ on the radial velocity direction and on the sky plane, $\xi(\pi)$ and $\xi(r_p)$, respectively (formally defined in § 3). Using numerical cosmological simulations, we show that if a substantial fraction of objects is in satellite galaxies, $\xi(\pi)$ is expected to be significantly in excess of $\xi(r_p)$ in the range of comoving distances $1 - 10\,h^{-1}$ Mpc (§ 30.3). The observed correlation functions do not show such an excess (§ 31.1), which we exploit to put an upper limit on the fraction of X-ray AGNs in the satellite galaxies. Finally, we explore the redshift evolution of the clustering

in the modeling of the spatial correlation function.

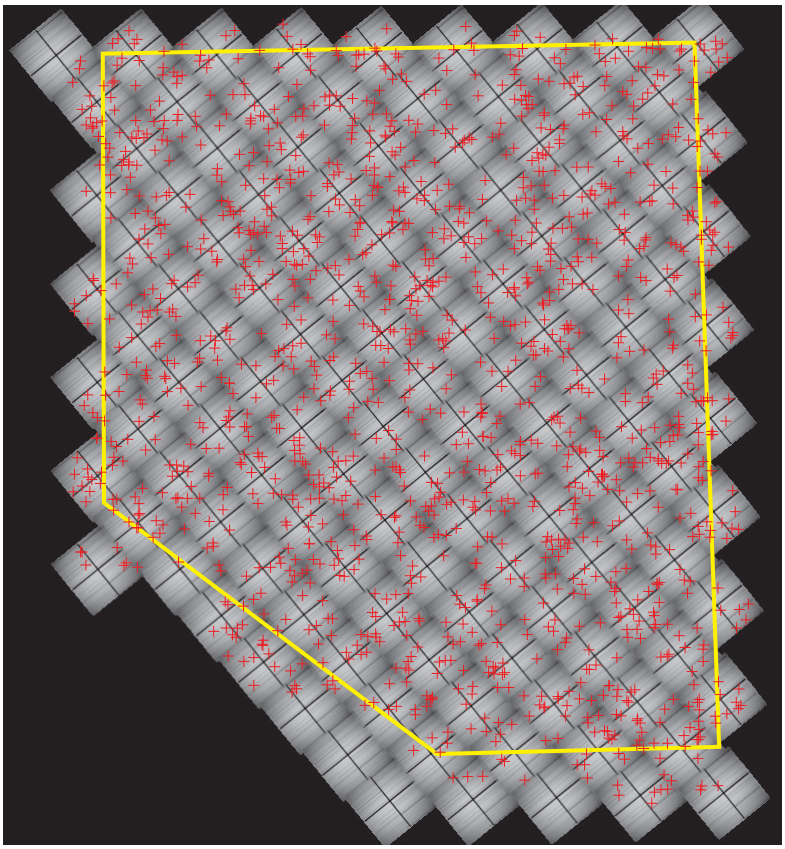


Figure 38. *Chandra* survey in the Boötes field. Grey scale map shows the combined *Chandra* sensitivity map (exposure maps for 126 individual pointings, each multiplied by the point source detection sensitivity as a function of off-axis distance derived in § 28.1). Crosses mark the location of X-ray sources used in this work. The polygon shows conservative boundaries of the spectroscopic survey. We consider only 1282 sources within the polygon boundaries in the clustering analysis.

length and compute the typical mass scale of the AGN host dark matter halos and put constraints on the AGN duty cycle.

## 28 AGN sample

We use a sample of high-$z$ AGNs derived from the *Chandra* X-ray survey in the 9.3 deg$^2$ Boötes field of the NOAO Deep Wide-Field Survey [171]. The region was uniformly covered with a grid of overlapping 5 ksec ACIS-I pointings providing a sensitivity of $4.7 \times 10^{-15}$ erg s$^{-1}$ cm$^{-2}$ in the 0.5–2 keV energy band [115]. Extensive optical data exist for this field and 98% of X-ray sources have optical or infrared counterparts [32]. The redshifts for X-ray sources with optical counterparts brighter than $I = 21.5$ were uniformly obtained with the MMT/Hectospec

instrument in the AGN and Galaxy Evolution Survey (AGES, C. S. Kochanek et al., in preparation). Further details on the X-ray and optical observations can be found in [102].

For clustering analysis, we use 1282 X-ray selected AGNs with spectroscopic redshifts. Redshift measurements effectively introduce an additional selection criterion of the sample, based on the optical magnitude of the counterparts, $I < 21.5$. This selection removes $\sim 35\%$ of the X-ray sources and likely introduces a high-$z$ cutoff in the redshift distribution of our sources (§ 28.2). Fortunately, the redshift measurements were done with Hectospec [70] in several passes, which excludes the problem with the so-called "fiber collisions". The locations of X-ray sources with spectroscopic redshifts are marked in Fig. 38.

To reconstruct the correlation function of sources, we need to simulate catalogs of randomly distributed sources whose distribution follows the sensitivity variations and the redshift distribution of our catalog. Below, we describe how these functions were derived from our *Chandra* sample.

### *28.1 Spatial variations of sensitivity*

*Chandra* sensitivity for the detection of point X-ray sources is not uniform across the field of view. Sensitivity variations ($\sim \pm 25\%$, see below) are imprinted on the distribution of detected sources. The typical spatial scale of the sensitivity variations is several arcmin, which corresponds to the comoving distance of $1 - 2\,h^{-1}$ Mpc at $z = 1$. This is comparable to the distance scale where we measure clustering, therefore these variations must be taken into account.

The sensitivity variations are caused mainly by two effects — 1) the vignetting of the *Chandra* X-ray telescopes, and 2) the degradation of the Point Spread Function (PSF) away from the optical axis. The mirror vignetting is well-calibrated and its effect can be computed for a given source population. However, the effects of PSF degradation on the source detection efficiency in the low-photon regime are very complex. Therefore, it is best to measure the combined effect of the sensitivity variations empirically.

Both the mirror vignetting and PSF degradation are approximately azimuthally symmetric. Therefore, we can assume that the detection sensitivity is a function of the source off-axis angle, $r_c$ and measure it from the radial profile of the

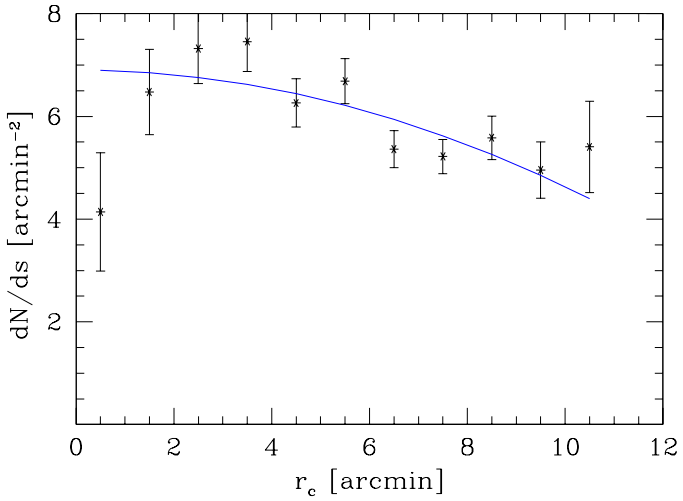



Figure 39. The observed surface density of detected X-ray sources as a function of the distance from the telescope optical axis. The sharp decrease at $r_c < 1'$ is attributed to the gap between the *Chandra* ACIS-I CCDs. The gradual decline at larger radii reflects the sensitivity variation caused by mirror vignetting and PSF degradation. The solid line is the best-fit second-order polynomial (eq. [6.1]).

surface density distribution of detected X-ray sources, averaged over all the 126 fields. The results is shown in Fig. 39. For a uniform sensitivity, we would expect a constant surface density, while in reality we find that the average surface density is $\sim 7$ arcmin$^{-2}$ near the optical axis, and falls to $\sim 5$ arcmin$^{-2}$ at off-axis distances of 10 arcmin. The derived radial profile of the source surface density can be modeled with a second-order polynomial,

$$\frac{dN}{ds} = 6.9\left(1 - (r_c/17.4')^2\right). \tag{6.1}$$

In addition to this gradual variation with radius, there are sharp features in the spatial sensitivity pattern related to the gaps between the *ACIS-I* CCDs. In particular, these gaps are responsible for the drop in the number of detected sources at $r_c < 1'$ in Fig. 39. These sensitivity variations can be adequately taken into account using the standard *Chandra* exposure maps.

Our final sensitivity map (Fig. 38) consists of the merged set of *Chandra* exposure maps computed for the individual 126 pointings, each multiplied by the radial sensitivity pattern given by eq. (6.1). This sensitivity map is taken into account in the derivation of the correlation function through the generation of the appropriate catalogs of random sources. The total area within the conservative boundaries of the spectroscopically surveyed region of the Boötes field

is 7.30 deg$^2$. The total effective area, taking into account the gaps between the *Chandra* CCDs and degradation of the detection efficiency, is 5.90 deg$^2$.

### *28.2 Redshift distribution*

The model for the source redshift distribution, $dN/dz$, should reflect both the intrinsic variations of the comoving number density with redshift and all selection effects of the catalog. A commonly used approach is to model the observed $dN/dz$ distribution with a high-order polynomial (see, e.g., [53, 210], for recent examples). This approach works well for catalogs with a large number of sources. However, for smaller catalogs, like ours, there is a danger that a high-order polynomial fit will follow statistical fluctuations in the observed $dN/dz$, while a low-order polynomial would be unable to adequately model the strong gradients at low $z$. Therefore, we fit the redshift distribution of AGNs in the Boötes field with a parametric model based on several physical assumptions.

The first component of the model represents the cosmological comoving volume per unit redshift, $dN_1/dz \propto dV/dz \propto D_c^2$, where $D_c(z)$ is the comoving distance to redshift $z$. The second component is a power law function of the minimum luminosity which corresponds to the *Chandra* flux limit at redshift $z$, $dN_2/dz \propto L_{\rm min}^{\alpha/2} \propto d_L^{\alpha}$, where $d_L$ is the luminosity distance, $d_L = D_c(1+z)$. This component represents the effect of the low-$L_x$ cutoff of the intrinsic luminosity function introduced by the selection which is primarily based on *Chandra* detections. It also can describe the evolution of the luminosity function at high $z$. The third component is a high-$z$ cutoff modeled by a broad Gaussian, $dN_3/dz \propto \exp\left(-d_L^2/C^2\right)$. This component can represent the high-$L_x$ cutoff or steepening of the intrinsic AGN luminosity function, and also can describe various observational limits implicitly built into our catalog (e.g., a lower efficiency of optical identifications and redshift measurements for the highest-$z$ X-ray sources). This simple analytic model,

$$\frac{dN}{dz} = \text{const} \times \frac{dV}{dz}\, d_L^{\alpha}\, \exp\left(-\left(d_L/C\right)^2\right), \tag{6.2}$$

which has only two free parameters provides a strikingly good fit to the observed redshift distribution of the Bootes X-ray selected AGNs (Fig. 40). The best-fit values are $\alpha = -1.07$ and $C = 1.50 \times 10^3\, h^{-1}$ Mpc. We also tried to introduce

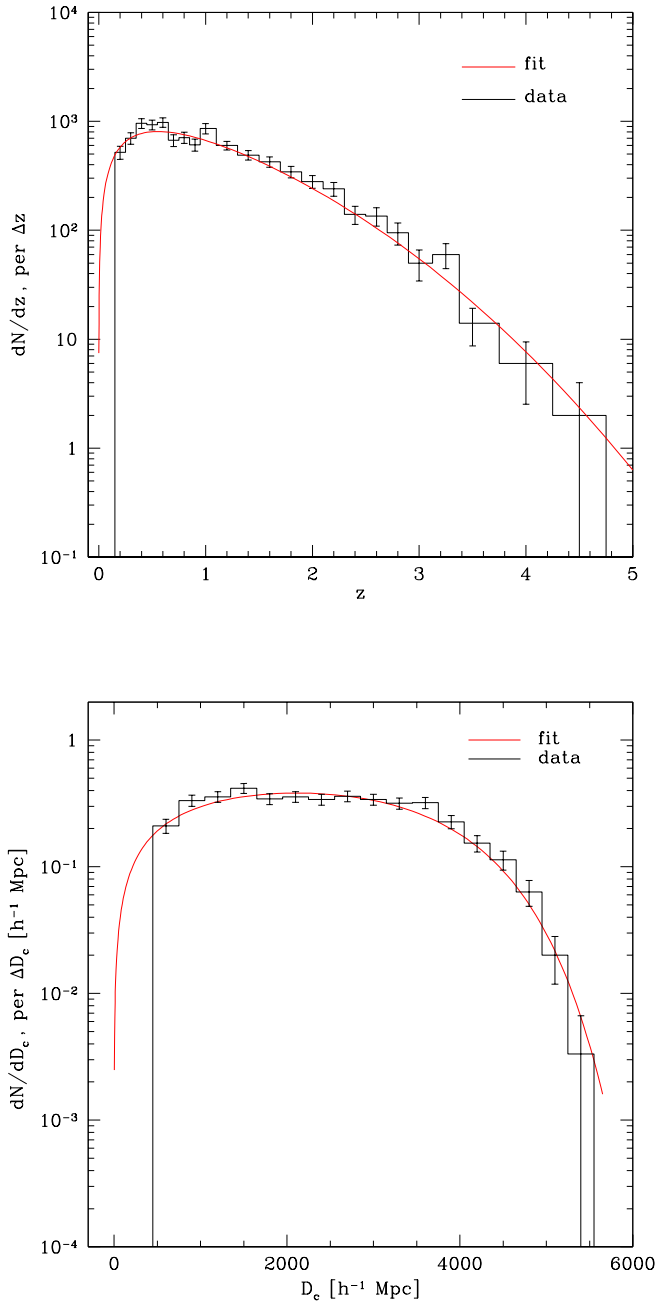



Figure 40. The observed distribution of *Chandra*/Boötes AGNs as a function of redshift *(top)* and comoving distance *(bottom)*. The lines show the analytic fit by eq. [6.2].

an intrinsic number density evolution in the first component, $dN_1/dz \propto (1+z)^{\beta}\, dV/dz$, but the best-fit values of $\beta$ were consistent with zero (no evolution). Note that this does not imply that the AGN luminosity function does not evolve (however we note that the observed evolution is indeed weak for low $L_x$, see e.g., Fig. 15 in [245]) because such an evolution can be also described by the $d_L^{\alpha}$ term since we do not tie $\alpha$ to the slope of the AGN luminosity function.

Figure 40 also demonstrates the general characteristics of our sample. The peak in the observed $dN/dz$ distribution is near $z \approx 0.6$. The median redshift of the sample is $z_{\rm med} = 1.04$. The tail in the redshift distribution extends to $z \approx 4.5$ but the fraction of AGNs with $z > 3$ is very small. Overall, the clustering properties of sources in our sample are most sensitive to the distribution of the X-ray AGN population near $z \approx 1$.

## 29 Two-point correlation function of *Chandra*/Boötes AGNs

The two-point correlation function for our complete sample is calculated as following. First, we estimate the two-dimensional correlation function, $\xi(r_{\rm p}, \pi)$ (see Chapter II, § 3 for details), on a grid of separations with equal logarithmic width for each cell, $\Delta \log d = 0.2$. We use the Landy & Szalay estimator (eq. 2.2). The random catalog is generated using the spatial sensitivity map and the model for the redshift distribution in our sample (§ 28.1 and 28.2). To minimize the additional noise, the number of objects in the random catalog is a factor of 100 larger than that for the AGN sample. The derived $\xi(r_{\rm p}, \pi)$ is then integrated in the radial and sky plane directions to obtain $w(r_{\rm p})$ and $w(\pi)$ (eq. 2.12 and 2.17). The integration is truncated at separation scales of $\pm 40\, h^{-1}$ Mpc to minimize noise (e.g. [92]). We then fit a power law (eq. 2.15) to $w(r_{\rm p})$ to measure the slope of the correlation function, $\gamma = 1.97 \pm 0.09$. The correlation length measured for the full sample is $r_0 = 6.27 \pm 0.33\, h^{-1}$ Mpc. We then compute the renormalization factor $A(\gamma)$ (eq. 2.16) and correct for the truncation of the integration in $w(r_{\rm p})$ and $w(\pi)$ at $40\, h^{-1}$ Mpc (eq. 2.18). With the renormalization factors determined, we convert $w(r_{\rm p})$ and $w(\pi)$ into $\xi(r_{\rm p})$ and $\xi(\pi)$ using eq. [2.19–2.20]. These functions are presented in Fig. 41a (filled and open circles, respectively). The shown statistical uncertainties include the Poissonian shot noise and intrinsic variance

terms (eq. 2.4). We prefer to use this analytic approximation instead of jackknife sampling because our sample has a small number of objects and the jackknife estimates are too uncertain.

