Astrophysics and Space Science Library 435

Mauro D'Onofrio
Roberto Rampazzo
Simone Zaggia *Editors*

From the Realm of the Nebulae to Populations of Galaxies

Dialogues on a Century of Research

ASSL

Springer

What is the expected metallicity of first galaxies?

The metallicity of the first galaxies can either be very low or high (solar and oversolar) depending on the morphological type of the primordial galaxies. A spheroid (elliptical galaxy or bulge of spiral) can attain a high metallicity in a very short time, if the star formation rate is strong enough. In fact, in such a case many SNe core-collapse producing metals can enrich the ISM in a few hundreds thousands years. This is the case of quasar (QSO) hosts, which are large ellipticals at high redshift. The metallicity inferred from the broad emission lines of QSOs indicate solar and oversolar abundances in these objects (see Maiolino et al. 2006). The QSO hosts confirm that different galaxies had very different star formation histories. On the other hand, if the star formation rate is low then the metal content can be comparable with the metallicity of the stars in the Galactic halo and be several orders of magnitude lower than the solar metallicity. However, a more important parameter than the absolute metallicity is represented by the abundance ratios, such as, for example, the already mentioned [α/Fe] ratios. These ratios, coupled with the metallicity, measured by [Fe/H], can give us an idea of the type of primordial object we are observing, since the [α/Fe] vs. [Fe/H] relations are different for different galaxies, as indicated in Fig. 4.20 of Chap. 4. On the base of the diagram of Fig. 4.20 we can therefore infer the nature of high redshift objects: in fact, in Fig. 4.20 overplotted on the theoretical curves are also data for Magellanic Irregulars and Damped Lyman-α systems (DLAs; objects observed at high redshift whose nature is unknown), and the comparison between models and data indicates that the unknown DLAs are very likely to be irregular galaxies. In particular, irregular galaxies show low [α/Fe] ratios at low [Fe/H] because of the slow star formation they suffered.

In the current standard model of cosmology baryons represent only $\sim$4 % of the energy density of the Universe, while up to $\sim$22 % is enclosed in the DM component. The dark energy seems to drive the whole energy density with up to $\sim$72 %.

Discovered during the seventy through its gravitational action, DM has up to now escaped any direct detection, posing the problem of their true existence. We enter with the following interviews in this almost unexplored territory of the modern astrophysics. DM dominated galaxies should manifest the presence of this component in their kinematics. We therefore start our approach to this subject with few questions about galaxy dynamics and kinematics.

7.7 Galaxy Dynamics: Dark Matter or MOND?

Questions for Luca Ciotti:
understanding galaxy dynamics has been one of the big efforts of the past century. What have been in your opinion the most significant progresses achieved in this research area? What are the limits of present day models and

luca.ciotti@unibo.it

simulations? In which way dynamical studies might provide constraints to the galaxy formation and evolution theories?

The proposed questions are very relevant in general, and even more from my particular point of view. However, I think any sufficiently simple answer is necessarily incomplete and questionable, considering the enormous range of applications of Stellar Dynamics when applied to the study of formation, evolution and structure of galaxies (the so-called Galaxy Dynamics). In any case, I will attempt to answer in a sufficiently objective and concise way.

The first point to be stressed is that the pillars upon which Stellar Dynamics is built are among the most studied branches of classical physics, i.e., Gravitation, Dynamics, Statistical Mechanics, Fluido-dynamics, with major contributions from giants like Newton, Maxwell, Boltzmann, Poincarè (just to recall some names, but the list would include almost all of the prominent physicists and mathematicians one can think of). From this point of view it is hardly believable that new relevant and original contributions can still be obtained in the field. However, this is not true, as *observations* of increasing quality and resolution provide a fresh flow of unexpected and very interesting problems, spurring whole new fields of application of the "old" theories. Before discussing some of them (not because I attribute to them any special significance, but just because it occurred to me to work on them), I would like to stress my view about two general points, that quite often arise in discussions about Galaxy Dynamics.

The first: Galaxy Dynamics as a tool. Over the years, talking to colleagues not directly involved in Galaxy Dynamics research, I found a common attitude to think about this subject as a "boring-but-sometimes-useful" collection of techniques to be used for the interpretation of the observations. I can easily accept that for someone Galaxy Dynamics is a boring discipline, but I strongly disagree with the view that it is just a "tool". Quite the opposite, the applications of Galaxy Dynamics to the interpretation of observational data are just one of the *two* relevant aspects of the discipline, the other being the *understanding* of the dynamics of astronomical systems made by a very large number of components and ruled by gravitation. From this last point of view, Galaxy Dynamics is connected with the study of non-linear and collective phenomena (like Plasma Physics), a field very far from being completely explored and understood.

The second: Galaxy Dynamics and computers. It is a quite widespread idea that "after-all-we-now-have-computers so let's use them to integrate equations, look at the results, and spare the time (and the painful work) needed for their understanding". The weakness of this position should be clear to everyone: the job of astrophysicists is to understand the physics behind the observed phenomena, not to contemplate the results of numerical simulations and be satisfied if they agree with observations. Computers and numerical simulations should be seen as *invaluable* tools for better understanding physics, not just to reproduce observations (or, even worst, to produce results that are studied instead of the physical Universe).

About the specific questions proposed. As fundamental progresses made in Galaxy Dynamics in the past century, I think a general consensus can be obtained for any list containing entries such as violent relaxation, construction of collisionless equilibria and study of their stability (with particularly important results for elliptical galaxies), integrability, relaxation times and dynamical friction, spiral density wave theory, dynamical evolution of weakly collisional systems (e.g., Globular Clusters, galactic nuclei with SMBHs). On a more observational side, I would list among the major accomplishments the discovery of the strict connection between structure and dynamics of the different galaxy morphological types, the evidences of Dark Matter on galactic scales, the existence of global Scaling Laws relating in a surprising way structural, dynamical, and stellar populations properties of ETGs, the presence of SMBHs at the centers of stellar spheroids. Among the relevant references, covering a large part of the subjects listed above, I would mention the books of Spitzer (1988); Binney and Tremaine (2008); Merritt (2013), and Bertin (2014).

