Organic Molecules in the Interstellar Medium

Jakob Held

Chalmers University of Technology
SE-41296 Göteborg, Sweden
(f96jahe@dd.chalmers.se)

*

Abstract

This paper presents a brief overview of the current knowledge about which organic molecules are present in the interstellar medium (ISM) and how these molecules are formed. The possibility of the formation and survival of amino acids in the ISM is also discussed. Amino acids are essential to life, as we know it, since they are the building blocks of proteins. Analysis of the Murchison meteorite shows that it contains an excess of left-handed amino acids. This excess could be a clue to why living things on earth almost exclusively use left-handed amino acids and would imply that life was given a head start by amino acids forming in the presolar nebula.

1 Introduction

The search for organic molecules in the interstellar medium has been given an increasing interest over the last years. Observational results are combined with laboratory experiments and theoretical models to enhance our understanding of how organic molecules evolve, from being single atoms in the interstellar medium (ISM) to perhaps amino acids incorporated into planetary systems.

2 The interstellar medium

The ISM is the birthsite of stars and planets and it is thought to account for 20-30% of the mass in our galaxy (Ehrenfreund & Charnley, 2000). It consists of single atoms, molecules, and dust grains between the stars. Most of the mass is distributed unevenly, concentrated in "clouds" (Irvine, 1998). These clouds are not uniform structures, their density is greatly varying and they evolve in time as old stars die and new stars are formed. The clouds are usually roughly classified as either diffuse or dense.

*Hot Topics in Astrophysics 2001/2002, Alessandro B. Romeo & Eva S. Karlsson (Eds.), Chalmers University of Technology and Göteborg University, 2002.

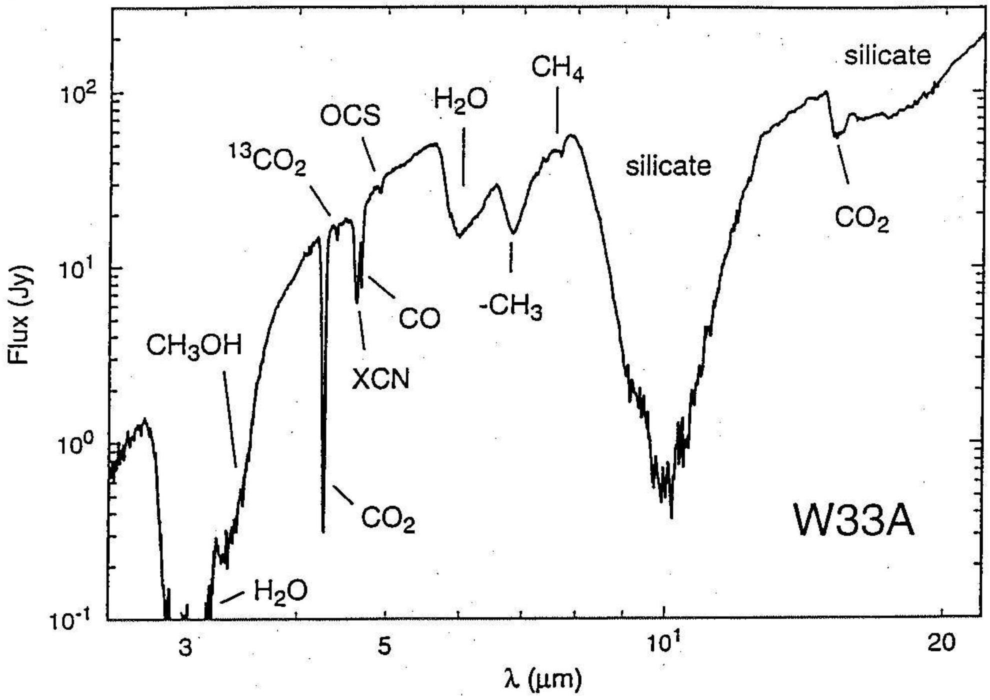


Figure 1: The ISO-SWS spectrum towards the protostar W33A. (From Ehrenfreund & Charnley, 2000.)