As discussed in Chapter II and briefly mentioned above, $\xi(r_p)$ should be close to the real 3D correlation function $\xi(r)$, while $\xi(\pi)$ should be distorted by peculiar motions of the AGN host galaxies. Indeed, we observe a strong suppression in $\xi(\pi)$ relative to $\xi(r_p)$ at separations $\pi < 1\,h^{-1}$ Mpc (Fig. 41a). Unfortunately, the corresponding radial velocity difference, $\Delta z \approx 0.0003$, is uncomfortably close to the uncertainties in the AGES redshift measurements (Kochanek et al., in preparation). Therefore, we ignore the $\xi(\pi)$ data at $\pi < 1\,h^{-1}$ Mpc. At intermediate separations ($r \approx [1-10]\,h^{-1}$ Mpc), $\xi(\pi)$ is expected to be enhanced by the peculiar motions (see below). No such enhancement is present in our data. In fact, the ratio $\xi(\pi)/\xi(r_p)$ is fully consistent with 1 at separations $> 1\,h^{-1}$ Mpc within the measurement uncertainties (Fig. 41b). Below, we show that this can be used to put an upper limit on the fraction of AGNs that can reside in satellite galaxies orbiting within massive dark matter halos.

*Comparison with previous measurements*

Our results represent the most accurate measurement of the spatial clustering of X-ray selected AGNs to date, so a comparison with earlier observations is useful.

The first detection of angular clustering of X-ray sources was reported by [251] based on the analysis of the *ROSAT* PSPC data. Using the Limber equation reconstruction [184], these authors estimated a correlation length of $19\pm5\,h^{-1}$ Mpc (comoving) at $z = 1.5$ using raw measurements, and $7.5\pm2\,h^{-1}$ Mpc after correcting for the "amplification" bias caused by the poor angular resolution of the *ROSAT* PSPC. The latter value is in good agreement with our results.

Our results also are in good agreement with direct measurements of the spatial clustering in the ROSAT NEP survey [39], *Chandra* surveys in Deep Fields North and South [91], and the *XMM-Newton* COSMOS field [92]. Correlation lengths for the X-ray AGNs have been estimated also from the Limber inversions of the angular clustering measured for the *XMM-Newton* and *Chandra* sources. A wide range of $r_0$ values can be found in the literature (e.g., [189] and references

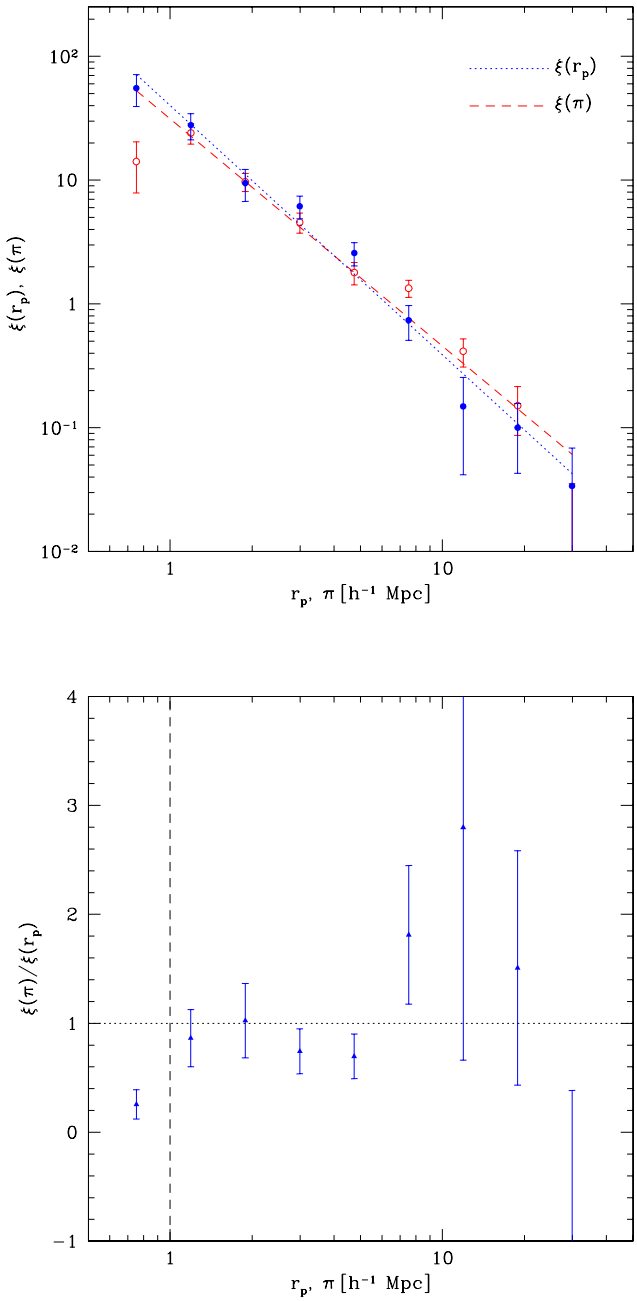



Figure 41. *(a)* The two-point correlation functions of Bootes/*Chandra* AGNs estimated from projections on the sky plane ($\xi(r_p)$, filled circles) and on the line of sight ($\xi(\pi)$, open circles). See §29 for the description of our procedure to derive $\xi(r_p)$ and $\xi(\pi)$. *(b)* The ratio of two projected functions. At small separation ($\pi \lesssim 1\,h^{-1}$ Mpc, indicated by the vertical dashed line), the ratio can be affected by uncertainties of the AGES redshift measurements. At larger separations, we observe no statistically significant difference between $\xi(r_p)$ and $\xi(\pi)$.

therein); our measurements are inconsistent with $r_0 > 10\,h^{-1}$ Mpc reported in some of these analyses.

## 30 AGN clustering model

In the past ten years, a very successful framework for modeling the nonlinear clustering properties of galaxies has been developed (the so-called Halo Occupation Distribution (HOD) model, see Chapter III, § 9.1 for details). The approach is based on the idea that the distribution of dark matter can be fully described through the mass function, linear bias, and density profiles of dark matter halos. These elements are well-calibrated using $N$-body simulations. The two additional ingredients of the model, which are less well known, are the probability distribution for a halo of mass $M$ to contain $N$ galaxies, and the distribution of galaxies within the halos. These functions can be parameterized by the functions suggested by the results of high-resolution numerical simulations (e.g., [127]), and some parameters of the model can be in fact determined by fitting the observed correlation functions. In particular, [127] and [265] show that the elements of HOD can be effectively decomposed into two components, separately describing the properties of central and satellite galaxies within the dark matter halos.

The HOD model is now very well developed for fitting the *shape* of the two-point correlation function. This technique has been applied to modeling the projected two-point correlation functions, $w(r_p)$, for objects ranging from Lyman-break galaxies at $z = 3 - 5$ [48] to relatively low-$z$ quasars from SDSS [180]. Recently, the HOD approach has been developed also for modeling the redshift-space distortions in the galaxy correlation functions [241, 239] — just the type of information we are aiming to use in this work to constrain the locations of Bootes/*Chandra* AGNs within the host dark matter halos.

In principle, the HOD models for galaxy clustering are analytic, and thus are convenient for those applications in which the cosmological parameters are varied. However, some of the most essential parameters of the HOD models are calibrated using numerical simulations. If one is interested in varying the parameters of galaxy distribution at a fixed redshift in a fixed cosmological

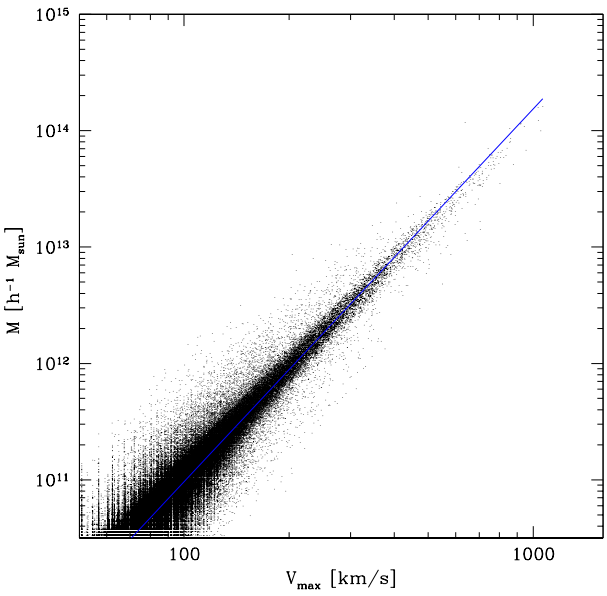



Figure 42. Mass and maximum circular velocity relation for dark matter halos from $\Lambda$CDM$_{120}$ - simulation (§7.1) output at redshift $z = 1$.

model, it is more accurate — and easier — to obtain the model correlation functions directly from numerical simulations rather than to rely on analytic approximations derived from analyzing the simulations. This is the approach we take here.

### *30.1 Numerical simulations*

The set of simulations we used for this work is described in chapter III, §7.1. These are high-resolution dissipationless simulations ($\Lambda$CDM$_{120}$) run in a flat $\Lambda$CDM cosmology within a box $120\,h^{-1}$ Mpc across. The locations, velocities, masses of the dark matter halos in the simulations were recorded at $z = 0.09$, 0.5, 1.0, 2.0, 3.1, 4.0, and 5.0, matching well the redshift distribution in the Bootes/*Chandra* AGN sample (see Fig. 40). Note that we choose to characterize the halos using their maximum circular velocities $v_{\rm max}$ rather than the virial mass for the reasons outlined in Chapter III, §7. However, $v_{\rm max}$ tightly correlates with the halo mass, as demonstrated in Figure 42.

The best-fit power law (blue solid line in Fig. 42 ) to the $v_{\rm max} - M_{180}$ relation[2] derived in these simulations at $z = 1$ is

$$\log M_{180} = 4.57 + 3.21 \log v_{\rm max}. \quad (6.3)$$

Here velocities are in km s$^{-1}$and masses are in units of $h^{-1}\,M_{\odot}$.

Because we are dealing with halos that exist in various environment, it is important to define once again (see for details § 7) that we use the term *host (distinct, parent)* halos to refer to objects whose centers are not located within any larger virialized systems, and *subhalos* which lie within the virial radius of a larger system. Briefly, a halo is classified as a subhalo if its center is within $r_{180}$ of the center of a more massive halo, where $r_{180}$ is the radius which corresponds to a mean spherical overdensity of 180 relative to the mean density at the given redshift. In the real Universe, centers of host halos can be identified as locations of the groups' central galaxies, while subhalos correspond to satellite galaxies.

### *30.2 Model of the AGN population*

Our data can constrain the following two basic properties of the AGN population. First, the overall clustering amplitude observed in $\xi(r_p)$ constrains the mass scale of the AGN host dark matter halos. Note that it is the mass scale of host halos, not of individual subhalos in which the AGNs might reside, which is constrained by the $\xi(r_p)$ amplitude. For example, a population of small subhalos with $v_{\rm max} = 50\,{\rm km\ s^{-1}}$ which are located within larger halos with $v_{\rm max} = 300\,h^{-1}\,{\rm km\ s^{-1}}$ has the clustering length which closely matches that for big host halos, and not that for the entire population of $v_{\rm max} = 50\,{\rm km\ s^{-1}}$ halos. Therefore, the first parameter of the model we use to characterize the spatial distribution of AGNs is the minimum $v_{\rm max}$ for the halos which can contain the X-ray AGNs. The AGN can be located either at the center of such a halo or in any of its smaller subhalos.

For a fixed threshold of the host halo $v_{\rm max}$, the correlation length of astronomical objects only weakly depends on whether these objects are located in

[2]Hereafter, we use the mass defined within an overdensity threshold of 180 with respect to the mean density of the Universe at the given redshift. The correspondence between $v_{\rm max}$ and mass is quoted for $z = 1$ unless the redshift is stated explicitly.

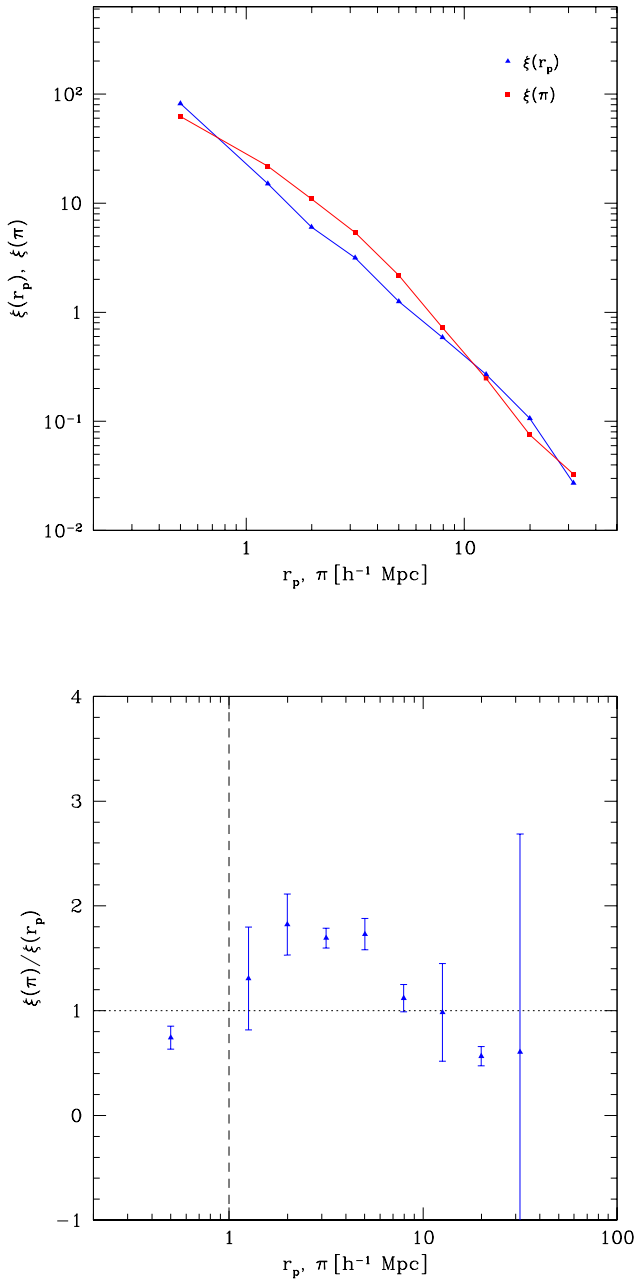



Figure 43. Correlation functions of all objects in the simulation output (both halos and subhalos) with $v_{\rm max} > 300\,{\rm km\ s^{-1}}$. The amplitude of the $\xi(r_p)$ function (left panel) matches the observed correlation length of the *Chandra*/Boötes sample, $r_0 = 6.27\,h^{-1}$ Mpc. However, many objects with such $v_{\rm max}$ are satellites in larger halos. This results in large peculiar motions leading to strong distortions of $\xi(\pi)$ with respect to $\xi(r_p)$. The ratio of $\xi(\pi)/\xi(r_p)$ averaged over three viewing directions of the simulation box at $z = 1$ is shown in the right panel (the uncertainties represent the scatter between the viewing directions). The predicted ration in the $2-6\,h^{-1}$ Mpc separation range is $\approx 1.8$, inconsistent with the observed correlation functions (see Fig. 41).