On a more personal side, I will focus on some aspects of the multifaced problem posed by the existence of Scaling Laws of ETGs, their interpretation, and the consequences for theories of galaxy formation. My interest in Stellar Dynamics began during the Ph.D. thesis work, when I was involved in the study of gas flows in ETGs under the supervision of Alvio Renzini. I was assigned to work on the dynamical modelization of the galaxies used for the hydrodynamical simulations. I was very interested by the mathematical aspects of Astrophysics, and I found my problem beautiful. In particular, in parallel with the Thesis work, I wrote a short paper on the dynamics of Sérsic models for elliptical galaxies (Ciotti 1991), that I sent for comments to Giuseppe Bertin (then at Scuola Normale Superiore in Pisa). This was the starting point not only of a lifelong friendship with Giuseppe, but also of a continuous source of learning and inspiration for my studies of Galaxy Dynamics. The two other persons that were fundamental for my scientific growth in this field were Jerry Ostriker and James Binney, and more recently Renzo Sancisi, Tjieerd van Albada, and Tim de Zeeuw.

During the galaxy modelization work for the thesis, Alvio maintained close contacts with George Djorgovski (Caltech), who gave us some advice on how to place our galaxy models on the Fundamental Plane of elliptical galaxies (hereafter FP) that was recently discovered (Djorgovski and Davis 1987; Dressler et al. 1987). The new galaxy models represented a big step forward in the field of cooling flow studies. From this work Alvio started to think on the implications of the FP itself for the understanding of formation and evolution of elliptical galaxies. We wrote a paper (Renzini and Ciotti 1993), and I also started to meditate about the meaning of the FP. I tried to work out the consequences by myself, and I soon realized that—with few notable exceptions—the astronomical community was quite confused about the meaning of the FP. For example there were papers dealing with the problem of the "deviation of the FP from Virial Theorem" (the so-called FP tilt), and even attempting to "reconcile" the two relations. This was clearly a physical nonsense, as the Virial Theorem by itself does not necessarily imply *any* Scaling Law for virialized systems (what about spiral galaxies then? are not they equilibrium—i.e. virialized—systems? but spiral galaxies certainly do not obey a FP!). I like to illustrate the case stressing the conceptual analogy between the

FP and the Main Sequence in the HR diagram: pretending that the FP is nothing else than the Virial Theorem in disguise is the same logical error as to pretend that the Main Sequence is nothing else than the locus of hydrostatic equilibria of gaseous spheres. Virialized galaxies, as gaseous spheres in equilibrium, could be placed *everywhere* in their respective parameter spaces: the existence of the FP and of the Main Sequence are actually telling us something completely different (and more important!) than an equilibrium condition. In fact, as the Main Sequence reveals something fundamental about energy generation in the stellar interiors, the FP contains important information about the structure, dynamics and formation of ETGs (Nipoti et al. 2002). The FP tilt does not measure any departure of ETGs from the Virial Theorem, instead it shows that for some reason ETGs "deviate" in a well determined way from structural and dynamical *homology*, i.e., they are not exact scaled-up/down versions of the same prototypical object, but the "deviations" actually depend in a very well defined way on their total luminosity and size, with a surprisingly small scatter (see, e.g., Ciotti 2009 and references therein). This systematic deviation it is called *weak homology* (e.g., see Bertin et al. 2002). It is never stressed enough that the very existence of the FP give another important clue about ETGs: in fact, in order to have the observed weak homology, not only structure and dynamics of the stellar component must be related, but also the *stellar mass-to-light ratio* of the stellar population and the amount and distribution of dark matter must be. I consider this empirical fact as one of the most strong arguments against merging as the main channel of formation of ETGs.

The last point opens another important consideration about the risk of a partial view of the problems addressed by Galaxy Dynamics. In fact, quite often dynamicists (and even more people working on numerical simulations) tend to think to galaxies as systems made by "gray dots", just ruled by gravity. Of course, this is quite true, but not completely, as stars are not grey dots: they have ages, luminosities, metallicities, and it may well happen that purely gravitational considerations can be ruled out by considering additional, non-gravitational aspects such as chemical evolution. I consider this one of the important lessons learned working with Alvio. One of such problems is that about the importance of *dry* and *wet* merging in the formation and evolution of ETGs. In fact, from the purely morphological point of view, merging appears to be—on average—able to reproduce some gross properties of ETGs. However, it is an elementary back-to-the-envelope exercise to show that in parabolic dry (i.e., in absence of gas dissipation) mergings with no substantial mass ejection the velocity dispersion of the final product cannot be larger than the larger one among those of the progenitors. This is in contrast with the Faber-Jackson law, therefore this simplistic picture is ruled out (e.g., Ciotti and van Albada 2001). Of course, gas dissipation (wet merging) can help to increase the velocity dispersion and reduce the size of the resulting system, however the more dissipation we need, the younger will be the stellar population of the new galaxy, and this must be reconciled with the empirical fact that more massive ETGs are older and more metallic that smaller systems. These kinds of problems are tentatively addressed by scenarios like the "anti-hierarchical formation of ETGs"; personally, I think that merging (in the usual sense attributed to the word) is much less important than commonly assumed in the formation of ETGs (Ciotti et al. 2007).

In my opinion, one of the future developments of Galaxy Dynamics (in addition to the study of very basic questions related to important mathematical and physical aspects of the theory), should be sought in the field where Stellar Dynamics meets Fluido-dynamics, i.e., in the problems concerned with the coexistence of pure gravitational (e.g., *N*-body) phenomena with hydrodynamical phenomena, such those involving the gas component of spiral galaxies and ETGs. Examples are those of the dynamics of spiral arms and of the extraplanar gas in spiral galaxies, and of the combined use of gas stellar dynamics to measure the amount and distribution of Dark Matter in ETGs. In fact, in spite of all the work done, and the many claims, the latter problem is far from being settled. I find quite curious to have papers dealing with Dark Matter in ETGs where not only the amount of Dark Matter is measured, but even its radial profile is discussed, sometimes at the level of minor details, against the expectations of "universality" as obtained by impressive cosmological numerical simulations of increasing resolution. While it is reasonable to contrast observations with the results of simulations, I would like to stress that the problem of the determination of the Dark Matter halos density profile is not even solved for *disk galaxies* (are the best solutions maximum disk? minimum disk?), in spite of their "simple" orbital structure (in first approximation almost circular velocity for the gas and stars in the disk). Therefore, my impression is that the current claims about ETGs, with their more complicate orbital structure (e.g., mass-anisotropy degeneracies, particularly affecting modeling techniques based on the Jeans equations), and with gaseous halos whose hydrostatic equilibrium (a common assumption made in mass studies based on X-ray observations) is far from being assured (see e.g., Pellegrini and Ciotti 2006), should be considered *at best* indications of the true radial distribution of Dark Matter halos. A significant progress in this field (where also gravitational lensing gives a fundamental contribution) will be reached when our global understanding of the physics of the mutual interaction between stellar, dark matter and gaseous components (e.g., Barnabé et al. 2006) will improve beyond the level where the different components are just "added" as to produce our preferred "pet galaxy models".