2.1 Diffuse Clouds

Low density (100-300 cm^{-1}) and an intense flux of ultraviolet photons characterize the diffuse clouds. The temperature is approximately 100K (Ehrenfreund & Charnley, 2000). This implies that the mean free paths are of the order 10^9 - 10^{15} cm, resulting in a time between collisions of 10^4 - 10^{10} s. This is much longer than the relaxation time for electronic and vibrational excitations so these do not influence the collisional formation processes (Snyder, 1997). The intense flux of UV photons destroy the complex molecules rapidly and only the simple molecules which are efficiently reformed can be unambiguously identified, through electronic absorptions of the continuum emission of radiation of background stars. The list of organic molecules detected in diffuse clouds includes HCO^+, CO, OH, C_2, HCN, HNC, CN, CS and H_2CO (Ehrenfreund & Charnley, 2000).

Molecules are thought to form primarily through simple exothermic gas-phase reactions, but the grains probably also contribute to the chemistry (Ehrenfreund & Charnley, 2000). Dust grains are formed in the envelopes of cool, evolved stars (Ehrenfreund & Charnley, 2000; Snyder, 1997) and are transported away from the star by stellar winds. They are thought to consist mainly of carbonaceous material and silicates. Supernovae shock waves help spread the grains over vast distances through the ISM where the grains are processed by grain-grain collisions, ultraviolet (UV) photons, cosmic rays, and grain growth by collisions with single molecules and atoms. The grains act as catalysts to help form more complex molecules (Ehrenfreund & Charnley, 2000; Irvine, 1998; Snyder, 1997). The identification of molecules incorporated in grains is difficult, since the spectroscopic methods (absorption spectra) give results based on functional groups of the grains rather than of single molecules (Irvine, 1998). Fig. 1 shows an absorption spectra of interstellar ices from the protostar W33A.

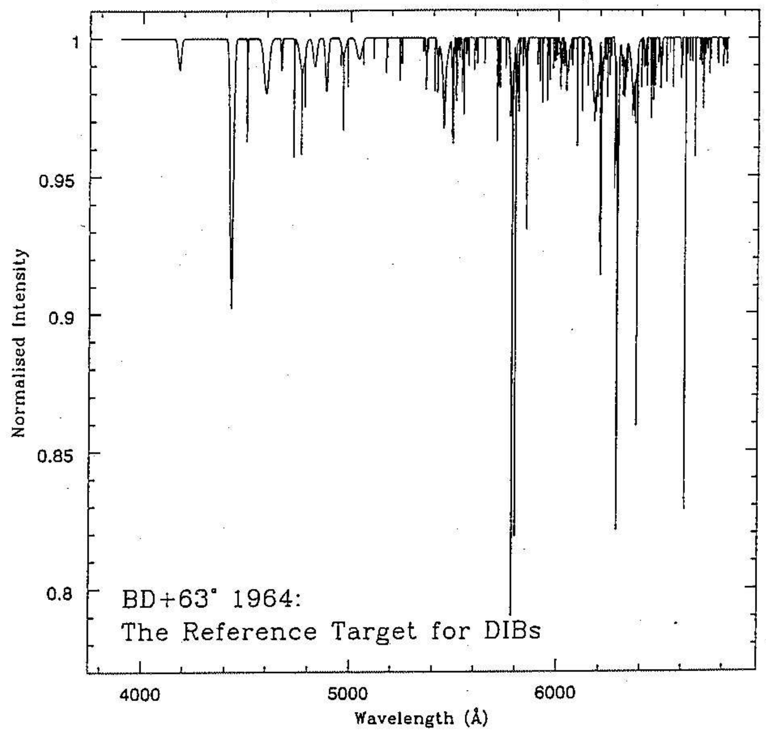


Figure 2: Spectrum towards the star BD+63 1964. (From Ehrenfreund & Charnley, 2000.)

There are several features of absorption spectra from diffuse clouds that are not well understood.

- UV BUMP at 220 nm. A strong feature in the absorption spectra of diffuse clouds is the bump att 220 nm. The proposed candidates for this feature are graphite gains, graphitic onions, hydrogenated amorphous carbon, and polycyclic aromatic hydrocarbons (PAHs) (Ehrenfreund & Charnley, 2000).