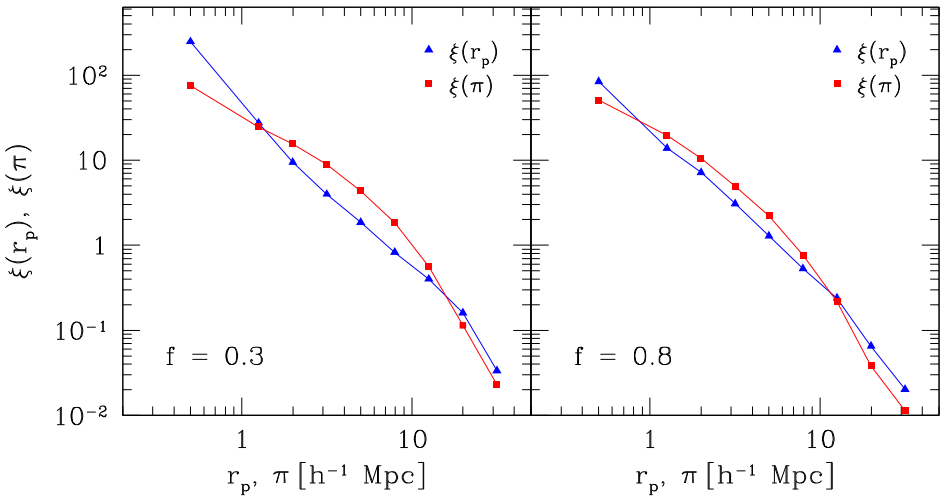



Figure 44. The projected correlation functions for halo samples at $z = 1$ with $v_{\rm max} > 250$ km s$^{-1}$, calculated with $f = 0.3$ and $f = 0.8$. The correlation functions for smaller $f$ (more AGNs in the satellite galaxies) show a stronger "1-halo" term at small separations [$\xi(r_p)$] and a stronger presence of the "finger of God" effect [$\xi(\pi)$].

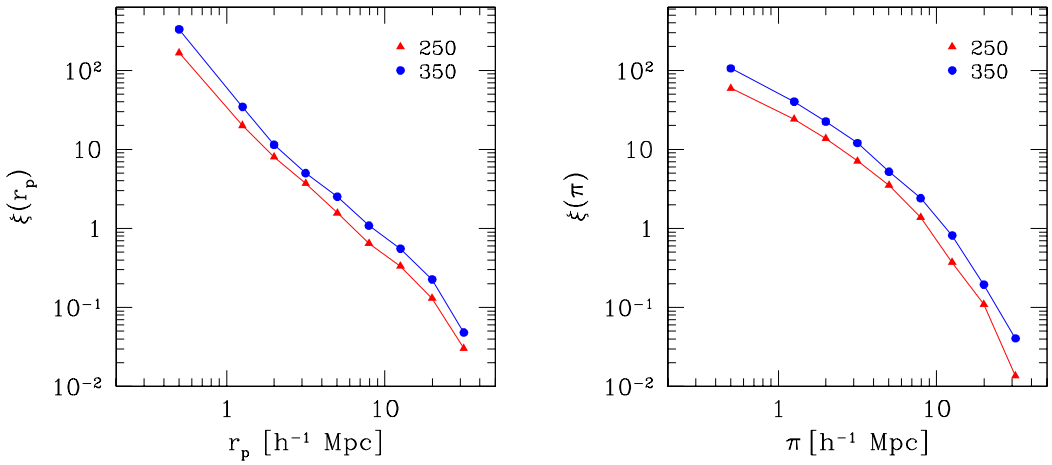



Figure 45. The projected correlation functions of halo samples at $z = 1$, calculated for $V_{\rm min} = 250$ and 350 km s$^{-1}$ and $f = 0.5$. The trend with $V_{\rm min}$ is mainly equivalent to a uniform scaling of the correlation amplitude at all separations.

the halo central galaxies or in smaller subhalos. As discussed above, the ratio $\xi(\pi)/\xi(r_p)$ should be a much more sensitive and direct indicator for the satellite fraction. We parameterize this fraction by the probability, $f$, for objects to reside in the central galaxies of host halos; $1-f$ is, therefore, the probability for objects to reside in any of the smaller subhalos. The two parameters, the $v_{\rm max}$ threshold, $V_{\rm min}$, and the probability $f$, fully specify the relation between our model AGN population and the dark matter halos and subhalos identified in the numerical simulations. $V_{\rm min}$ is primarily constrained by the correlation length of $\xi(r_p)$, and $f$ is mostly constrained by the $\xi(\pi)/\xi(r_p)$ ratio.

Algorithmically, we simulate the AGN locations by randomly drawing the halos and subhalos from the simulation box according to the parameters $V_{\rm min}$ and $f$. First, we select all bound structures (both halos and subhalos) with $v_{\rm max} > 80\,{\rm km\ s^{-1}}$. This threshold is slightly higher than the resolution limit in the simulations. We verified that the final results are nearly the same when this initial threshold is varied between $50\,{\rm km\ s^{-1}}$ and $100\,{\rm km\ s^{-1}}$. We then select only those subhalos which are contained within halos with $v_{\rm max}$ above the given value of $V_{\rm min}$. We then randomly put a small number of objects (10–100) within the simulation box[3]. With the probability $1-f$, the object is associated with one of the selected subhalos, and with the probability $f$ it is put in the center of one of the halos. This procedure is repeated multiple times randomly selecting one of the box axes as the line of sight. Using these simulated objects, we derive a model of $\xi(r_p)$ and $\xi(\pi)$ for each combination of $V_{\rm min}$ and $f$.

If we attempt to populate the halos in these simulations with AGNs without making a distinction of main halos and subhalos, we obtain a model correlation functions which are inconsistent with the data. For example, we can adjust the amplitude of the model correlation function by changing the $v_{\rm max}$ threshold for the halo to be able to host an AGN. The observed correlation length, $r_0 = 6.27\,h^{-1}\,{\rm Mpc}$, is matched if we consider all objects with $V_{\rm min} > 300\,{\rm km\ s^{-1}}$. However, the population of such halos and subhalos in the simulations shows significant redshift-space distortions (Fig. 43). In particular, the ratio $\xi(\pi)/\xi(r_p)$ is $\approx 1.8$ in the range of separations $2-6\,h^{-1}\,{\rm Mpc}$, which is clearly inconsistent

[3]This is done to approximate the low space density of *Chandra*/Boötes AGNS.

with observations (Fig. 41). Therefore, application of our more complicated model to distinguish the locations of central and satellite galaxies, is warranted.

### 30.3 Model correlation functions $\xi(r_p)$ and $\xi(\pi)$

The correlation functions $\xi(r_p)$ and $\xi(\pi)$ were derived from the simulation outputs on a grid of parameters within the range $V_{\rm min} \in [200; 370]$ km s$^{-1}$ and $f \in [0; 1]$. Examples are shown in Fig. 44 and 45. Comparison of the two panels in Fig. 44 illustrates the effect of $f$ on the correlation functions. For smaller $f$ (more objects located in satellite galaxies, left panel), the projected correlation function $\xi(r_p)$ shows a stronger component at $d < 1\,h^{-1}$ Mpc in excess of a power law extrapolation from larger separations. This excess is attributed to the "1-halo" term in the analytic halo model of the correlation function. At the same time, $\xi(\pi)$ shows a larger suppression of the correlation amplitude at $d < 1\,h^{-1}$ Mpc with respect to $\xi(r_p)$ at the same separations, and a stronger enhancement over $\xi(r_p)$ at $d = 2-10\,h^{-1}$ Mpc for small $f$. This is the consequence of a stronger "finger of God" effect in the case when more objects are located in the satellite galaxies. Unfortunately, the statistical uncertainties in the real data do not allow a detailed modeling of the observed $\xi(r_p)$ at small separations. Modeling of the $\xi(\pi)$ at small separations is further complicated by the effect of uncertainties in the redshift measurements (§ 29). At large separations, $\gtrsim 0.1$ of the simulation box size, the correlation functions derived from the simulations are not reliable [127, 44, 48]. Taking all these considerations into account, we will match the observed and model correlation functions in the intermediate range of radii, $d = 1 - 12\,h^{-1}$ Mpc.

First, we compute the correlation length, $r_0$, for each combination $(V_{\rm min}, f)$. This is done by fitting a power law function, $(r_p/r_0)^{-\gamma}$, to $\xi(r_p)$ in the range $r_p = 1 - 12\,h^{-1}$ Mpc. We then compute the ratio $\xi(\pi)/\xi(r_p)$ (an example is shown in Fig. 47) and fit it in the same range of separations with a modified log-normal function,

$$\phi(d) = 1 + A \exp\left(-\left[\frac{(\log d - \log d_0)^2}{2D^2}\right]^g\right). \qquad (6.4)$$

The index $g$ is fixed at the mean best-fit value for all $(V_{\rm min}, f)$ combinations, $g = 1.2$. The $\xi(\pi)/\xi(r_p)$ ratio derived from the simulations shows substantial

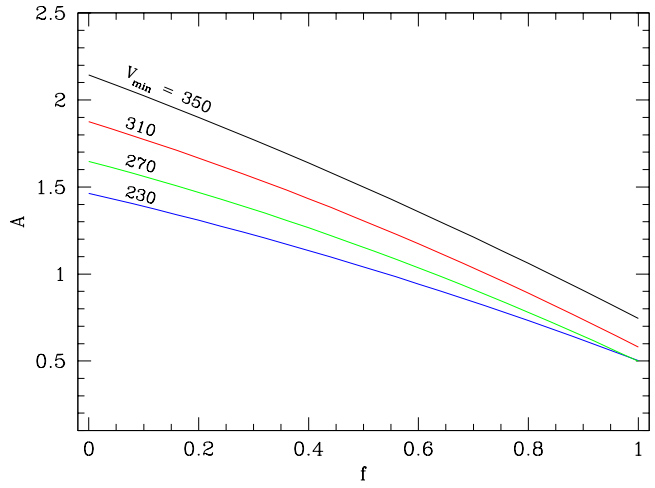



Figure 46. Dependence of amplitude $A$ (eq. 6.4) on the circular velocity threshold, $V_{\rm min}$, and the probability $f$.

variations related to cosmic variance (this can be estimated by comparing the results for three viewing angles, see errorbars in Fig. 47). Therefore, we need to smooth the results of fit given by eq. 6.4. This is achieved by fitting low-order polynomials to the parameters $D$, $d_0$, and $A$ as a function of $V_{\rm min}$ and $f$. We examined the dependence of $D$ and $d_0$ on $V_{\rm min}$ and $f$, and found that they show no obvious trend with $V_{\rm min}$ and may be approximated by linear functions of $f$. Then, we obtained the values of $A$ for each combination of $(V_{\rm min}, f)$ and fitted them with a second-order polynomial (Fig. 46). This procedure provides a smooth map of parameters from eq. 6.4 over the $(V_{\rm min}, f)$ range of interest. An example of the fitting function derived from this smooth map is shown in Fig. 47.

### *30.4 Application to the data*

In applying the correlation function model to the data we avoid including any sensitivity to the *slope* of the correlation function. The primary motivation is that the best-fit slope to the correlation functions derived from numerical simulations is slightly different from our best-fit value — the average power law slope of the model $\xi(r_p)$ functions is in the range $\gamma = 1.6 - 2.0$ for $d < 20\,h^{-1}$ Mpc separations vs. our best-fit value of $1.97 \pm 0.09$ (§ 29). This may indicate that our method of assigning the AGN locations to the dark matter halos is overly

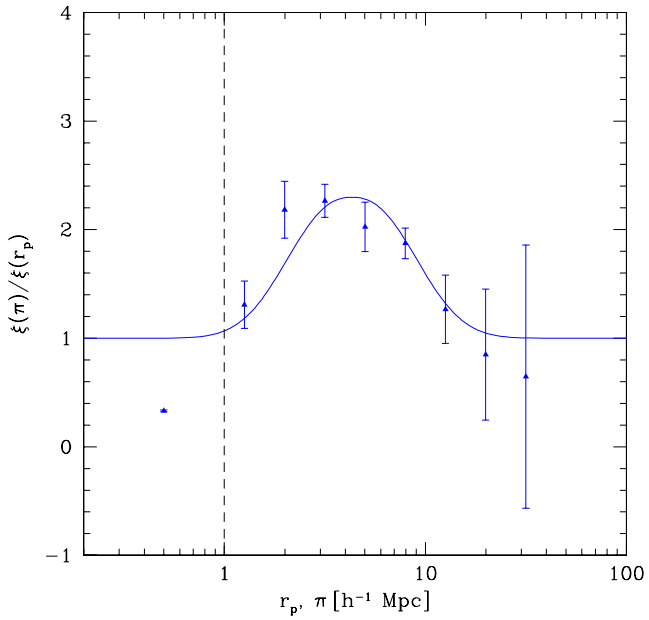



Figure 47. The ratio of the model correlation functions computed for $f = 0.5$ and $V_{\rm min} = 310\,{\rm km\ s^{-1}}$. Uncertainties at each separation are estimated using the variance of the ratio computed for three different viewing angles and reflect mostly the cosmic variance. The solid line shows the analytic approximation (eq. 6.4).

simplistic. Alternatively, this disagreement may arise because a combination of the cosmological parameters used in the simulation is slightly different from the currently accepted values, resulting in a possible mismatch between the slopes of the model and the real correlation function of astronomical objects. The most relevant combination of parameters is the product $\Omega_M h$. The most recent results from the WMAP 7-year data release [124] indicate $\Omega_M h = 0.19$ while in the simulations $\Omega_M h = 0.21$. Therefore, we rely on those properties of the correlation function which are less sensitive to the details of the modeling procedure and exact values of the cosmological parameters.

First, we use the value of the correlation length derived from fitting the $\xi(r_p)$ function (§ 29). Second, we use the ratio $x = \xi(\pi)/\xi(r_p)$ in the range of separations $1 - 12\,h^{-1}\,{\rm Mpc}$. For halos with circular velocities $v_{\rm max} \sim 300\,{\rm km\ s^{-1}}$ (as indicated by the amplitude of the AGN correlation function, see below), the "fingers of God" extend to $v_{\rm max}/H \sim 3\,h^{-1}\,{\rm Mpc}$, just in the middle of this range of separations. Formally, the constraints on the parameter of the AGN population

model, $V_{\rm min}$ and $f$, are derived using a $\chi^2$- function computed as

$$\chi^2 = \frac{(r_0 - r_0^{\rm mod})^2}{\sigma_{r0}^2} + \sum \frac{(x - x^{\rm mod})^2}{\sigma_x^2} \tag{6.5}$$

where the summation in the second term is over the data points in the $1 - 10\,h^{-1}$ Mpc separation range, and the model functions are those described in § 30.3.

*Redshift trends*

The procedure described above provides a correlation function model at the redshifts where the simulation outputs were saved. To measure the correlation length and the ratio of the $\xi(\pi)$ and $\xi(r_p)$ functions, we need to bin the *Chandra*/Boötes sample into several wide redshift intervals. Therefore we need to address an issue of a possible presence of $z$-dependent trends of the correlation function models.

Fortunately, for our choice of observables, $r_0$ and the $\xi(\pi)/\xi(r_p)$ ratio, the redshift trends are very weak. This is illustrated in Fig. 48. The upper panel shows $r_0$ as a function of $V_{\rm min}$ for a population model with $f = 1$ (all objects are at the centers of distinct halos) for the simulation outputs at $z = 0.5$, 1, 2, and 3. Obviously, there is almost no change in $r_0$ for a fixed $V_{\rm min}$ at $z < 3$. Any changes are much smaller than the uncertainties of our $r_0$ measurement even for the full samle. Therefore, we conclude that the model $r_0$ as a function of $V_{\rm min}$ does not evolve over our redshift range of interest. The ratio $\xi(\pi)/\xi(r_p)$ also shows little, if any, evolution with redshift. The bottom panel in Fig. 48 shows the results for the population model with $V_{\rm min} = 310\,{\rm km\ s^{-1}}$ and $f = 0.5$ (50% of objects are in the satellite subhalos of $v_{\rm max} > 310\,{\rm km\ s^{-1}}$ halos). Any difference between the simulation outputs is within the uncertainties (estimated from analyzing three different projections for each simulation output).