With the next two interviews we enter in the still ample debate about the existence and nature of the DM component.

Questions for Jaan Einasto:
the nature of Dark Matter (DM) is still a mystery after more than 20 years. You have deeply investigated in your recent book "Dark Matter and Cosmic Web Story" the presence and the role of DM in galaxies. What are the most significant observational proofs of its existence in galaxies?

Actually there are two problems with DM—its existence and nature.

The existence of DM is a problem already for 80 years when Fritz Zwicky discovered the mass paradox in the Coma cluster of galaxies (Zwicky 1933). A similar paradox was found in galaxies—the rotation curve of galaxies remained flat at large galactocentric distances in contrast to the expected Keplerian decrease in the SMoC, because it may be added to the SMoPP naturally in accounting for the masses of the active neutrinos.[12]

Thus, a very conservative cosmological model appears to be emerging (Kroupa 2014a): it is based on GR, has inflation and dark energy, and the exotic cold or warm dark matter particles of the SMoC are naturally avoided by vacuum effects yielding SID with sterile neutrinos entering as a hot dark matter component. This model has been introduced by Garry Angus, who constraints the mass of the sterile neutrino to be about 11 eV (Angus 2009; Angus and Diaferio 2011; Angus et al. 2010, 2013). As of a few months ago full-scale purely hydrodynamic simulations of structure formation in MOND have become possible with the code *Phantom of Ramses* (PoR) developed by Fabian Lüghausen in collaboration with Benoit Famaey and myself as a patch to Romain Teyssier's RAMSES (Lüghausen et al. 2015).

Given that the SMoC is ruled out, new models are being studied which naturally account for the observed gravitational and dynamical properties of galaxies without dark matter. However, given the pressure in the community against non-SMoC research, progress is very slow and even haltering. The widely-held interpretation that the population of galaxies we observed today is a result of mergers is certainly wrong; mergers play a minor role only (Sect. 7.7). *Galaxies, as they are observed, cannot be reproduced in the SMoC.* Observational cosmology is providing important new constraints, but applications of the redshift—distance and redshift—age relations from the SMoC to interpret these data is almost certainly wrong in that *the true nature of the physical systems at high redshift is likely to be distorted when interpreted within the framework of the SMoC.*

Among the alternative ideas on the DM problem, the Modified Newtonian Dynamics (MOND) has reached a considerable number of supporters. The next interview to Luca Ciotti will present his point of view on this theory.

Questions for Luca Ciotti:
up to now, several alternative theoretical explanation to the DM have been attempted. MOND is a theory that does not require exotic particles but a modification of the Newtonian law at specific conditions. Could you briefly explain MOND and address the question of the advantages and disadvantages of this approach? Are you aware of other theoretical approach to the DM problem that could be promising?

As universally known, Modified Newtonian Dynamics (nowadays there are several variants, but here I will generically refer to all of them as MOND) is a theory developed to avoid the introduction of Dark Matter in Astronomy, obtained by suitably modifying the gravity law (e.g. Bekenstein and Milgrom 1984; Milgrom 1983). In practice, in the most known formulation, the Poisson equation $\nabla^2\phi = 4\pi G\rho$ relating the gravitational potential ϕ to the given mass distribution ρ, is

[12]Note that the often invoked limits on neutrino masses using standard-cosmological arguments become invalid by discarding the SMoC.

replaced by the field equation

$$\nabla \cdot [\mu(||\nabla\phi||/a_0)\,\nabla\phi] = 4\pi G\rho, \tag{7.4}$$

where $\mu \sim 1$ for $||\nabla\phi|| >> a_0$ (the Newtonian limit) and $\mu \sim ||\nabla\phi||/a_0$ for $||\nabla\phi|| << a_0$ (the so-called "deep-MOND" regime). The characteristic acceleration scale—empirically determined—is $a_0 \simeq 1.2\,10^{-8}$ cm s^{-2}.

My interest in MOND started when Bob Sanders (Groningen University) visited Bologna for a few days following an invitation of Renzo Sancisi. He gave a talk on the general principles behind MOND, and I recall well the first impression I get. I was struck by the fact that MOND was clearly based on a quite beautiful and deep mathematical structure, and had the enormous merit (admittedly, a rarer and rarer property in the steadily increasing number of astrophysical papers!) to be right or wrong. During the talk, I realized that the two-body relaxation time in MOND (and dynamical friction as well) should be very different than in Newtonian gravity (with or without dark matter), essentially because the field strength in the weak regime goes like r^{-1} instead of r^{-2}, so that we do not have the Coulomb logarithm and the MOND effects should be much stronger than in classical gravitation. In turns, this should imply that stellar systems in MOND should be much more collisional than in Newtonian gravity, with all the internal dynamical evolution much faster than routinely assumed, with important observational consequences. But I also soon realized that the standard approach to the computation of relaxation times could not be applied to MOND, as the theory is intrinsecally non-linear, and you cannot just "sum" the effects of gravitational interactions as in the classical Chandrasekhar approach. I studied the problem in detail, and finally—after a visit to James Binney in Oxford where we solved some non-trivial issue due to non-linearity, we wrote a paper (Ciotti and Binney 2004), showing that indeed MOND is more collisional than Newtonian gravity, with interesting (but not so dramatic as I initially expected) consequences.

After this work, I concluded that in order to have a better understanding of MOND in realistic cases, one should move away from the study of spherically symmetric systems where—for technical reasons that I cannot touch here—the computations are quite simple, but also quite special. By extending a technique developed with Bertin to produce non-spherical exact density-potential pairs in Newtonian gravity, it was possible to introduce a mapping method that allows the construction of ellipsoidal-like, analytical density-potential pairs in MOND, so that their orbital properties can be studied with standard numerical codes. In fact, one of the main aspects of MOND, its non-linearity, is reflected by the fact that the associated field equation in the deep-MOND regime reduces to the so-called "p-laplacian", a non linear partial differential operator. After this analytical work it was clear that the next step would be the development of a numerical MOND solver, and with the fundamental contribution of dr. Pasquale Londrillo (INAF—OABo), in a joint work with dr. Carlo Nipoti (Bologna University) we finally wrote NMODY, a numerical code able to run N-body simulations in MOND. This was the first MOND code extensively used in literature, and we made a free distribution of it.

With this code we studied for the first time several interesting problems in MOND, such as violent relaxation radial orbit instability, dynamical friction, with the hope to find some important discrepancy with respect to the corresponding "Equivalent Newtonian Systems".[13] Remarkably, we found instead that the differences were not so big, even though well detectable and going in the sense indicated by theoretical arguments (such as the rigorous time-independence of the virial function W in time-dependent, deep-MOND systems, and the long-lasting virial oscillations of MOND systems approaching equilibrium, e.g., see Nipoti et al. 2008). On a more theoretical side, in collaboration with Tim de Zeeuw (Leiden University) and H.S. Zhao I also studied the properties of Stäckel systems in MOND. Also in this case we find, quite surprisingly, that the so-called "Kuzmin theorem" (i.e., the fact that for a Stäckel potential the assignment of density along the short axis fixes the density everywhere) is not a peculiar property of Newtonian gravity but also holds for MOND or, more technically, also of the *p-laplacian* applied to potentials separable in ellipsoidal confocal coordinates (Ciotti et al. 2012).

The general lesson I learned from the studies mentioned above (and others on disk galaxies not discussed here) is that MOND is on one side conceptually very different from Newtonian gravity plus dark matter, *however*, when applied to systems similar to the observed ones, the MOND predictions are in general surprisingly similar (see, e.g., Sanders and McGaugh 2002). Of course, there are differences in the predictions of the two theories, but none of them appears fully inconsistent with observations or, in any case, the discrepancies can be debated (even in the most problematic cases for MOND, such as the "Bullett Cluster"). This is a most remarkable because MOND was initially introduced just to explain the flat rotation curves of spiral galaxies, so that all other predictions for different systems (ETGs, cluster of galaxies, etc.) can be considered impressive accomplishments (see Sanders book). The successes are even more surprising considering that in MOND you just have a *single* free parameter (the acceleration scale a_0), and everything is determined by the baryonic distributions.

So, what is my view of MOND? I think that MOND is mathematically deep and elegant, and it has some interesting aspect worth to be discussed against Classical Gravitation and General Relativity (plus dark matter, plus the even more mysterious "Dark Energy"), especially because it is falsifiable. On the physical side, I'm personally very conservative, and I believe that MOND is *not* the correct description of gravity on large scale objects such galaxies, which I believe is strictly Newtonian and obeys the *superposition principle*, with all the related consequences (for example, the fact that the internal dynamics of a system is independent of the acceleration in case of free fall). May be MOND is just telling us that Dark matter and baryons in real systems "talked" each other at the epoch of the formation

[13]Obviously, proper comparison of MOND with standard gravity must be done with Newtonian systems plus an amount of DM as would be predicted by the application of MOND to the purely baryonic component. These systems are what we called "Equivalent Newtonian Systems".

of cosmic structures much more than what we usually think, and this mutual relationship is qualitatively captured by the existence of a"universal" constant, a_0.

I think the most interesting future development of MOND on the theory side will be the possibility to run simulations of structure formation in the cosmological setting. Of course, this will depend on the numerical ability to treat baryonic physics, as in MOND we do not have (in principle) Dark Matter!

7.8 To Summarize

The aim of this chapter was that of showing that our idea of galaxies is still changing today, after one century since their discovery. The first beautiful image of galaxies as unperturbed Island Universes floating in a uniformly expanding space is already behind our shoulders. The time of the primitive scenario provided by the monolithic collapse of a protogalaxy regulated by relative simple physical mechanisms that originated the main galaxy components is definitively over. It is necessary to accept the idea that galaxies are the results of a complex evolution and that we have only a partial knowledge of the mechanisms that operated the transformation of galaxies across the Hubble time. Galaxies, like men, live in a complex and evolving society that sometime change forever their properties.

Galaxies are also the site of the most energetic phenomena observed in the Universe. We have discussed some of them in this chapter trying to understand their relation with the host galaxy. How much these phenomena influence the whole galaxy evolution? What are the consequences of these tremendous energy deliver for the galaxy itself and for its environment?

Like for the other chapters we try here to summarize the most important facts emerged in our interviews.

- *After the pioneering observations of Zwicky, Holmberg, Arp, Vorontsov-Velyaminov and many others, astronomers began conscious that gravitational interactions might change the morphology of galaxies and produce a number of new structures that sometimes are stable on medium-large dynamical timescale. The Toomre' simulations opened the way towards the modern large N-body simulations that are now able to explain several of the observed morphologies of interacting systems: rings, induced spirals, bridges, tails, shells, loops, etc.*
- *Tidal dwarf galaxies are an example of objects created by gravitational interactions. These galaxies might reach stable configurations and are likely not embedded in a dark halo.*
- *Environmental interactions can be as important as close encounters in affecting the evolution of galaxies. The gas stripping and the consequent starvation are only one of the mechanisms with which the environment can operate. Other effects connected to the environment are the compression of removed gas clouds igniting a star formation in the ISM, or the stressing and wave generation in the gas not removed from the disks. Galaxy harassment is also very important. It is the result*

central supermassive black hole? Each of them suggests a totally different physical mechanism for quenching, and yet they are all tightly correlated with each other, so it gets very hard to observationally identify the culprit! Yet, it is even possible that mass and environment quenching may be two different manifestation of a same, underlying physical process (Knobel et al. 2015). I hope we can solve this problem within a few years.