The lack of evolution in the clustering model over our redshift interval (and also the lack of detectable evolution of $r_0$ with $z$, see § 31.2 below) indicates that we can safely combine the data over the entire redshift range in the sample. Furthermore, there is no need to weight the models with the redshift distribution — one simply can use the results for the simulation output at $z = 1$. We take this approach in fitting the parameters of the population model, $V_{\rm min}$ and $f$

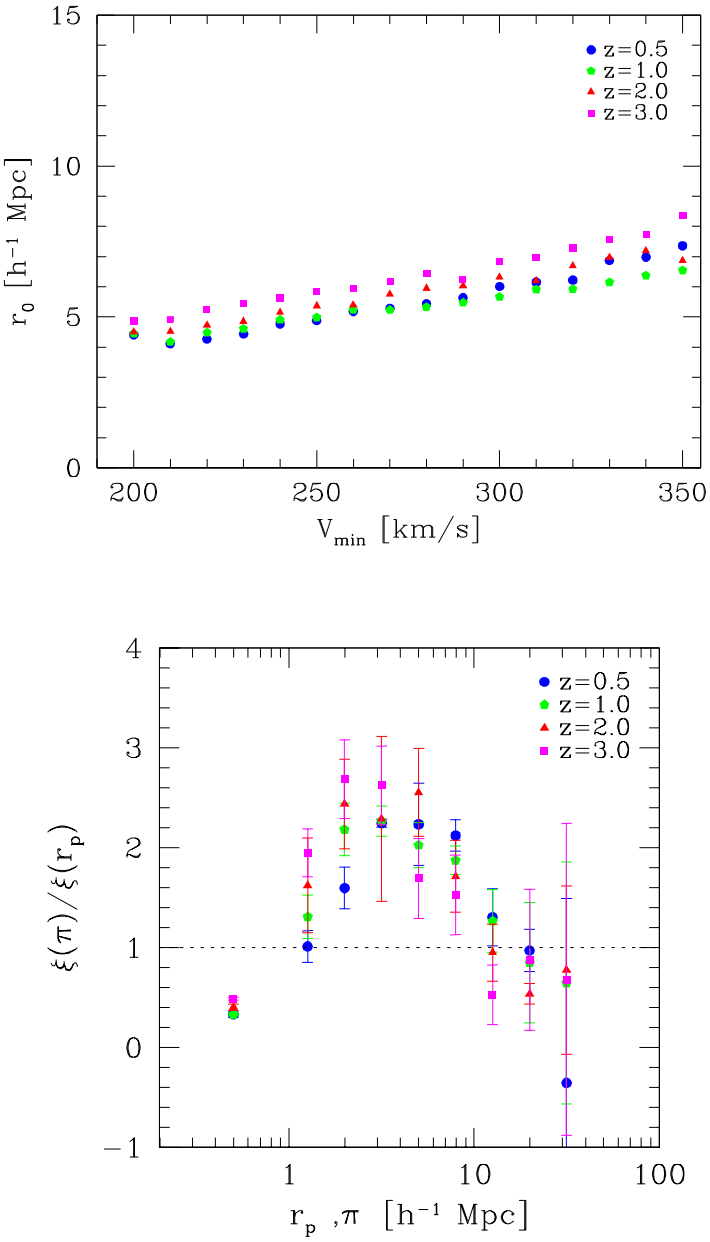



Figure 48. *Top:* Dependence of $r_0$ on $V_{\rm min}$ for the objects at the centers of distinct halos ($f = 1$) for the simulation outputs at $z = 0.5$, 1, 2, and 3. There is almost no change with redshift for $z \lesssim 2.5$. *Bottom:* the ratio of $\xi(\pi)/\xi(r_p)$ for the population model with $V_{\rm min} = 310\,{\rm km\,s^{-1}}$ and $f = 0.5$ (50% of objects are in the satellite galaxies of the $v_{\rm max} > 310\,{\rm km\,s^{-1}}$ halos). Any difference in this ratio between different simulation snapshots is within the uncertainties (estimated from analyzing 3 different projections for each snapshot).

Table VI.1. Clustering modeling results as a function of redshift

| Redshift range | $z_{\rm med}$ | $\langle \log L_x \rangle$ | $r_0$ | $V_{\rm min}$ | $M_{\rm lim}$ | $n$ | $n_{\rm halo}$ | $n/n_{\rm halo}$ |
|---|---|---|---|---|---|---|---|---|
| | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| $0.172 - 0.555..$ | 0.374 | 42.58 | $5.40 \pm 0.46$ | $271 \pm 23$ | $3.5 \pm 0.9$ | $1.54 \times 10^{-4}$ | $1.40^{+0.46}_{-0.32}$ | $0.110^{+0.032}_{-0.027}$ |
| $0.555 - 1.000..$ | 0.738 | 43.34 | $5.67 \pm 0.64$ | $287 \pm 33$ | $3.4 \pm 1.3$ | $6.12 \times 10^{-5}$ | $1.23^{+0.62}_{-0.38}$ | $0.050^{+0.022}_{-0.017}$ |
| $1.000 - 1.630..$ | 1.279 | 43.85 | $6.96 \pm 0.80$ | $362 \pm 45$ | $5.2 \pm 2.1$ | $2.56 \times 10^{-5}$ | $0.46^{+0.35}_{-0.19}$ | $0.056^{+0.039}_{-0.024}$ |
| $1.630 - 5.0....$ | 2.099 | 44.28 | $6.46 \pm 1.46$ | $308 \pm 70$ | $2.1 \pm 1.2$ | $8.61 \times 10^{-6}$ | $0.82^{+1.43}_{-0.50}$ | $0.011^{+0.016}_{-0.007}$ |

Note. — Column (4) — best-fit correlation length assuming a fixed slope of $\gamma = 1.97$, in units of $h^{-1}$ comoving Mpc. Column (5) — threshold maximum circular velocity of the host dark matter halos, in units of km s$^{-1}$ (§31.2). Column (6) — the halo virial mass corresponding to $V_{\rm min}$, in units of $10^{12}\,h^{-1}\,M_\odot$. Column (7) — comoving number density of X-ray sources at $z_{\rm med}$, in units of $h^3\,{\rm Mpc}^{-3}$. Column (8) — comoving number density of dark matter halos with $v_{\rm max} > V_{\rm min}$ at $z_{\rm med}$, in units of $10^{-3}\,h^3\,{\rm Mpc}^{-3}$. Column (9) — probability for the *Chandra*/Boötes AGNs to be in active state computed as the ratio of number density of AGNs and the dark matter halos with $v_{\rm max} > V_{\rm min}$.

(§ 31.1). In addition, we constrain the evolution of $V_{\rm min}$ with $z$ (§ 31.2) under the assumption that the fraction of AGNs in the subhalos does not evolve (i.e., using a fixed $f$ derived from the analysis of the entire sample).

*Adjusting the results to a low-$\sigma_8$ cosmology*

Finally, we note that the numerical simulations we use were run assuming a high value of the power spectrum normalization, $\sigma_8 = 0.9$ at $z = 0$. This results in an incorrect prediction of the correlation amplitude for halos of a given mass, and thus slightly biases the derived parameters of the AGN population model, in particular, $V_{\rm min}$. Obviously, it would be best to use the simulations performed for the currently favored cosmological model with $\sigma_8 \approx 0.8$ but in general, these were unavailable at the time of this investigation. We applied an analytical procedure, described in Appendix in [233], which allows one to scale results obtained from a simulation performed within a given cosmological model to any desired cosmology. In particular, if we use the best-fit flat ΛCDM cosmological model derived from the joint analysis of the galaxy cluster mass function and other cosmological datasets, $\sigma_8 = 0.786$ and $\Omega_M = 0.268$ [252], the masses reported below should be scaled by a factor of 0.69, the $V_{\rm min}$'s decreased by 10%, and the number density of objects increased by 20%.

## 31 Modeling results

### *31.1 $V_{\rm min}$ and the satellite sraction for the entire sample*

Figure 49 shows the combined constraints on the population model parameters, $V_{\rm min}$ and $f$, obtained from fitting the full sample of *Chandra*/Boötes AGNs. The best-fit velocity threshold is $V_{\rm min} = 310 \pm 13\,{\rm km\ s^{-1}}$ (68% CL one-parameter uncertainty). At $z = 1$, this corresponds to $M_{180} = 3.7 \times 10^{12}\,h^{-1}\,M_\odot$, or $2.2 \times 10^{12}\,h^{-1}\,M_\odot$ after correcting for the lower-$\sigma_8$ cosmology (see § 30.4 and Appendix in [233])[4]. We thus conclude, in agreement with the earlier studies [92, 102], that

[4] [102] quote a higher mass, $\sim 10^{13}\,h^{-1}\,M_\odot$ for an X-ray AGNs sample with the mean $\langle z\rangle = 0.51$, from which they measure $r_0 = 4.8\,h^{-1}$ Mpc. The main source of the difference, as explained in [46, see their p.1195], is a more accurate model for the matter power spectrum at galactic scales deployed in modern simulations such as those we use here.

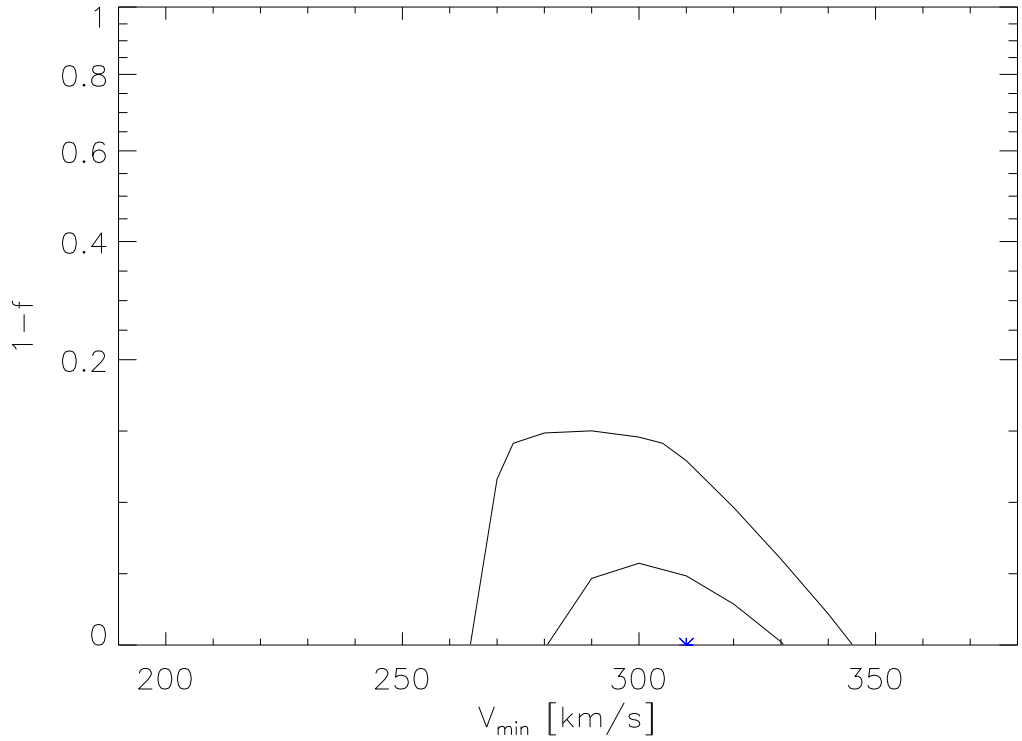



Figure 49. Combined constraints on halo occupation parameters, $V_{\rm min}$ and $f$, obtained from the full sample of the *Chandra*/Boötes AGNs. The best fit ($V_{\rm min} = 310\,{\rm km\ s^{-1}}$ and $f = 1$) is indicated with a star. The contours correspond to 68% and 95% confidence regions for two interesting parameters ($\Delta\chi^2 = 2.3$ and 6.2, respectively). The quantity $1 - f$ represents the fraction of AGNs residing in the satellite galaxies.

the X-ray AGNs at high redshifts reside in small galaxy groups with masses of a factor of a few above the present-day mass of the Milky Way.

As expected from the striking agreement of the observed $\xi(r_p)$ and $\xi(\pi)$ projected correlation function, the best fit is $f = 1$, i.e. *all AGNs are located at the centers of distinct dark matter halos*. The 95% CL upper limit on the satellite fraction is $1 - f < 0.09$. This limit is lower than the 11% fraction of $v_{\rm max} > 310\,{\rm km\ s^{-1}}$ objects located within bigger halos in the simulation. We thus find that the X-ray AGNs at $z = 1$ tend to *avoid* big galaxies in the outskirts of yet more massive groups and clusters. They definitely avoid satellite galaxies in the $v_{\rm max} \gtrsim 300\,{\rm km\ s^{-1}}$ galaxy groups because otherwise we would find $1 - f = 0.3 - 0.6$.[5] Instead, the X-ray AGNs are preferentially located at the centers of distinct dark matter halos.

[5]For example, there are 908 subhalos with $v_{\rm max} > 201\,{\rm km\ s^{-1}}$ located within 1650 $V_{\rm min} = 310\,{\rm km\ s^{-1}}$ halos, leading to $f = 0.65$.

### 31.2 $V_{\rm min}$ as a function of redshift

Splitting our *Chandra*/Boötes AGN sample into 4 subsamples, $\sim 320$ objects in each, we can measure the correlation length, $r_0$, as a function of redshift if we hold the slope of the correlation function fixed at the best-fit value obtained for the entire sample, $\gamma = 1.97$. The results are shown in Fig. 50. There is almost no change in $r_0$ over the redshift interval $z = 0.35 - 2.1$. A formal linear fit to the $r_0(z)$ measurements gives a change from $r_0 = 5.5\,h^{-1}$ Mpc at $z = 0.35$ to $r_0 \approx 7\,h^{-1}$ Mpc at $z = 2$. However, all the measurements are consistent with the average $r_0 = 6.27\,h^{-1}$ Mpc derived for the entire sample, so the trend with redshift is not significant.

Assuming further a fixed $f = 1$ at each $z$, as indicated by modeling of the redshift-space distortions in the entire sample, we can convert the best-fit values of $r_0$ at each redshift to the threshold circular velocity for the parent dark matter halos. The results are shown in Fig. 51. We find no detectable trend of $V_{\rm min}$ or the corresponding mass threshold, $M_{\rm lim}$ (see Table VI.1), with redshift, either. This appears somewhat counterintuitive because in a flux-limited sample, such as ours, the objects at higher redshift have higher intrinsic luminosities, and we might expect them to be located in more massive dark matter halos. We note, however, that the studies of optically selected QSOs also indicate a weak or no trend of clustering length with the object luminosity [53, 222]. Clustering analysis of the SDSS quasars [210] shows mild or no evolution of the real-space correlation length at $z \lesssim 2$.

### 31.3 Constraints on AGN duty cycle

Having estimated the mass scale and therefore the space density of *Chandra*/Boötes AGNs, we can formally compute their duty cycle following the approach of [160]. Under the simplifying assumption that the halo lifetime is approximately independent of mass, the probability for the AGN to be active is simply

$$P_{\rm active} = \frac{n_{\rm AGN}}{n_{\rm halo}}. \tag{6.6}$$

The limiting X-ray luminosity is ill-defined for our sample because the X-ray detection is extended down to very low limits in terms of the number of detected

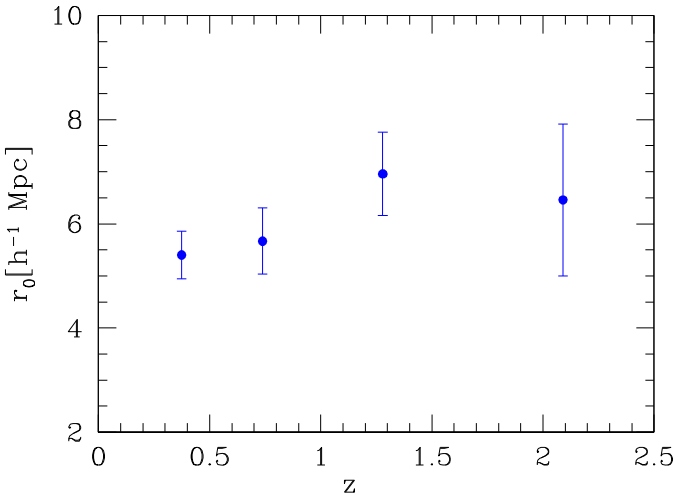



Figure 50. The correlation lengths, $r_0$, measured in four redshift intervals, $z$ = [0.172 − 0.555], [0.555 − 1.000], [1.000 − 1.630], and [1.630 − 5.0], containing 320, 307, 344, and 311 objects, respectively. The values of $r_0$ were obtained by fitting the projected correlation function, $w(r_p)$, in each bin assuming the fixed slope of the correlation function $\gamma$ = 1.97 (the best-fit value for the entire sample).