What is the quenching time scale? Is it the same for all galaxies?

Many groups are trying to measure the quenching timescale, which may be different for mass and environment quenching, but there is no answer yet to this question. If quenching is due to gas ejection from the galaxy, e.g., as resulting from some sort of AGN feedback, then the quenching timescale may be quite short, of the order of the dynamical time, or $\sim 10^8$ years. If instead quenching results from cutting off gas supply from the environment, then the quenching timescale could be quite long, of the order of the gas depletion timescale, i.e., $M_{\rm gas}$/SFR, or some $\sim 10^9$ years, with $M_{\rm gas}$ being the mass of gas inside the galaxy at the beginning of the quenching process. We can gather an estimate of the quenching timescale from the number of galaxies caught in such transition, but, as I mentioned earlier, the green valley can be also populated, at all redshifts, by occasional visitors and intruders.

Which is the relation between the quenching of SF and the morphological transformation?

Empirically, we see that most quenched galaxies show an early-type morphology (i.e., they are elliptical or S0 galaxies) and most early-type galaxies are quenched. But why quenching is accompanied by morphological transformation we don't not know for sure, yet. This is indeed another open question. Integral field spectroscopy of local early-type galaxies has demonstrated that the vast majority of them ($\sim 86\,\%$) are fast rotators, whereas only the residual minority are slow rotators (Emsellem et al. 2011). There is general consensus that the slow rotators are the result of merging, which then can be considered responsible for the morphological transformation for only a minority of galaxies. The fast rotators instead are likely to be the result of the evolution of the disk, via some kind of disk instability (Dekel and Burkert 2014).

We examine now another aspect of galaxy evolution, that related to the so-called feedback. With this term astronomers summarize all the processes occurring in galaxies that are energetic enough to significantly affect their evolution.

8.6 The Role of Feedback

Questions for Luca Ciotti:
the AGN feedback is claimed to be an important physical mechanism in galaxy evolution. Could you explain why and trace a short history of this idea ? Which observations prove that such feedback indeed occurred? How is galaxy

luca.ciotti@unibo.it

evolution affected by the feedback? Is this mechanism active in all galaxies or only in some morphological types?

The topic of AGN feedback in galaxies (in particular, in early-type galaxies, hereafter ETGs) has been, and it is right now, a relevant aspect of my research activity. As a consequence, in the following the presentation may reflect quite a personal point of view, which is not necessarily shared by all other researchers in the field. Overall, looking back over the past 25 years, since when I started to work on the subject (together with J.P. Ostriker during the sojourn at Princeton University as a PhD student), I can say that the attitude of a large part of the scientific community has been quite peculiar, ranging from initial positions like "there is no AGN feedback in ETGs", to the present "AGN feedback is the main actor in shaping the formation and evolution of ETGs, and to produce their properties as we observe them today". Well, I quite disagree with both views. I will present some arguments supporting the claim that AGN feedback *was* known to be important even 25 years ago, a necessary conclusion of elementary empirical arguments. At the same time, I claim that the main effects of AGN feedback are *not on* the galaxies, hosting at their centers the Supermassive Black Holes (hereafter SMBHs), but are essentially of more local nature, mainly affecting the growth of the SMBHs and extending at most to the galactic centers, in a $\simeq$ kpc-size region around the SMBH, and of course regulating star formation in the centers of ETGs.

In 1989–1992 I was working on my PhD thesis in an excellent research group, lead by Alvio Renzini. Annibale D'Ercole (then Astronomer at the Bologna Astronomical Observatory) and Silvia Pellegrini (also PhD student) were also in the group. Alvio was very enthusiastic about a new idea he had for the explanation of some puzzling observational property of the X-ray emission of the hot atmospheres surrounding ETGs. In particular, it was clear that, in absence of some form of heating, the gaseous halos of ellipticals, *produced by the mass ejected by the stellar mass losses of the aging stellar population* at the rate $\dot{M}_*$, i.e. the "secular evolution" of these systems, would necessarily lead to massive *cooling flows* in all elliptical galaxies, with the consequent prediction of systematically high and *unobserved* X-ray luminosities (L_X). In fact, from the well established and tested theory of stellar evolution, it is known that the mass losses of an old, passively evolving stellar population of present-day total luminosity L_B (in blue solar units) can be well approximated as

$$\dot{M}_* \simeq 1.5\,10^{-11} L_B t_{15}^{-1.35} \quad M_\odot \mathrm{yr}^{-1}, \tag{8.2}$$

where t_{15} is time in 15 Gyr units.

The cooling flow model (Cowie and Binney 1977; Fabian and Nulsen 1977), with the prediction of high values of L_X, was the paradigm at the epoch, but it is important to recall some important facts. That the stellar evolution would inject over cosmological times an *enormous* amount of mass in the host galaxies (summing up to 20 to 30 % of the initial stellar mass M_* of the galaxy) was so obvious that in the '70s the very important model of Supernova driven galactic wind (Mathews and