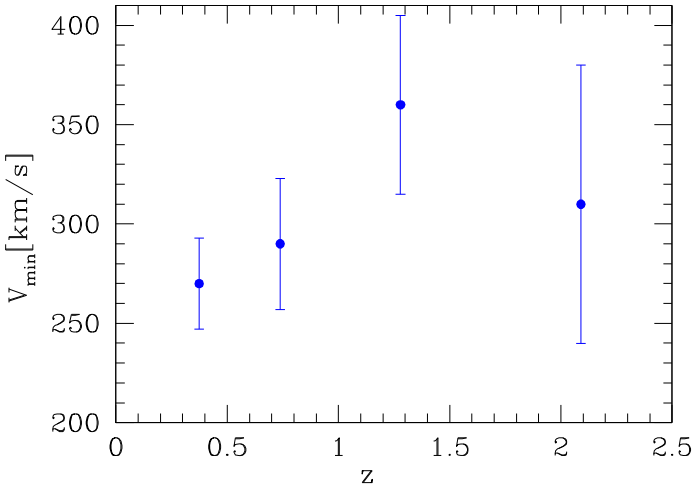



Figure 51. Minimum rotational velocity for dark matter halos, corresponding to the correlation lengths in Fig. 50.

X-ray photons ($\geq$ 4, see [115]), and in this regime, a wide range of possible intrinsic intensities corresponds to the given number of photons [116]. However, as a guide for the typical luminosity scale one can use the median $\log L_x$ reported in Table VI.1. With these caveats, the probabilities given by eq. 6.6 and reported in column (9) of Table VI.1 correspond to the probability for the dark matter halos with $v_{\rm max} > V_{\rm min}$ to host an AGN with the instantaneous soft-band X-ray luminosity of order $\langle \log L_x \rangle$ or above. Our results indicate that $P_{\rm active}$ generally declines with mean luminosity and/or redshift. Interestingly, our results indicate that AGNs are quite common at low redshifts — approximately 10% of dark matter halos with $v_{\rm max} > 270\,{\rm km\ s^{-1}}$ (or mass $M_{180} > 3.5 \times 10^{12}\,h^{-1}\,M_\odot$) host an AGN with a soft-band X-ray luminosity of $\sim 4 \times 10^{42}\,{\rm erg\ s^{-1}}$ or above.

## 32 Conclusions

We measured the clustering properties of X-ray selected AGNs using a sample of 1282 sources with spectroscopic redshifts from the 9.3 deg$^2$ *Chandra* survey in the Boötes region — the most accurate such measurement to date. In agreement with previous studies of X-ray and optically selected AGNs, we find that the real-space correlation function can be approximated in the radial range $1-20\,h^{-1}$ comoving Mpc by a power law with a slope of $\gamma = 1.97 \pm 0.09$. The correlation length is $\approx 6.5\,h^{-1}\,{\rm Mpc}$, showing only weak trends with redshift at $z = 0.3 - 2$ (or with X-ray luminosity).

Matching the observed clustering properties of the *Chandra*/Boötes AGNs to those of dark matter halos in the high-resolution cosmological simulations, we find that the X-ray AGNs reside in halos with the maximum rotational velocity $\approx 310\,{\rm km\ s^{-1}}$, or with total masses $\sim 3.7 \times 10^{12}\,h^{-1}\,M_\odot$, also with no detectable redshift trend. The lack of a redshift or luminosity dependence of the AGN clustering is inconsistent with the scenarios in which the AGN luminosities in the active state are similar fractions of the Eddington luminosity. However, it can be explained in the scenario in which the AGN activity is triggered by major mergers of gas-rich galaxies, and the instantaneous luminosity passes through many levels after each trigger [105].

Our results reveal another interesting aspect of the AGN clustering which

was predicted in [106]. The redshift measurements in our sample are sufficiently accurate to detect peculiar motions of objects in excess of $\sim 100\,\mathrm{km\ s^{-1}}$. The comparison of the two-point correlation functions projected on the line of sight and on the sky plane reveals no signatures of the redshift-space distortions, which allows us to put limits on the fraction of AGNs located in the satellite subhalos *within* the host dark matter halos. We find that the X-ray AGNs are predominantly located in the central galaxies of the host dark matter halos and tend to *avoid* satellite galaxies. Quantitatively, we limit the fraction of AGNs in non-central galaxies to be $< 0.09$ at the 95% CL, less than that expected from the abundance of satellite subhalos found in the numerical simulations (§ 31.1). The central locations of the quasar host galaxies are expected in the trigger model because mergers of equally-sized galaxies preferentially occur at the centers of dark matter halos [106].

Finally, we compared the number densities of the *Chandra*/Boötes AGNs to that of the dark matter halos with the mass corresponding to the AGN clustering amplitude. We find that the fraction of halos with active X-ray AGNs decreases with increasing $z$ — and, correspondingly, with $L_x$ — probably reflecting the lower probability for an object to have a higher instantaneous luminosity. At the lowest redshifts in our sample, *Chandra* probes such a low luminosity that X-ray AGNs become quite common. At $z = 0.37$, the *Chandra*-detected sources are located in more than 10% of the dark matter halos with $v_{\mathrm{max}} > 270\,\mathrm{km\ s^{-1}}$ or $M > 3.55 \times 10^{12}\,h^{-1}\,M_{\odot}$.

# VII Conclusions

We studied clustering of cosmic sources detected in near- and mid-infrared by the *Spitzer* Space Telescope and in X-rays by the *Chandra* X-ray Observatory. The analysis is focused on three classes of high-redshift ($z > 1$) sources — star-forming and passively evolving galaxies, and active galactic nuclei. Depending on the quality of available data, we applied the spatial or angular correlation function formalism to derive the clustering strengths, from which we put constraints on the halo occupation parameters for these sources using the correlation function models for the dark matter halos derived from cosmological numerical simulations.

The analysis of the first class of sources, high-$z$ star-forming galaxies, is based on the sample of $\sim 20000$ objects detected at 24 $\mu$m in the SWIRE Lockman Hole field ($\sim 9\,\mathrm{deg}^2$) above a flux limit of $S_\nu(24\,\mu\mathrm{m}) > 310\,\mu\mathrm{Jy}$. Using optical/near-IR colors, we divided the sample into low-redshift ($z_{\mathrm{mean}} = 0.68$) and high-redshift ($z_{\mathrm{mean}} = 1.72$) subsamples. The spatial clustering lengths derived for these subsamples, $r_0 = 4.74 \pm 0.16\,h^{-1}\,\mathrm{Mpc}$ and $7.87 \pm 0.63\,h^{-1}\,\mathrm{Mpc}$, respectively, imply the total mass scales of parent dark matter halos $M_{\mathrm{tot}} \gtrsim 5 \times 10^{11}\,h^{-1}\,M_\odot$ at $z_{\mathrm{mean}} = 0.68$ and $M_{\mathrm{tot}} \gtrsim 2 \times 10^{12}\,h^{-1}\,M_\odot$ at $z_{\mathrm{mean}} = 1.72$. The number density of such halos is a factor of $\sim 5$ higher than the number density of 24 $\mu$m sources in both redshift bins. Approximate estimates of the IR-luminosities based on the catalogues of 24 $\mu$m sources in the GOODS fields [207] showed that our high-$z$ subsample represents a population of "distant ultra-luminous infrared galaxies" (ULIRGs) with $L_{\mathrm{IR}} > 10^{12} L_\odot$, while the low-$z$ subsample mainly consists of "LIRGs", $L_{\mathrm{IR}} \sim 10^{11} L_\odot$. Based on these results, we concluded that our low- and high-redshift 24 $\mu$m-selected objects represent different populations of galaxies hosted by halos from different mass ranges. Their high level of mid-IR lumi-

nosities may be caused by similar physical processes (e.g., triggered by mergers or interactions), but occurring in different environments.

Further, we investigated the clustering properties of massive, passively evolving galaxies which already formed large amount of stellar mass by $z \sim 2$. This sample was previously selected in the SWIRE ELAIS-S1 field ($\sim 4\,\mathrm{deg}^2$) using the presence of the 1.6 $\mu$m stellar peak redshifted into the 5.8 $\mu$m *Spitzer*/IRAC band. The angular correlation function of these objects, called "IR-peakers", is detected with a $\gtrsim 4\sigma$ significance at small scales $\sim 0.01^\circ$. We showed that the modeling of these measurements based on the power-law fit to the correlation function led to a biased (low) estimation of the spatial correlation length $r_0$. The proper results can be obtain if one models the observed angular correlation function using the full correlation function model for the dark matter halos derived from the cosmological numerical simulation. This analysis showed that the true correlation length of IR-peakers is $r_0 = 7.7 \pm 1.1\,h^{-1}$ Mpc, implying that they are hosted by those $10-25\%$ dark matter halos with total mass $M_{\rm tot} \approx [0.5{-}1]{\times}10^{13} M_\odot$, in which the star formation efficiency reaches $\sim 10{-}20\%$, which is somewhat higher than the value estimated for present-day galaxies of a similar mass [96]. The clustering properties of IR-peakers also can be used to rule out that these objects are located in only central or only satellite galaxies of the parent dark matter halos.

Lastly, we discussed the halo occupation properties of high-redshift galaxies with active nuclei selected by their X-ray emission. We measured the spatial clustering of X-ray sources selected from the *Chandra* survey in the Boötes field. For the whole sample (with $z_{\rm med} \sim 1$), we measure the correlation length $r_0 \approx 6.5\,h^{-1}$ Mpc, and find no statistically significal evolution in $r_0$ over $z = 0.37 - 2.1$, evem though the typical X-ray luminosity in the *Chandra* sample varies by a factor of $\sim 50$ between these redshifts. From matching the observed clustering properties of the *Chandra* sources to those of dark matter halos in the high-resolution cosmological simulations, we found that the X-ray AGNs reside in halos with total masses $\sim 2 \times 10^{12}\,h^{-1}\,M_\odot$, also with no detectable redshift trend. Peculiar motions of galaxies within the halos of this mass would have caused detectable distortions in the observed correlation function along the line of sight. We have not detected such distortions, and the upper limit allowed us to show

that the high-$z$ X-ray emitting AGNs preferentially reside in the centers of their host dark matter halos and tend to avoid satellite galaxies.

Thus, the three classes of objects studied in this thesis have a similar correlation length ($r_0 \approx 7 - 8\,h^{-1}$ Mpc), and hence reside in dark matter halos of similar mass, a few $\times 10^{12}\,h^{-1}\,M_{\odot}$. However, they represent distinct and mostly non-overlapping populations. The investigation of possible connections between them will be the subject of a future work. Taken together, the objects studied in this thesis populate $\sim 30 - 50\%$ of the massive dark matter halos at $z = 1 - 2$.

# Bibliography


[1] Aarseth, S. J., *NBODY2: A direct N-body integration code*. 2001, New Astronomy, 6, 277.

[2] Adelberger, K., *Star Formation and Structure Formation at* $1 < z < 3$. In *KITP Program: Galaxy Formation and Evolution* (2000).

[3] Adelberger, K. L., Steidel, C. C., Pettini, M., Shapley, A. E., Reddy, N. A., Erb, D. K., *The Spatial Clustering of Star-forming Galaxies at Redshifts* $1.4 <\sim z <\sim 3.5$. 2005, ApJ, 619, 697.

[4] Allgood, B., Flores, R. A., Primack, J. R., Kravtsov, A. V., Wechsler, R. H., Faltenbacher, A., Bullock, J. S., *The shape of dark matter haloes: dependence on mass, redshift, radius and formation*. 2006, MNRAS, 367, 1781.

[5] Auger, M. W., Treu, T., Bolton, A. S., Gavazzi, R., Koopmans, L. V. E., Marshall, P. J., Moustakas, L. A., Burles, S., *The Sloan Lens ACS Survey. X. Stellar, Dynamical, and Total Mass Correlations of Massive Early-type Galaxies*. 2010, ApJ, 724, 511.

[6] Bagla, J. S., *Cosmological N-Body simulation: Techniques, Scope and Status*. 2005, Current Science, 88, 1088.

[7] Ballinger, W. E., Peacock, J. A., Heavens, A. F., *Measuring the cosmological constant with redshift surveys*. 1996, MNRAS, 282, 877.

[8] Barkana, R., Loeb, A., *In the beginning: the first sources of light and the reionization of the universe*. 2001, Phys. Rep., 349, 125.

[9] Barrow, J. D., Bhavsar, S. P., Sonoda, D. H., *A bootstrap resampling analysis of galaxy clustering*. 1984, MNRAS, 210, 19P.

[10] Basu-Zych, A. R., Schiminovich, D., Heinis, S., et al., *Studying Large- and Small-Scale Environments of Ultraviolet Luminous Galaxies*. 2009, ApJ, 699, 1307.

[11] Baugh, C. M., Lacey, C. G., Frenk, C. S., Granato, G. L., Silva, L., Bressan, A., Benson, A. J., Cole, S., *Can the faint submillimetre galaxies be explained in the* $\Lambda$ *cold dark matter model?* 2005, MNRAS, 356, 1191.

[12] Berlind, A. A., Weinberg, D. H., *The Halo Occupation Distribution: Toward an Empirical Determination of the Relation between Galaxies and Mass*. 2002, ApJ, 575, 587.

[13] Berlind, A. A., Weinberg, D. H., Benson, A. J., et al., *The Halo Occupation Distribution and the Physics of Galaxy Formation*. 2003, ApJ, 593, 1.

[14] Bernardeau, F., Colombi, S., Gaztañaga, E., Scoccimarro, R., *Large-scale structure of the Universe and cosmological perturbation theory*. 2002, Phys. Rep., 367, 1.

[15] Berrier, J. C., Bullock, J. S., Barton, E. J., Guenther, H. D., Zentner, A. R., Wechsler, R. H., *Close Galaxy Counts as a Probe of Hierarchical Structure Formation*. 2006, ApJ, 652, 56.

[16] Berta, S., Fritz, J., Franceschini, A., Bressan, A., Lonsdale, C., *Photometric estimates of stellar masses in high-redshift galaxies*. 2004, A&A, 418, 913.

[17] Berta, S., Fritz, J., Franceschini, A., Bressan, A., Pernechele, C., *Spatially-resolved spectrophotometric analysis and modelling of the Superantennae*. 2003, A&A, 403, 119.

[18] Berta, S., Lonsdale, C. J., Polletta, M., et al., *The contribution of very massive high-redshift SWIRE galaxies to the stellar mass function*. 2007, A&A, 476, 151.

[19] Berta, S., Lonsdale, C. J., Siana, B., et al., *Keck spectroscopy of z = 1-3 ULIRGs from the Spitzer SWIRE survey*. 2007, A&A, 467, 565.

[20] Berta, S., Rubele, S., Franceschini, A., et al., *The ESO-Spitzer Imaging extragalactic Survey (ESIS). II. VIMOS I, z wide field imaging of ELAIS-S1 and selection of distant massive galaxies*. 2008, A&A, 488, 533.

[21] Bertin, E., Arnouts, S., *SExtractor: Software for source extraction.* 1996, A&AS, 117, 393.

[22] Bertschinger, E., *Simulations of Structure Formation in the Universe*. 1998, ARA&A, 36, 599.

[23] Blain, A. W., Chapman, S. C., Smail, I., Ivison, R., *Clustering of Submillimeter-selected Galaxies*. 2004, ApJ, 611, 725.

[24] Blake, C., Collister, A., Lahav, O., *Halo-model signatures from 380000 Sloan Digital Sky Survey luminous red galaxies with photometric redshifts*. 2008, MNRAS, 385, 1257.

[25] Blumenthal, G. R., Faber, S. M., Primack, J. R., Rees, M. J., *Formation of galaxies and large-scale structure with cold dark matter*. 1984, Nature, 311, 517.

[26] Bolzonella, M., Miralles, J., Pelló, R., *Photometric redshifts based on standard SED fitting procedures*. 2000, A&A, 363, 476.