Baker 1971) was proposed as the natural solution to the conundrum posed on one hand by the unquestioned prediction of stellar evolution about mass losses, and the apparent lack of detection of gas in ETGs on the other. The whole astronomical community was well aware that ETGs, at least from the point of view of the mass budget, are certainly not *dead and red* objects. In the '80s, the detection of X-ray emission around ETGs by *Einstein* (see, e.g., Fabbiano (2012); Mathews and Brighenti (2003)) finally showed that the mass *was* there, and the *cooling flow model* became the paradigm to study this kind of problems. However, it was soon realized that if the mass injected was cooling, the final state of such cooling gas should be found somewhere in the galaxy, in form of new stars, or dark objects, or free floating baryon condensations. In addition, it became also clear that the X-ray emission L_X of medium-to-low mass ellipticals was systematically *lower* than what expected by the *standard* cooling flow model, that instead worked better (although with significant dispersion—almost two dex—in the predicted values of L_X) for massive ellipticals. Remarkably, *all* the proposed solutions attempting to reconcile the pure cooling flow scenario with observations failed, for a combination of theoretical and empirical arguments. Renzini coagulated a research group, with complementary competences, to work on the problem. In particular, by building realistic galaxy models that at the epoch were state-of-the-art (e.g., laying on the Fundamental Plane), and using the most robust prescriptions of stellar evolution, we concluded that elliptical galaxies are—from the energetic point of view—very peculiar systems, i.e., the energy needed to steadily extract the injected gas from the galaxy gravitational potential, and the energy injected per unit time in the hot ISM by SNIa explosions and thermalization of stellar motions, are *almost the same*, so that the X-ray halos are in a metastable energetic configuration. Moreover, we also found that, due to the Faber-Jackson relation, the binding energy per unit mass of the ISM (roughly proportional to the stellar velocity dispersion of the host galaxy) in large ETGs is higher than in low mass systems, so that while the latter systems should be in a global galactic wind state (in practice, mass losses from the evolving stars are ejected from the galaxy being heated to a super-virial temperatures), massive ETGs should be in the cooling flow state, with the consequent high L_X. These energy-based estimates were nicely confirmed by our hydrodynamical simulations (Ciotti 1991) that however revealed a scenario more complicated than that depicted above (for example, the remarkable fact that the time evolution of the SNIa explosion rate is very similar to the time evolution of $\dot{M}_*$, a fact without obvious physical explanation). In summary, at that time, in addition to have learnt a lot of physics from Alvio and numerics from Annibale, I had clear in mind that (1) even in isolated ETGs (i.e., in absence of major/minor merging, cold flows, etc., objects that today would be called "red and dead"), there are internal, time-decreasing, significant sources of mass just provided by stellar evolution, and (2) while the cooling flow was not the state of the atmospheres of ETGs of low/medium mass, a large fraction of the massive ellipticals (say objects with a central velocity dispersion of the order of 250 km/s or more), should be in a cooling-flow like state (for a full account of the situation see, e.g., Pellegrini (2012), and references therein).

In particular, while the work in our group in Bologna was clearly a significant step forward in understanding the evolution of the gaseous component of "red and dead" galaxies, yet the fate of the $\simeq 1 M_{\odot}/yr$ produced internally and flowing towards the center in *massive* ETGs remained unsolved. It was exactly at this time that I started my sojourn in Princeton. After my arrival at the beginning of 1992 and a few weeks of "testing", Jerry decided that I would be assigned to study the problem of the fate of the cooling flows in big ellipticals. This was particularly timely, considering the important discovery that at the center of ETGs there are SMBHs with a mass of the order of $M_{\rm BH} \simeq 10^{-3} M_{*}$ (Magorrian et al. 1998), successively confirmed and reinforced by the discovery of the $M_{\rm BH} - \sigma$ relation (see, e.g., Ferrarese and Merritt (2000); Gebhardt et al. (2000); Yu and Tremaine (2002)). It is clear that in these systems AGN feedback is necessary, not as a consequence of complicated arguments, but just because of the extreme *smallness* of the mass of the central SMBHs. In fact, a rough calculation easily shows that the SMBH masses are approximately *two orders of magnitude* smaller than the gas made available by stellar evolution in isolated ETGs (and the argument is only reinforced in case of external accretion/merging). In practice, *AGN feedback is required by mass arguments, not by energetic arguments*. We started to work on the theory of AGN feedback, supported by numerical simulations of increasing quality (with improvements in the input physics still ongoing, thanks to the involvement over the years of several other researchers) to test observational predictions. In fact, for a mass accretion rate of $\dot{M}_{\rm BH}$, the emitted luminosity—for a given electromagnetic efficiency ϵ—is

$$L_{\rm BH} = \epsilon \dot{M}_{\rm BH} c^2 \simeq \epsilon (\dot{M}_{\rm BH}/M_{\odot} yr)\, 5.7\, 10^{46} \quad {\rm erg/s}, \qquad (8.3)$$

high enough to suppress the potential cooling flow and interrupt accretion (see also Binney and Tabor (1995)). The question we addressed in this first exploration of AGN feedback was why we do not observe quasars at the center of all massive ETGs as a consequence of the expected accretion. The answer was obtained and refined in a series of papers, based on numerical hydrodynamical simulations of gas flows in ETGs including radiative transport, with the spatial and temporal resolution needed to probe the resulting flows on cosmological times and on spatial scales ranging from galactic sizes down to the parsec scale near the central SMBH (well inside the Bondi radius, so that no "ad hoc" treatment for accretion, common in similar studies, was required). We showed that gas accretion on the central SMBH, due to the onset of a "cooling flow" phase, releases and transfers to the ISM enough energy to stop the cooling flow itself, and to evacuate the inner kpc-scale region around the SMBH. After a characteristic time, needed to replenish the central zone of the galaxy, and to increase the ISM to values large enough to start another "cooling catastrophe", the cycle repeats (for a full description of the simulations and the results, see (Ciotti 2009a,b; Ciotti and Ostriker 2012; Ostriker and Ciotti 2005)).

Quite surprisingly (for the current view), we found a strong and negative reaction to our proposal (with the exception of a few notable cases, such as Alvio Renzini and James Binney, one of the fathers of the cooling flow model then visiting

Princeton, where I met him for the first time) as in general the community was fiercely defending the cooling flow paradigm (already in crisis due to SNIa heating for low/medium mass galaxies, and now also questioned for the remaining galaxies). The reactions went so far as to claim that "ETGs were lacking signs of feedback", or proposing that the SMBHs were actually steady accreting in the "obscured modality" (i.e., without emission of significant radiation, with no feedback, and so in a sense still consistent with the cooling flow paradigm). But all these criticisms missed the point, i.e., that the low mass of central SMBHs *is* a clear observational signature of feedback, and that obscured accretion *cannot* be the solution, because the SMBH mass would grow to unobserved values (in fact, that obscured accretion cannot be used to reconcile the cooling flow model with the physics of SMBH accretion is also proved beyond discussion by the Soltan argument, coupled with the well known theoretical upper bounds on accretion efficiency of compact objects, see e.g., see e.g. Yu and Tremaine (2002)).