[27] Borgani, S., *Scaling in the Universe.* 1995, Phys. Rep., 251, 1.

[28] Borgani, S., *Advanced Stastical Methods for Large-Scale Structure Studies*. In S. Bonometto, J. R. Primack, & A. Provenzale, ed., *Dark Matter in the Universe*, (pp. 209–+) (1996).

[29] Borgani, S., Kravtsov, A., *Cosmological simulations of galaxy clusters*. 2009, ArXiv e-prints.

[30] Bouwens, R. J., Illingworth, G. D., Franx, M., Ford, H., *UV Luminosity Functions at z˜4, 5, and 6 from the Hubble Ultra Deep Field and Other Deep Hubble Space Telescope ACS Fields: Evolution and Star Formation History*. 2007, ApJ, 670, 928.

[31] Boylan-Kolchin, M., Springel, V., White, S. D. M., Jenkins, A., Lemson, G., *Resolving cosmic structure formation with the Millennium-II Simulation*. 2009, MNRAS, 398, 1150.

[32] Brand, K., Brown, M. J. I., Dey, A., et al., *The Chandra XBoötes Survey. III. Optical and Near-Infrared Counterparts*. 2006, ApJ, 641, 140.

[33] Brand, K., Moustakas, J., Armus, L., et al., *The Origin of the 24 μm Excess in Red Galaxies*. 2009, ApJ, 693, 340.

[34] Brodwin, M., Dey, A., Brown, M. J. I., et al., *Clustering of Dust-Obscured Galaxies at z ˜ 2*. 2008, ApJ, 687, L65.

[35] Brown, M. J. I., Zheng, Z., White, M., et al., *Red Galaxy Growth and the Halo Occupation Distribution*. 2008, ApJ, 682, 937.

[36] Bryan, G. L., Norman, M. L., *Statistical Properties of X-Ray Clusters: Analytic and Numerical Comparisons*. 1998, ApJ, 495, 80.

[37] Bullock, J. S., Dekel, A., Kolatt, T. S., Kravtsov, A. V., Klypin, A. A., Porciani, C., Primack, J. R., *A Universal Angular Momentum Profile for Galactic Halos*. 2001, ApJ, 555, 240.

[38] Cacciato, M., van den Bosch, F. C., More, S., Li, R., Mo, H. J., Yang, X., *Galaxy clustering and galaxy-galaxy lensing: a promising union to constrain cosmological parameters*. 2009, MNRAS, 394, 929.

[39] Cappelluti, N., Böhringer, H., Schuecker, P., Pierpaoli, E., Mullis, C. R., Gioia, I. M., Henry, J. P., *The soft X-ray cluster-AGN spatial cross-correlation function in the ROSAT-NEP survey*. 2007, A&A, 465, 35.

[40] Cappi, M., Mazzotta, P., Elvis, M., et al., *Chandra Study of an Overdensity of X-Ray Sources around Two Distant (Z˜0.5) Clusters*. 2001, ApJ, 548, 624.

[41] Cen, R., Ostriker, J. P., *Physical Bias of Galaxies from Large-Scale Hydrodynamic Simulations*. 2000, ApJ, 538, 83.

[42] Chary, R., Elbaz, D., *Interpreting the Cosmic Infrared Background: Constraints on the Evolution of the Dust-enshrouded Star Formation Rate*. 2001, ApJ, 556, 562.

[43] Coles, P., Lucchin, F., *Cosmology: The Origin and Evolution of Cosmic Structure, Second Edition* (Wiley-VCH, 2002).

[44] Colín, P., Valenzuela, O., Klypin, A., *Bars and Cold Dark Matter Halos*. 2006, ApJ, 644, 687.

[45] Conroy, C., Prada, F., Newman, J. A., et al., *Evolution in the Halo Masses of Isolated Galaxies between $z \sim 1$ and $z \sim 0$: From DEEP2 to SDSS*. 2007, ApJ, 654, 153.

[46] Conroy, C., Shapley, A. E., Tinker, J. L., Santos, M. R., Lemson, G., *The Varied Fates of $z \sim 2$ Star-forming Galaxies*. 2008, ApJ, 679, 1192.

[47] Conroy, C., Wechsler, R. H., *Connecting Galaxies, Halos, and Star Formation Rates Across Cosmic Time*. 2009, ApJ, 696, 620.

[48] Conroy, C., Wechsler, R. H., Kravtsov, A. V., *Modeling Luminosity-dependent Galaxy Clustering through Cosmic Time*. 2006, ApJ, 647, 201.

[49] Cooray, A., *Second Moment of Halo Occupation Number*. 2002, ApJ, 576, L105.

[50] Cooray, A., Amblard, A., Wang, L., et al., *HerMES: Halo occupation number and bias properties of dusty galaxies from angular clustering measurements*. 2010, A&A, 518, L22+.

[51] Cooray, A., Sheth, R., *Halo models of large scale structure*. 2002, Phys. Rep., 372, 1.

[52] Crocce, M., Pueblas, S., Scoccimarro, R., *Transients from initial conditions in cosmological simulations*. 2006, MNRAS, 373, 369.

[53] Croom, S. M., Boyle, B. J., Shanks, T., et al., *The 2dF QSO Redshift Survey - XIV. Structure and evolution from the two-point correlation function*. 2005, MNRAS, 356, 415.

[54] Croton, D. J., Gao, L., White, S. D. M., *Halo assembly bias and its effects on galaxy clustering*. 2007, MNRAS, 374, 1303.

[55] Daddi, E., Dickinson, M., Morrison, G., et al., *Multiwavelength Study of Massive Galaxies at $z \sim 2$. I. Star Formation and Galaxy Growth*. 2007, ApJ, 670, 156.

[56] Davé, R., Finlator, K., Oppenheimer, B. D., Fardal, M., Katz, N., Kereš, D., Weinberg, D. H., *The nature of submillimetre galaxies in cosmological hydrodynamic simulations*. 2010, MNRAS, 404, 1355.

[57] Davis, M., Efstathiou, G., Frenk, C. S., White, S. D. M., *The evolution of large-scale structure in a universe dominated by cold dark matter*. 1985, ApJ, 292, 371.

[58] Davis, M., Peebles, P. J. E., *Large Scale Structure of galaxies*. 1983, ApJ, 267, 465.

[59] de Bernardis, P., Ade, P. A. R., Bock, J. J., et al., *A flat Universe from high-resolution maps of the cosmic microwave background radiation*. 2000, Nature, 404, 955.

[60] de la Torre, S., Guzzo, L., Kovač, K., et al., *Understanding the shape of the galaxy two-point correlation function at $z \simeq 1$ in the COSMOS field*. 2010, MNRAS, 409, 867.

[61] de la Torre, S., Le Fèvre, O., Arnouts, S., et al., *VVDS-SWIRE. Clustering evolution from a spectroscopic sample of galaxies with redshift $0.2 < z < 2.0$ selected from Spitzer IRAC 3.6 $\mu$m and 4.5 $\mu$m photometry*. 2007, A&A, 475, 443.

[62] de Lapparent, V., Geller, M. J., Huchra, J. P., *A slice of the universe*. 1986, ApJ, 302, L1.

[63] Desai, V., Soifer, B. T., Dey, A., et al., *Redshift Distribution of Extragalactic* 24 $\mu$m *Sources*. 2008, ApJ, 679, 1204.

[64] Dey, A., Soifer, B. T., Desai, V., et al., *A Significant Population of Very Luminous Dust-Obscured Galaxies at Redshift* $z \sim 2$. 2008, ApJ, 677, 943.

[65] Dickinson, M., Papovich, C., Ferguson, H. C., Budavári, T., *The Evolution of the Global Stellar Mass Density at* $0 < z < 3$. 2003, ApJ, 587, 25.

[66] Efstathiou, G., Eastwood, J. W., *On the clustering of particles in an expanding universe*. 1981, MNRAS, 194, 503.

[67] Einasto, J., *Large scale structure of the Universe*. In R. Ruffini & G. Vereshchagin, ed., *American Institute of Physics Conference Series*, vol. 1205 of *American Institute of Physics Conference Series*, (pp. 72–81) (2010).

[68] Eisenstein, D. J., Zaldarriaga, M., *Correlations in the Spatial Power Spectra Inferred from Angular Clustering: Methods and Application to the Automated Plate Measuring Survey*. 2001, ApJ, 546, 2.

[69] Eke, V. R., Cole, S., Frenk, C. S., *Cluster evolution as a diagnostic for Omega*. 1996, MNRAS, 282, 263.

[70] Fabricant, D., Fata, R., Roll, J., et al., *Hectospec, the MMT's 300 Optical Fiber-Fed Spectrograph*. 2005, PASP, 117, 1411.

[71] Fall, S. M., Efstathiou, G., *Formation and rotation of disc galaxies with haloes*. 1980, MNRAS, 193, 189.

[72] Farrah, D., Lonsdale, C. J., Borys, C., et al., *The Spatial Clustering of Ultraluminous Infrared Galaxies over* $1.5 < z < 3$. 2006, ApJ, 641, L17.

[73] Farrah, D., Lonsdale, C. J., Weedman, D. W., et al., *The Nature of Star Formation in Distant Ultraluminous Infrared Galaxies Selected in a Remarkably Narrow Redshift Range*. 2008, ApJ, 677, 957.

[74] Fazio, G. G., Hora, J. L., Allen, L. E., et al., *The Infrared Array Camera (IRAC) for the Spitzer Space Telescope*. 2004, ApJS, 154, 10.

[75] Ferrarese, L., Ford, H., *Supermassive Black Holes in Galactic Nuclei: Past, Present and Future Research*. 2005, Space Sci. Rev., 116, 523.

[76] Fiolet, N., Omont, A., Polletta, M., et al., *Multi-wavelength properties of Spitzer selected starbursts at* $z \sim$. 2009, A&A, 508, 117.

[77] Fisher, K. B., Davis, M., Strauss, M. A., Yahil, A., Huchra, J. P., *Clustering in the 1.2-JY IRAS Galaxy Redshift Survey - Part Two - Redshift Distortions and Xi/r/p, PI*. 1994, MNRAS, 267, 927.

[78] Förster Schreiber, N. M., Genzel, R., Lutz, D., Kunze, D., Sternberg, A., *Near-Infrared Integral Field Spectroscopy and Mid-Infrared Spectroscopy of the Starburst Galaxy M82*. 2001, ApJ, 552, 544.

[79] Franceschini, A., Hasinger, G., Miyaji, T., Malquori, D., *On the relationship between galaxy formation and quasar evolution*. 1999, MNRAS, 310, L5.

[80] Franceschini, A., Manners, J., Polletta, M. d. C., et al., *A Complete Multiwavelength Characterization of Faint Chandra X-Ray Sources Seen in the Spitzer Wide-Area Infrared Extragalactic (SWIRE) Survey*. 2005, AJ, 129, 2074.

[81] Franceschini, A., Rodighiero, G., Vaccari, M., Berta, S., Marchetti, L., Mainetti, G., *Galaxy evolution from deep multi-wavelength infrared surveys: a prelude to Herschel*. 2010, A&A, 517, A74+.

[82] Franx, M., Labbé, I., Rudnick, G., et al., *A Significant Population of Red, Near-Infrared-selected High-Redshift Galaxies*. 2003, ApJ, 587, L79.

[83] Fritz, J., Franceschini, A., Hatziminaoglou, E., *Revisiting the infrared spectra of active galactic nuclei with a new torus emission model*. 2006, MNRAS, 366, 767.

[84] Gao, D., Zhang, Y., Zhao, Y., *The Application of kd-tree in Astronomy*. In R. W. Argyle, P. S. Bunclark, & J. R. Lewis, ed., *Astronomical Data Analysis Software and Systems XVII*, vol. 394 of *Astronomical Society of the Pacific Conference Series*, (pp. 525–+) (2008).

[85] Gao, L., Springel, V., White, S. D. M., *The age dependence of halo clustering*. 2005, MNRAS, 363, L66.

[86] Gaztanaga, E., *High-Order Galaxy Correlation Functions in the APM Galaxy Survey*. 1994, MNRAS, 268, 913.

[87] Gebhardt, K., Bender, R., Bower, G., et al., *A Relationship between Nuclear Black Hole Mass and Galaxy Velocity Dispersion*. 2000, ApJ, 539, L13.

[88] Genzel, R., Lutz, D., Sturm, E., et al., *What Powers Ultraluminous IRAS Galaxies?* 1998, ApJ, 498, 579.

[89] Giavalisco, M., *Lyman-Break Galaxies*. 2002, ARA&A, 40, 579.

[90] Gilli, R., Daddi, E., Chary, R., et al., *The spatial clustering of mid-IR selected star forming galaxies at $z \sim$ in the GOODS fields*. 2007, A&A, 475, 83.

[91] Gilli, R., Daddi, E., Zamorani, G., et al., *The spatial clustering of X-ray selected AGN and galaxies in the Chandra Deep Field South and North*. 2005, A&A, 430, 811.

[92] Gilli, R., Zamorani, G., Miyaji, T., et al., *The spatial clustering of X-ray selected AGN in the XMM-COSMOS field*. 2009, A&A, 494, 33.

[93] Gott, J. R., III, Turner, E. L., Aarseth, S. J., *N-body simulations of galaxy clustering. III - The covariance function*. 1979, ApJ, 234, 13.

[94] Granato, G. L., De Zotti, G., Silva, L., Bressan, A., Danese, L., *A Physical Model for the Coevolution of QSOs and Their Spheroidal Hosts*. 2004, ApJ, 600, 580.

[95] Grazian, A., Fontana, A., de Santis, C., et al., *The GOODS-MUSIC sample: a multicolour catalog of near-IR selected galaxies in the GOODS-South field*. 2006, A&A, 449, 951.

[96] Guo, Q., White, S., Li, C., Boylan-Kolchin, M., *How do galaxies populate dark matter haloes?* 2010, MNRAS, 404, 1111.

[97] Hamana, T., Ouchi, M., Shimasaku, K., Kayo, I., Suto, Y., *Properties of host haloes of Lyman-break galaxies and Lyman $\alpha$ emitters from their number densities and angular clustering*. 2004, MNRAS, 347, 813.

[98] Hamilton, A. J. S., *Toward Better Ways to Measure the Galaxy Correlation Function*. 1993, ApJ, 417, 19.

[99] Hauser, M. G., Arendt, R. G., Kelsall, T., et al., *The COBE Diffuse Infrared Background Experiment Search for the Cosmic Infrared Background. I. Limits and Detections*. 1998, ApJ, 508, 25.

[100] Hauser, M. G., Dwek, E., *The Cosmic Infrared Background: Measurements and Implications*. 2001, ARA&A, 39, 249.

[101] Heymans, C., Bell, E. F., Rix, H., et al., *A weak lensing estimate from GEMS of the virial to stellar mass ratio in massive galaxies to $z \sim 0.8$*. 2006, MNRAS, 371, L60.

[102] Hickox, R. C., Jones, C., Forman, W. R., et al., *Host Galaxies, Clustering, Eddington Ratios, and Evolution of Radio, X-Ray, and Infrared-Selected AGNs*. 2009, ApJ, 696, 891.

[103] Hogg, D. W., *Distance measures in cosmology*. 1999, ArXiv Astrophysics e-prints.

[104] Hopkins, A. M., *On the Evolution of Star-forming Galaxies*. 2004, ApJ, 615, 209.

[105] Hopkins, P. F., Hernquist, L., Cox, T. J., Di Matteo, T., Robertson, B., Springel, V., *A Unified, Merger-driven Model of the Origin of Starbursts, Quasars, the Cosmic X-Ray Background, Supermassive Black Holes, and Galaxy Spheroids*. 2006, ApJS, 163, 1.

[106] Hopkins, P. F., Hernquist, L., Cox, T. J., Kereš, D., *A Cosmological Framework for the Co-Evolution of Quasars, Supermassive Black Holes, and Elliptical Galaxies. I. Galaxy Mergers and Quasar Activity*. 2008, ApJS, 175, 356.

[107] Hu, E. M., Cowie, L. L., McMahon, R. G., *The Density of Ly alpha Emitters at Very High Redshift*. 1998, ApJ, 502, L99+.

[108] Hu, W., Kravtsov, A. V., *Sample Variance Considerations for Cluster Surveys*. 2003, ApJ, 584, 702.