A few important aspects of AGN feedback should be considered. First, the time interval from the beginning of central accretion, to its shutdown due to AGN feedback, is found to be of the order of 10^7 yrs, in nice accordance with observational estimates of the "on" phase of quasars. Second, in the simulations these feedback events becomes more and more rare as the galaxy age increases (see Fig. 8.7), because the stellar mass losses need longer and longer time to produce the critical density required for a global ISM cooling event (see Eq. 8.2). Third, as the major feedback events in the life of a galaxy are just a few, it results that the *duty-cycle* of AGN activity (i.e., the time fraction so that the AGN luminosity is above some fraction of the Eddington luminosity) is much less than unity ($\simeq 10^{-2}$ or even less), thus explaining why we do not see quasars in galaxies in the local Universe, i.e., because the probability to catch a SMBH in the "on" phase is very small, and it decreases with increasing cosmic time.

As already stressed, an important aspect of the AGN feedback physics - not always appreciated—is that the main issue of the problem is not whether there is enough energy to stop a cooling flow (see Eq. 8.3), but *how much* of the energy emitted in a given accretion event can be transmitted to the ISM in the host galaxy. Theoretical estimates and physically based numerical simulations of AGN feedback show that in fact the fraction of energy transferred to the ISM (and so able to stop gas cooling) is *very small*. In other words, the energy emitted by the AGN in a given accretion event is very large (much bigger than the energy required to eject all the ISM from the galaxy in the intergalactic space), but the captured fraction (both as radiation and kinetic coupling with the conical nuclear wind launched by the AGN) is only able to momentarily stop the gas cooling.

This is a very interesting fact, as nowadays AGN feedback, after having been initially ignored or even discarded as an important aspect of the evolution of ETGs, is invoked as the final explanation of why ETGs are the systems with the characteristics we observe. For example, AGN feedback is considered the main actor in quenching star formation at the epoch of galaxy formation. My impression is that this is more an expectation than a proved statement. In fact, numerical simulations in spherical symmetry (when feedback effects are maximum for geometric reasons),

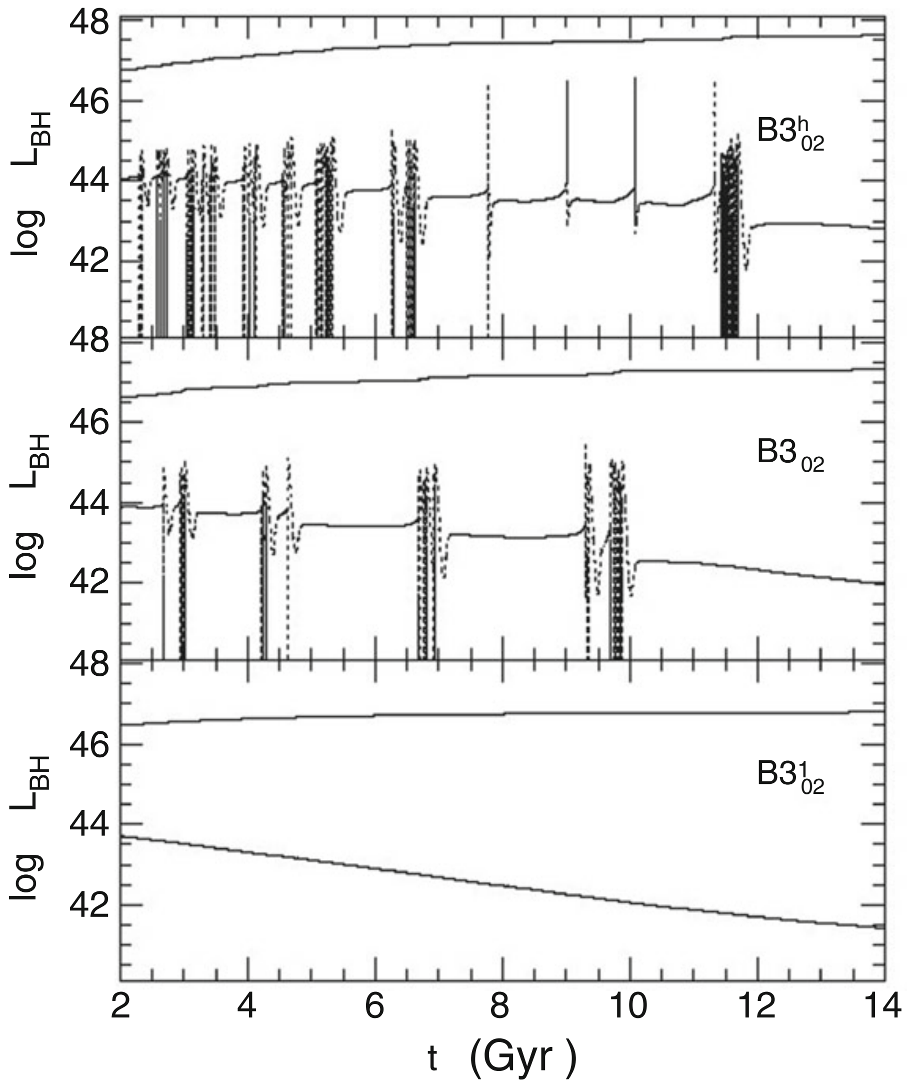

Fig. 8.7 *Dotted lines* are the optical SMBH luminosity corrected for absorption (i.e, as would be observed from infinity) for three galaxy models with central velocity dispersion of 280 km/s ($B3^h_{02}$), 260 km/s ($B3_{02}$), and 240 km/s ($B3^l_{02}$). The almost *horizontal solid line* represents the Eddington luminosity. Note how the less massive galaxy is in a state of SNIa driven permanent galactic wind, and the AGN accretion luminosity remains low (Adapted by permission of the AAS from Ciotti et al. 2010)

and with realistic coupling between radiation and matter (obtained by solving radiative transport equations) are systematically found unable to eject from massive galaxies the ISM produced by stellar evolution, even worse if we imagine the galaxy filled with all the gas needed for star formation, and a more realistic (and less efficient) non-spherical feedback geometry.