[109] Huang, J., Faber, S. M., Daddi, E., et al., *Infrared Spectrograph Spectroscopy and Multi-Wavelength Study of Luminous Star-Forming Galaxies at $z \simeq 1.9$*. 2009, ApJ, 700, 183.

[110] Jenkins, A., Frenk, C. S., White, S. D. M., Colberg, J. M., Cole, S., Evrard, A. E., Couchman, H. M. P., Yoshida, N., *The mass function of dark matter haloes*. 2001, MNRAS, 321, 372.

[111] Kaiser, N., *On the spatial correlations of Abell clusters*. 1984, ApJ, 284, L9.

[112] Kaiser, N., *Clustering in real space and in redshift space*. 1987, MNRAS, 227, 1.

[113] Kang, X., Jing, Y. P., Mo, H. J., Börner, G., *Semianalytical Model of Galaxy Formation with High-Resolution N-Body Simulations*. 2005, ApJ, 631, 21.

[114] Kauffmann, G., Colberg, J. M., Diaferio, A., White, S. D. M., *Clustering of galaxies in a hierarchical universe - I. Methods and results at $z = 0$*. 1999, MNRAS, 303, 188.

[115] Kenter, A., Murray, S. S., Forman, W. R., et al., *XBootes: An X-Ray Survey of the NDWFS Bootes Field. II. The X-Ray Source Catalog*. 2005, ApJS, 161, 9.

[116] Kenter, A. T., Murray, S. S., *A New Technique for Determining the Number of X-Ray Sources per Flux Density Interval*. 2003, ApJ, 584, 1016.

[117] Kerscher, M., *Statistical Analysis of Large-Scale Structure in the Universe*. In K. R. Mecke & D. Stoyan, ed., *Statistical Physics and Spatial Statistics. The Art of Analyzing and Modeling Spatial Structures and Pattern Formation*, vol. 554 of *Lecture Notes in Physics, Berlin Springer Verlag*, (pp. 37–+) (2000).

[118] Kerscher, M., Szapudi, I., Szalay, A. S., *A Comparison of Estimators for the Two-Point Correlation Function*. 2000, ApJ, 535, L13.

[119] Klypin, A., *Numerical Simulations in Cosmology I: Methods*. 2000, ArXiv Astrophysics e-prints.

[120] Klypin, A., Holtzman, J., *Particle-Mesh code for cosmological simulations*. 1997, ArXiv Astrophysics e-prints.

[121] Klypin, A., Trujillo-Gomez, S., Primack, J., *Halos and galaxies in the standard cosmological model: results from the Bolshoi simulation*. 2010, ArXiv e-prints.

[122] Knebe, A., *How to Simulate the Universe in a Computer*. 2005, Publ. of Astron. Soc. of Australia, 22, 184.

[123] Komatsu, E., Dunkley, J., Nolta, M. R., et al., *Five-Year Wilkinson Microwave Anisotropy Probe Observations: Cosmological Interpretation*. 2009, ApJS, 180, 330.

[124] Komatsu, E., Smith, K. M., Dunkley, J., et al., *Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation*. 2010, ArXiv e-prints.

[125] Koulouridis, E., Plionis, M., *Luminous X-ray Active Galactic Nuclei in Clusters of Galaxies*. 2010, ApJ, 714, L181.

[126] Kravtsov, A. V. 1999. *High-resolution simulations of structure formation in the universe*. Ph.D. thesis, NEW MEXICO STATE UNIVERSITY.

[127] Kravtsov, A. V., Berlind, A. A., Wechsler, R. H., Klypin, A. A., Gottlöber, S., Allgood, B., Primack, J. R., *The Dark Side of the Halo Occupation Distribution*. 2004, ApJ, 609, 35.

[128] Kravtsov, A. V., Klypin, A. A., *The Origin and Evolution of Halo Bias in Linear and Nonlinear Regimes*. 1999, ApJ, 520, 437.

[129] Kravtsov, A. V., Klypin, A. A., Khokhlov, A. M., *Adaptive Refinement Tree: A New High-Resolution N-Body Code for Cosmological Simulations*. 1997, ApJS, 111, 73.

[130] Kuhlen, M., Diemand, J., Madau, P., Zemp, M., *The Via Lactea INCITE simulation: galactic dark matter substructure at high resolution*. 2008, Journal of Physics Conference Series, 125, 012008.

[131] Lacey, C. G., Baugh, C. M., Frenk, C. S., Benson, A. J., Orsi, A., Silva, L., Granato, G. L., Bressan, A., *Predictions for Herschel from Λ-cold dark matter: unveiling the cosmic star formation history*. 2010, MNRAS, 405, 2.

[132] Lacy, M., Storrie-Lombardi, L. J., Sajina, A., et al., *Obscured and Unobscured Active Galactic Nuclei in the Spitzer Space Telescope First Look Survey*. 2004, ApJS, 154, 166.

[133] Lagache, G., Puget, J., Dole, H., *Dusty Infrared Galaxies: Sources of the Cosmic Infrared Background*. 2005, ARA&A, 43, 727.

[134] Landy, S. D., Szalay, A. S., *Bias and variance of angular correlation functions*. 1993, ApJ, 412, 64.

[135] Le Floc'h, E., Papovich, C., Dole, H., et al., *Infrared Luminosity Functions from the Chandra Deep Field-South: The Spitzer View on the History of Dusty Star Formation at* $0 <\sim z <\sim 1$. 2005, ApJ, 632, 169.

[136] Lee, K., Giavalisco, M., Gnedin, O. Y., Somerville, R. S., Ferguson, H. C., Dickinson, M., Ouchi, M., *The Large-Scale and Small-Scale Clustering of Lyman Break Galaxies at* $3.5 \leq z \leq 5.5$ *from the GOODS Survey*. 2006, ApJ, 642, 63.

[137] Limber, D. N., *The Analysis of Counts of the Extragalactic Nebulae in Terms of a Fluctuating Density Field*. 1953, ApJ, 117, 134.

[138] Lin, Y., Mohr, J. J., Stanford, S. A., *K-Band Properties of Galaxy Clusters and Groups: Luminosity Function, Radial Distribution, and Halo Occupation Number*. 2004, ApJ, 610, 745.

[139] Loh, Y., Rich, R. M., Heinis, S., et al., *The UV-optical colour dependence of galaxy clustering in the local universe*. 2010, MNRAS, 407, 55.

[140] Longair, M. S., *Galaxy Formation* (Springer, 2008).

[141] Lonsdale, C. J., Polletta, M., Farrah, D., et al., *Distant ULIRGs in the SWIRE Survey*. In R.-R. Chary, H. I. Teplitz, & K. Sheth, ed., *Infrared Diagnostics of Galaxy Evolution*, vol. 381 of *Astronomical Society of the Pacific Conference Series*, (pp. 461–+) (2008).

[142] Lonsdale, C. J., Polletta, M. d. C., Omont, A., et al., *MAMBO 1.2 mm Observations of Luminous Starbursts at $z \sim 2$ in the SWIRE Fields*. 2009, ApJ, 692, 422.

[143] Lonsdale, C. J., Smith, H. E., Rowan-Robinson, M., et al., *SWIRE: The SIRTF Wide-Area Infrared Extragalactic Survey*. 2003, PASP, 115, 897.

[144] Ma, C., Fry, J. N., *Deriving the Nonlinear Cosmological Power Spectrum and Bispectrum from Analytic Dark Matter Halo Profiles and Mass Functions*. 2000, ApJ, 543, 503.

[145] Macciò, A. V., Dutton, A. A., van den Bosch, F. C., *Concentration, spin and shape of dark matter haloes as a function of the cosmological model: WMAP1, WMAP3 and WMAP5 results*. 2008, MNRAS, 391, 1940.

[146] Madau, P., *Cosmic star formation history*. In S. S. Holt & L. G. Mundy, ed., *American Institute of Physics Conference Series*, vol. 393 of *American Institute of Physics Conference Series*, (pp. 481–490) (1997).

[147] Madau, P., Ferguson, H. C., Dickinson, M. E., Giavalisco, M., Steidel, C. C., Fruchter, A., *High-redshift galaxies in the Hubble Deep Field: colour selection and star formation history to $z \sim 4$*. 1996, MNRAS, 283, 1388.

[148] Maddox, S. J., Dunne, L., Rigby, E., et al., *Herschel-ATLAS: The angular correlation function of submillimetre galaxies at high and low redshift*. 2010, A&A, 518, L11+.

[149] Madgwick, D. S., Hawkins, E., Lahav, O., et al., *The 2dF Galaxy Redshift Survey: galaxy clustering per spectral type*. 2003, MNRAS, 344, 847.

[150] Magliocchetti, M., Cirasuolo, M., McLure, R. J., et al., *On the evolution of clustering of 24-$\mu$m-selected galaxies*. 2008, MNRAS, 383, 1131.

[151] Magliocchetti, M., Porciani, C., *The halo distribution of 2dF galaxies*. 2003, MNRAS, 346, 186.

[152] Magliocchetti, M., Silva, L., Lapi, A., de Zotti, G., Granato, G. L., Fadda, D., Danese, L., *A highly obscured and strongly clustered galaxy population discovered with the Spitzer Space Telescope*. 2007, MNRAS, 375, 1121.

[153] Magorrian, J., Tremaine, S., Richstone, D., et al., *The Demography of Massive Dark Objects in Galaxy Centers*. 1998, AJ, 115, 2285.

[154] Makovoz, D., Marleau, F. R., *Point-Source Extraction with MOPEX*. 2005, PASP, 117, 1113.

[155] Mandelbaum, R., Seljak, U., Kauffmann, G., Hirata, C. M., Brinkmann, J., *Galaxy halo masses and satellite fractions from galaxy-galaxy lensing in the Sloan Digital Sky Survey: stellar mass, luminosity, morphology and environment dependencies*. 2006, MNRAS, 368, 715.

[156] Marín, F. A., Wechsler, R. H., Frieman, J. A., Nichol, R. C., *Modeling the Galaxy Three-Point Correlation Function*. 2008, ApJ, 672, 849.

[157] Martínez, V. J., *The Large-Scale Structure in the Universe: From Power Laws to Acoustic Peaks*. In V. J. Martínez, E. Saar, E. Martínez-González, & M.-J. Pons-Bordería, ed., *Data Analysis in Cosmology*, vol. 665 of *Lecture Notes in Physics, Berlin Springer Verlag*, (pp. 269–289) (2009).

[158] Martínez, V. J., Saar, E., *Statistics of the Galaxy Distribution* (Chapman & Hall/CRC, 2002).

[159] Martini, P., Kelson, D. D., Kim, E., Mulchaey, J. S., Athey, A. A., *Spectroscopic Confirmation of a Large Population of Active Galactic Nuclei in Clusters of Galaxies*. 2006, ApJ, 644, 116.

[160] Martini, P., Weinberg, D. H., *Quasar Clustering and the Lifetime of Quasars*. 2001, ApJ, 547, 12.

[161] McCracken, H. J., Peacock, J. A., Guzzo, L., et al., *The Angular Correlations of Galaxies in the COSMOS Field*. 2007, ApJS, 172, 314.

[162] Melott, A. L., *Two-dimensional simulation of the gravitational superclustering of collisionless particles*. 1983, MNRAS, 202, 595.

[163] Meneux, B., Guzzo, L., de la Torre, S., et al., *The zCOSMOS survey. The dependence of clustering on luminosity and stellar mass at* $z = 0.2 - 1$. 2009, A&A, 505, 463.

[164] Milliard, B., Heinis, S., Blaizot, J., et al., *Clustering Properties of Rest-Frame UV-selected Galaxies. I. the Correlation Length Derived from GALEX Data in the Local Universe*. 2007, ApJS, 173, 494.

[165] Mo, H. J., White, S. D. M., *An analytic model for the spatial clustering of dark matter haloes*. 1996, MNRAS, 282, 347.

[166] Molnar, S. M., Hughes, J. P., Donahue, M., Joy, M., *Chandra Observations of Unresolved X-Ray Sources around Two Clusters of Galaxies*. 2002, ApJ, 573, L91.

[167] Moore, A. W., Connolly, A. J., Genovese, C., et al., *Fast Algorithms and Efficient Statistics: N-Point Correlation Functions*. In A. J. Banday, S. Zaroubi, & M. Bartelmann, ed., *Mining the Sky*, (pp. 71–+) (2001).

[168] More, S., van den Bosch, F. C., Cacciato, M., Mo, H. J., Yang, X., Li, R., *Satellite kinematics - II. The halo mass-luminosity relation of central galaxies in SDSS*. 2009, MNRAS, 392, 801.

[169] More, S., van den Bosch, F. C., Cacciato, M., Skibba, R., Mo, H. J., Yang, X., *Satellite kinematics - III. Halo masses of central galaxies in SDSS*. 2011, MNRAS, 410, 210.

[170] Moster, B. P., Somerville, R. S., Maulbetsch, C., van den Bosch, F. C., Macciò, A. V., Naab, T., Oser, L., *Constraints on the Relationship between Stellar Mass and Halo Mass at Low and High Redshift*. 2010, ApJ, 710, 903.

[171] Murray, S. S., Kenter, A., Forman, W. R., et al., *XBootes: An X-Ray Survey of the NDWFS Bootes Field. I. Overview and Initial Results*. 2005, ApJS, 161, 1.

[172] Nagai, D., Kravtsov, A. V., *The Radial Distribution of Galaxies in* $\Lambda$ *Cold Dark Matter Clusters*. 2005, ApJ, 618, 557.

[173] Narayanan, D., Dey, A., Hayward, C. C., et al., *A physical model for z ˜ 2 dust-obscured galaxies*. 2010, MNRAS, 407, 1701.

[174] Navarro, J. F., Frenk, C. S., White, S. D. M., *The Structure of Cold Dark Matter Halos*. 1996, ApJ, 462, 563.

[175] Norberg, P., Baugh, C. M., Gaztañaga, E., Croton, D. J., *Statistical analysis of galaxy surveys - I. Robust error estimation for two-point clustering statistics*. 2009, MNRAS, 396, 19.

[176] Oegerle, W. R., Hill, J. M., *Dynamics of cD Clusters of Galaxies. IV. Conclusion of a Survey of 25 Abell Clusters*. 2001, AJ, 122, 2858.

[177] Ouchi, M., Shimasaku, K., Akiyama, M., et al., *The Subaru/XMM-Newton Deep Survey (SXDS). IV. Evolution of Ly*$\alpha$ *Emitters from* $z = 3.1$ *to* 5.7 *in the* $1\mathrm{deg}^2$ *Field: Luminosity Functions and AGN*. 2008, ApJS, 176, 301.

[178] Overzier, R. A., Röttgering, H. J. A., Rengelink, R. B., Wilman, R. J., *The spatial clustering of radio sources in NVSS and FIRST; implications for galaxy clustering evolution*. 2003, A&A, 405, 53.

[179] Padilla, N. D., Baugh, C. M., *The power spectrum of galaxy clustering in the APM Survey*. 2003, MNRAS, 343, 796.

[180] Padmanabhan, N., White, M., Norberg, P., Porciani, C., *The real-space clustering of luminous red galaxies around* $z < 0.6$ *quasars in the Sloan Digital Sky Survey*. 2009, MNRAS, 397, 1862.

[181] Peacock, J. A., *Large-scale surveys and cosmic structure*. 2003, ArXiv Astrophysics e-prints.

[182] Peacock, J. A., Cole, S., Norberg, P., et al., *A measurement of the cosmological mass density from clustering in the 2dF Galaxy Redshift Survey*. 2001, Nature, 410, 169.

[183] Peacock, J. A., Smith, R. E., *Halo occupation numbers and galaxy bias*. 2000, MNRAS, 318, 1144.

[184] Peebles, P., *The Large Scale Structure of the Universe* (Princeton University Press, 1980).

[185] Peebles, P. J. E., *Structure of the Coma Cluster of Galaxies*. 1970, AJ, 75, 13.

[186] Peebles, P. J. E., *Statistical Analysis of Catalogs of Extragalactic Objects. I. Theory*. 1973, ApJ, 185, 413.

[187] Peebles, P. J. E., *The Nature of the Distribution of Galaxies*. 1974, A&A, 32, 197.

[188] Phleps, S., Peacock, J. A., Meisenheimer, K., Wolf, C., *Galaxy clustering from COMBO-17: the halo occupation distribution at* $\langle z \rangle = 0.6$. 2006, A&A, 457, 145.

[189] Plionis, M., Rovilos, M., Basilakos, S., Georgantopoulos, I., Bauer, F., *Luminosity-dependent X-Ray Active Galactic Nucleus Clustering?* 2008, ApJ, 674, L5.

[190] Poggianti, B. M., Bressan, A., Franceschini, A., *Star Formation and Selective Dust Extinction in Luminous Starburst Galaxies*. 2001, ApJ, 550, 195.

[191] Polletta, M., Tajer, M., Maraschi, L., et al., *Spectral Energy Distributions of Hard X-Ray Selected Active Galactic Nuclei in the XMM-Newton Medium Deep Survey*. 2007, ApJ, 663, 81.

[192] Pollo, A., Meneux, B., Guzzo, L., et al., *The VIMOS VLT deep survey. Computing the two point correlation statistics and associated uncertainties*. 2005, A&A, 439, 887.

[193] Pons-Bordería, M., Martínez, V. J., Stoyan, D., Stoyan, H., Saar, E., *Comparing Estimators of the Galaxy Correlation Function*. 1999, ApJ, 523, 480.

[194] Pope, A., Chary, R., Alexander, D. M., et al., *Mid-Infrared Spectral Diagnosis of Submillimeter Galaxies*. 2008, ApJ, 675, 1171.

[195] Porciani, C., Catelan, P., Lacey, C., *How Dark Matter Halos Cluster in Lagrangian Space*. 1999, ApJ, 513, L99.

[196] Porciani, C., Giavalisco, M., *The Clustering Properties of Lyman Break Galaxies at Redshift* $z \sim 3$. 2002, ApJ, 565, 24.

[197] Porciani, C., Magliocchetti, M., Norberg, P., *Cosmic evolution of quasar clustering: implications for the host haloes*. 2004, MNRAS, 355, 1010.

[198] Prunet, S., Pichon, C., Aubert, D., Pogosyan, D., Teyssier, R., Gottloeber, S., *Initial Conditions For Large Cosmological Simulations*. 2008, ApJS, 178, 179.

[199] Puget, J., Abergel, A., Bernard, J., Boulanger, F., Burton, W. B., Desert, F., Hartmann, D., *Tentative detection of a cosmic far-infrared background with COBE.* 1996, A&A, 308, L5+.

[200] Quadri, R., van Dokkum, P., Gawiser, E., et al., *Clustering of K-selected Galaxies at* $2 < z < 3.5$*: Evidence for a Color-Density Relation*. 2007, ApJ, 654, 138.

[201] Quadri, R. F., Williams, R. J., Lee, K., Franx, M., van Dokkum, P., Brammer, G. B., *A Confirmation of the Strong Clustering of Distant Red Galaxies at* $2 < z < 3$. 2008, ApJ, 685, L1.

[202] Revnivtsev, M., Vikhlinin, A., Sazonov, S., *Resolving the Galactic ridge X-ray background*. 2007, A&A, 473, 857.

[203] Ricciardelli, E., Franceschini, A., *GECO: Galaxy Evolution COde - A new semi-analytical model of galaxy formation*. 2010, A&A, 518, A14+.

[204] Rieke, G. H., Young, E. T., Engelbracht, C. W., et al., *The Multiband Imaging Photometer for Spitzer (MIPS)*. 2004, ApJS, 154, 25.

[205] Rigby, J. R., Rieke, G. H., Maiolino, R., et al., *24 Micron Properties of X-Ray-selected Active Galactic Nuclei*. 2004, ApJS, 154, 160.

[206] Roche, N., Shanks, T., Metcalfe, N., Fong, R., *The Angular Correlation Function of Galaxies with B=25MAG*. 1993, MNRAS, 263, 360.

[207] Rodighiero, G., Vaccari, M., Franceschini, A., et al., *Mid- and far-infrared luminosity functions and galaxy evolution from multiwavelength Spitzer observations up to $z \sim 2.5$*. 2010, A&A, 515, A8+.

[208] Ross, A. J., Percival, W. J., Brunner, R. J., *Evolution of the clustering of photometrically selected SDSS galaxies*. 2010, MNRAS, 407, 420.

[209] Ross, N. P., da Ângela, J., Shanks, T., et al., *The 2dF-SDSS LRG and QSO Survey: the LRG 2-point correlation function and redshift-space distortions*. 2007, MNRAS, 381, 573.

[210] Ross, N. P., Shen, Y., Strauss, M. A., et al., *Clustering of Low-redshift ($z < 2.2$) Quasars from the Sloan Digital Sky Survey*. 2009, ApJ, 697, 1634.

[211] Rowan-Robinson, M., Babbedge, T., Oliver, S., et al., *Photometric redshifts in the SWIRE Survey*. 2008, MNRAS, 386, 697.

[212] Rudnick, G., Labbé, I., Förster Schreiber, N. M., et al., *Measuring the Average Evolution of Luminous Galaxies at $z < 3$: The Rest-Frame Optical Luminosity Density, Spectral Energy Distribution, and Stellar Mass Density*. 2006, ApJ, 650, 624.

[213] Sacchi, N., La Franca, F., Feruglio, C., et al., *Spectroscopic Identifications of Spitzer Sources in the SWIRE/XMM-Newton/ELAIS-S1 Field: A Large Fraction of Active Galactic Nucleus with High F(24 $\mu$m)/F(R) Ratio*. 2009, ApJ, 703, 1778.

[214] Sanders, D. B., Mirabel, I. F., *Luminous Infrared Galaxies*. 1996, ARA&A, 34, 749.

[215] Scannapieco, E., Barkana, R., *An Analytical Approach to Inhomogeneous Structure Formation*. 2002, ApJ, 571, 585.

[216] Scoccimarro, R., Sheth, R. K., Hui, L., Jain, B., *How Many Galaxies Fit in a Halo? Constraints on Galaxy Formation Efficiency from Spatial Clustering*. 2001, ApJ, 546, 20.

[217] Scott, S. E., Dunlop, J. S., Serjeant, S., *A combined re-analysis of existing blank-field SCUBA surveys: comparative 850-$\mu$m source lists, combined number counts, and evidence for strong clustering of the bright submillimetre galaxy population on arcminute scales*. 2006, MNRAS, 370, 1057.

[218] Seljak, U., *Analytic model for galaxy and dark matter clustering*. 2000, MNRAS, 318, 203.

[219] Seljak, U., *Redshift-space bias and $\beta$ from the halo model*. 2001, MNRAS, 325, 1359.

[220] Seljak, U., Sugiyama, N., White, M., Zaldarriaga, M., *Comparison of cosmological Boltzmann codes: Are we ready for high precision cosmology?* 2003, Phys. Rev. D, 68, 083507.

[221] Shankar, F., Weinberg, D. H., Miralda-Escudé, J., *Self-Consistent Models of the AGN and Black Hole Populations: Duty Cycles, Accretion Rates, and the Mean Radiative Efficiency*. 2009, ApJ, 690, 20.

[222] Shen, Y., Strauss, M. A., Ross, N. P., et al., *Quasar Clustering from SDSS DR5: Dependences on Physical Properties*. 2009, ApJ, 697, 1656.

[223] Sheth, R. K., Tormen, G., *Large-scale bias and the peak background split*. 1999, MNRAS, 308, 119.

[224] Shupe, D. L., Rowan-Robinson, M., Lonsdale, C. J., et al., *Galaxy Counts at 24 $\mu$m in the SWIRE Fields*. 2008, AJ, 135, 1050.

[225] Sijacki, D., Springel, V., Di Matteo, T., Hernquist, L., *A unified model for AGN feedback in cosmological simulations of structure formation*. 2007, MNRAS, 380, 877.

[226] Silva, L., Granato, G. L., Bressan, A., Danese, L., *Modeling the Effects of Dust on Galactic Spectral Energy Distributions from the Ultraviolet to the Millimeter Band*. 1998, ApJ, 509, 103.

[227] Silverman, J. D., Green, P. J., Barkhouse, W. A., et al., *Comoving Space Density of X-Ray-selected Active Galactic Nuclei*. 2005, ApJ, 624, 630.

[228] Silverman, J. D., Green, P. J., Barkhouse, W. A., et al., *The Luminosity Function of X-Ray-selected Active Galactic Nuclei: Evolution of Supermassive Black Holes at High Redshift*. 2008, ApJ, 679, 118.

[229] Silverman, J. D., Lamareille, F., Maier, C., et al., *Ongoing and Co-Evolving Star Formation in zCOSMOS Galaxies Hosting Active Galactic Nuclei*. 2009, ApJ, 696, 396.

[230] Skrutskie, M. F., Cutri, R. M., Stiening, R., et al., *The Two Micron All Sky Survey (2MASS)*. 2006, AJ, 131, 1163.

[231] Springel, V., Hernquist, L., *Cosmological smoothed particle hydrodynamics simulations: a hybrid multiphase model for star formation*. 2003, MNRAS, 339, 289.

[232] Springel, V., White, S. D. M., Jenkins, A., et al., *Simulations of the formation, evolution and clustering of galaxies and quasars*. 2005, Nature, 435, 629.

[233] Starikova, S., Cool, R., Eisenstein, D., et al., *Constraining halo occupation properties of X-ray AGNs using clustering of Chandra sources in the Bootes survey region*. 2010, ArXiv e-prints.

[234] Stern, D., Eisenhardt, P., Gorjian, V., et al., *Mid-Infrared Selection of Active Galaxies*. 2005, ApJ, 631, 163.

[235] Surace, J. A., Shupe, D. L., Fang, F., Evans, T., Alexov, A., Frayer, D., Lonsdale, C. J., SWIRE Team, *Data Processing and Validation of the SWIRE Survey*. In *Bulletin of the American Astronomical Society*, vol. 37 of *Bulletin of the American Astronomical Society*, (pp. 1246–+) (2005).

[236] Swanson, M. E. C., Tegmark, M., Blanton, M., Zehavi, I., *SDSS galaxy clustering: luminosity and colour dependence and stochasticity*. 2008, MNRAS, 385, 1635.

[237] Tasitsiomi, A., Kravtsov, A. V., Wechsler, R. H., Primack, J. R., *Modeling Galaxy-Mass Correlations in Dissipationless Simulations*. 2004, ApJ, 614, 533.

[238] The SDSS Collaboration, Zehavi, I., Zheng, Z., et al., *Galaxy Clustering in the Completed SDSS Redshift Survey: The Dependence on Color and Luminosity*. 2010, ArXiv e-prints.

[239] Tinker, J. L., *Redshift-space distortions with the halo occupation distribution - II. Analytic model*. 2007, MNRAS, 374, 477.

[240] Tinker, J. L., Wechsler, R. H., Zheng, Z., *Interpreting the Clustering of Distant Red Galaxies*. 2010, ApJ, 709, 67.

[241] Tinker, J. L., Weinberg, D. H., Zheng, Z., *Redshift-space distortions with the halo occupation distribution - I. Numerical simulations*. 2006, MNRAS, 368, 85.

[242] Tinker, J. L., Weinberg, D. H., Zheng, Z., Zehavi, I., *On the Mass-to-Light Ratio of Large-Scale Structure*. 2005, ApJ, 631, 41.

[243] Totsuji, H., Kihara, T., *The Correlation Function for the Distribution of Galaxies*. 1969, PASJ, 21, 221.

[244] Treister, E., Urry, C. M., Van Duyne, J., et al., *Spitzer Number Counts of Active Galactic Nuclei in the GOODS Fields*. 2006, ApJ, 640, 603.

[245] Ueda, Y., Akiyama, M., Ohta, K., Miyaji, T., *Cosmological Evolution of the Hard X-Ray Active Galactic Nucleus Luminosity Function and the Origin of the Hard X-Ray Background*. 2003, ApJ, 598, 886.

[246] Vale, A., Ostriker, J. P., *Linking halo mass to galaxy luminosity*. 2004, MNRAS, 353, 189.

[247] Vale, A., Ostriker, J. P., *The non-parametric model for linking galaxy luminosity with halo/subhalo mass*. 2006, MNRAS, 371, 1173.

[248] van Dokkum, P. G., Förster Schreiber, N. M., Franx, M., et al., *Spectroscopic Confirmation of a Substantial Population of Luminous Red Galaxies at Redshifts $z \geq 2$*. 2003, ApJ, 587, L83.

[249] van Dokkum, P. G., Quadri, R., Marchesini, D., et al., *The Space Density and Colors of Massive Galaxies at $2 < z < 3$: The Predominance of Distant Red Galaxies*. 2006, ApJ, 638, L59.

[250] Viero, M. P., Ade, P. A. R., Bock, J. J., et al., *BLAST: Correlations in the Cosmic Far-Infrared Background at 250, 350, and 500 $\mu$m Reveal Clustering of Star-forming Galaxies*. 2009, ApJ, 707, 1766.

[251] Vikhlinin, A., Forman, W., *Detection of the Angular Correlation of Faint X-Ray Sources*. 1995, ApJ, 455, L109.

[252] Vikhlinin, A., Kravtsov, A. V., Burenin, R. A., et al., *Chandra Cluster Cosmology Project III: Cosmological Parameter Constraints*. 2009, ApJ, 692, 1060.

[253] Vikhlinin, A., McNamara, B. R., Forman, W., Jones, C., Quintana, H., Hornstrup, A., *A Catalog of 200 Galaxy Clusters Serendipitously Detected in the ROSAT PSPC Pointed Observations*. 1998, ApJ, 502, 558.

[254] Waddington, I., Oliver, S. J., Babbedge, T. S. R., et al., *Clustering of galaxies at 3.6 $\mu$m in the Spitzer Wide-area Infrared Extragalactic legacy survey*. 2007, MNRAS, 381, 1437.

[255] Wall, J. V., Jenkins, C. R., *Practical Statistics for Astronomers* (CAMBRIDGE University Press, 2003).

[256] Weiß, A., Kovács, A., Coppin, K., et al., *The Large Apex Bolometer Camera Survey of the Extended Chandra Deep Field South*. 2009, ApJ, 707, 1201.

[257] White, M., Hernquist, L., Springel, V., *The Halo Model and Numerical Simulations*. 2001, ApJ, 550, L129.

[258] White, S. D. M., Rees, M. J., *Core condensation in heavy halos - A two-stage theory for galaxy formation and clustering*. 1978, MNRAS, 183, 341.

[259] Yan, L., Chary, R., Armus, L., et al., *Spitzer Detection of Polycyclic Aromatic Hydrocarbon and Silicate Dust Features in the Mid-Infrared Spectra of $z \sim 2$ Ultraluminous Infrared Galaxies*. 2005, ApJ, 628, 604.

[260] Yang, X., Mo, H. J., van den Bosch, F. C., *Constraining galaxy formation and cosmology with the conditional luminosity function of galaxies*. 2003, MNRAS, 339, 1057.

[261] Zehavi, I., Blanton, M. R., Frieman, J. A., et al., *Galaxy Clustering in Early Sloan Digital Sky Survey Redshift Data*. 2002, ApJ, 571, 172.

[262] Zehavi, I., Eisenstein, D. J., Nichol, R. C., et al., *The Intermediate-Scale Clustering of Luminous Red Galaxies*. 2005, ApJ, 621, 22.

[263] Zehavi, I., Weinberg, D. H., Zheng, Z., et al., *On Departures from a Power Law in the Galaxy Correlation Function*. 2004, ApJ, 608, 16.

[264] Zel'Dovich, Y. B., *Fragmentation of a homogeneous medium under the action of gravitation*. 1970, Astrophysics, 6, 164.

[265] Zheng, Z., Berlind, A. A., Weinberg, D. H., et al., *Theoretical Models of the Halo Occupation Distribution: Separating Central and Satellite Galaxies*. 2005, ApJ, 633, 791.

[266] Zheng, Z., Zehavi, I., Eisenstein, D. J., Weinberg, D. H., Jing, Y. P., *Halo Occupation Distribution Modeling of Clustering of Luminous Red Galaxies*. 2009, ApJ, 707, 554.

[267] Zwicky, F., Herzog, E., Wild, P., *Catalogue of galaxies and of clusters of galaxies, Vol. I* (1961).