Another related interesting result that emerged from our work (Ciotti and Ostriker 2007), was the fact that actually AGN feedback can *induce* star formation, at the beginning of each major feedback event. In fact, each event (of a total duration of $\approx 10^7$ yrs), when observed at sufficiently high time resolution, is made of a series of sub-burst of increasing intensity (e.g., see the last burst in the top panel of Fig. 8.7), due to a complex hydrodynamical structure of the ISM in the $\simeq 300 - 500$ pc around the SMBH. In this region, the sequence of shock waves (direct and reflected) leads to the formation of a gaseous cold shell, with a few hundred parsecs radius, that in turns form stars at peak rates of $10^2 M_{\odot}/yr$ or more. The final sub burst in the series finally ends the sequence, and stops star formation: therefore, we found that AGN feedback is—at the same time—able to *induce* and *suppress* star formation. We also found that the new stars produced by the periodic central starbursts are distributed in the central regions of the models with a profile remarkably similar in shape and values to the observed stellar cusps in the central regions of ETGs (Graham et al. (2003), see also Ciotti (2009a,b), and references therein). It is interesting to speculate that the so-called "E+A galaxies" may be somewhat related to this recurrent activity.

In the spirit of this book, I conclude presenting a list of major results about AGN feedback that I think are quite robust, followed by a list of points that I feel should be the focus of future investigations, theoretical and observational.

(R1) AGN feedback in galaxies is required by simple mass arguments, not by energy arguments: the mass of SMBHs at the center of big ETGs is approximately two orders of magnitude smaller than the gas that would be accreted by a non-impeded cooling flow. Therefore, obscured /or radiatively inefficient accretion is not a solution to the problem of missing quasars in massive ETGs.

(R2) Sporadic quasar activity *is* present in ETGs, even in perfect isolation, due to the immense amount of material secularly injected in the galaxy by stellar evolution. Therefore, quasar statistics cannot be straightforwardly used as a measure of frequence of gas-rich ("wet") merging events, as it can be produced purely by secular internal evolution of "red and dead" galaxies.

(R3) AGN feedback is, *empirically*, fundamental to maintain the mass of SMBHs "small", however it is unable to fully evacuate the host galaxy by the mass injected by the aging stars. SNIa heating, being distributed over the galaxy body, and released at a continuous rate, is much more important. All the available indications from numerical simulations where the feedback is calculated from first principles seem to suggest that the effects at early times can be similarly small, in absence of some additional physical effects. Possibly, SNII are more important in terminating star formation at early times.

(R4) Stellar evolution has the nice property that the amount of material injected *scales linearly* with the stellar mass of the galaxy, so that the accretion of some fraction of this material on the central SMBH does not destroy (or even improves) a proportionality possibly established at the end of the period of galaxy formation.

(R5) The efficiency of AGN feedback and the rates of gas injection and cooling are essentially *unrelated* phenomena: a long-time balance between the two is impossible, so that steady-state configurations are practically impossible in massive ETGs. A possible exception is represented by low mass ellipticals, where SMBHs accretion proceeds at very low $L_{\rm BH}$, with Bondi-like accretion from hot and low-density atmosphere, as the galaxies are in SNIa assisted global winds.

Among the questions that I would like to see addressed (and solved!) in a near future:

(Q1) What is the role of angular momentum in the structure and evolution of gas flows in galaxies with some rotation? It is known that in these systems, in absence of additional heating phenomena, gas cooling would lead to the formation of massive, centrifugally-supported, kpc-size disks of cold gas, unable to reach the center. What happens of these disks? Are they consumed by star formation? Are they massive enough to become self-gravitating and unstable? If yes, will they develop non axisymmetric features, break angular momentum conservation and collapse toward the center fueling the SMBH? What kind of feedback the AGN will produce when fed by such disks?

(Q2) How can we describe in acceptable physical terms the "granularity" of the galaxy stellar distribution within the inner tens of parsec around the SMBH? Of course, a spatial and temporal smooth description of the stellar distribution and of the mass and energy injection becomes more and more unrealistic as the number of stars involved decreases.

(Q3) What is the relative role of radiative and kinetic energy in AGN feedback? What are the observational signatures of AGN induced and suppressed star formation (the so-called positive and negative feedback)? What is the relative importance for feedback of the starburst energy compared to the AGN energy?

(Q4) The contribution of AGN feedback to quench star formation at the epoch of galaxy assembly was really fundamental? Or it was just an additional contribution to SNII and SNIa activity?

Questions for Francesca Matteucci:
SNe have been indicated as possible sources of feedback mechanisms. Could you explain why? Which is the role of SNe in galaxy evolution? Which observations confirm these ideas?

As already mentioned, supernovae influence galaxy evolution through chemical enrichment and energy feedback, namely the energy that they can transfer into the ISM. The explosion energy of SNe is large, although in some cases most of it can be lost via cooling, and clearly contributes to increase the thermal energy of the ISM. Because of this, the interstellar gas can reach the escape velocity and escape from the potential well of the galaxy and in this case we speak of galactic wind, but the gas can also be temporarily removed and fall back again, in such a case we speak of galactic fountains (Bregman 1980; Spitoni et al. 2008). These fountains are likely to occur in spiral disks and are triggered by multiple explosions of massive stars. The evidence of galactic winds is given by the metals found in the ICM and IGM and they have also been observed in dwarf irregular galaxies. In particular, the observations of dwarf starburst galaxies indicate that these winds are linked to SN explosions. Martin (2002) reported `Chandra` observations of the dwarf starburst galaxy NGC 1569 in the Local Group showing the gas which is escaping from the

Astrophysics and Space Science Library 435

Mauro D'Onofrio · Roberto Rampazzo · Simone Zaggia *Editors*

From the Realm of the Nebulae to Populations of Galaxies

Dialogues on a Century of Research

In order to outline possible future directions in galaxy research, this book wants to be a short stopover, a moment of self-reflection of the past century of achievements in this area.

Since the pioneering years of galaxy research in the early 20th century, the research on galaxies has seen a relentless advance directly connected to the parallel exponential growth of new technologies.

Through a series of interviews with distinguished astronomers the editors provide a snapshot of the achievements obtained in understanding galaxies. While many initial questions about their nature have been addressed, many are still open and require new efforts to achieve a solution. The discussions may reveal paradigms worthwhile revisiting. With the help of some of those scientists who have contributed to it, the editors sketch the history of this scientific journey and ask them for inspirations for future directions of galaxy research.

Astronomy / Astrophysics

ISBN 978-3-319-31004-6

9 783319 310046

▸ springer.com