- Diffuse interstellar bands (DIBs). The diffuse interstellar bands lie between 400 and 1000nm. At the present more 250 DIBs are known but the carriers of these bands are not identified. Laboratory studies point toward PAHs, carbon chains, and carbon rings as potential carriers for some DIBs (Ehrenfreund & Charnley, 2000). Fig. 2 shows an absorption spectra due to DIBs.

- Extended red emission. The extended red emission seems to be a general property of interstellar dust. The proposed cadidates for this emission in the infrared are the same as those for UV bump.

These unexplained features all point toward large carbonaceous compounds and this would imply that a complex but not well understood chemistry is present in the diffuse clouds.

2.2 Dense clouds

The dense clouds are the sites of stellar and planetary formation. Their composition is mainly hydrogen gas and hence they are also called molecular clouds (Irvine, 1998). They show a complex structure with densities varying between 10^2 - 10^8 cm^{-3} and temperatures between 10 - 200K. They also show a high density of grains, which absorb and scatter UV photons, leading to a lower UV intensity than in the diffuse clouds.

The dust grains accrete mantles of frozen volatiles (ices). These mantles are formed by accretion of molecules followed by surface processes altering the molecules. The mantles are primarily composed of simple molecules such as H_20, NH_3 and CH_4, formed by elementary hydrogen reactions (Ehrenfreund & Charnley, 2000). In dense regions it is expected that most gas phase molecules will freeze out onto the grains. To sustain a gas phase abundance of molecules there has to exist a continuos emission process of molecules from the grains. These mechanisms of selective desorption and mantle explosion are not well understood (Irvine, 1998).

As the grains move into denser star forming regions, the higher temperature and increased UV intensity becomes important for the chemistry. Some of the mantles evaporate, leading to a more complex chemistry in the gas phase. This could lead to the formation of large molecules of biological importance (Ehrenfreund & Charnley, 2000).

Table 1 shows a list of molecules that up to now have been identified in the ISM.

3 Amino acids in the interstellar medium

The amino acids could be said to be the most basic building blocks of life, as we know it. They are the key components in proteins, which are fundamental for the creation of cells. Therefore it would be very interesting to find out if amino acids can be formed and survive in the interstellar medium. Amino acids have been found in meteorites and carbonaceous chondrites. They show an enhanced deutrium fractionation ratio similar to that observed in molecular clouds. This gives more credibility to the idea that life on Earth began with complex organic molecules being formed in the presolar nebula, and thereafter transferred to the early Earth.

3.1 Formation of amino acids

The first question to ask is whether it is possible for large molecules like amino acids to form in the ISM. A theoretical model predicts that this is possible in grains irradiated by UV light. Laboratory experiments also show that amino acids are produced in an environment similar to that of the ISM in a star forming region.

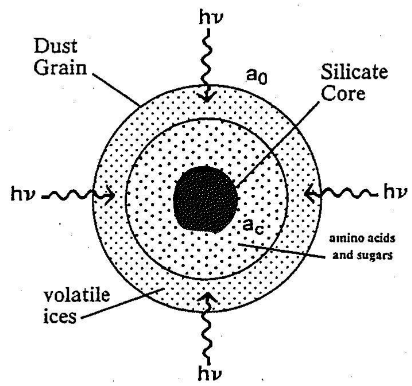
A theoretical model for how amino acids are formed in interstellar grains has been presented (Sorrell, 2001). According to this model UV photons irradiate an ice mantle composed of H_2O, CH_4, NH_3, and CO. Free OH, CH_3, and NH_2 radicals are formed. The radicals are very reactive and create simple amino acids and sugars, for example the simplest amino acid glycine, as can be seen in fig 3. The mantle protects the large molecules from UV radiation.

Bernstein et al (2002) have performed an experiment where H_2O ices containing CH_3OH, HCN, and NH_3 were vapor deposited at a temperature and pressure similar to that of a dense cloud (15K and 10^{-8} torr). The sample was also irradiated by UV light. Upon warming the ice sublimes, leaving a residue which was analyzed using chromatography-mass spectrometry. N-formyl glycine, cycloserine and glycerol were detected before hydrolysis and glycine, alanine, serine rea, ethanolamine were detected after hydrolysis. A similar experiment conducted by Munoz Caro et al (7) resulted in the identification of 16 different amino acids.

Number of Atoms										
2	3	4	5	6	7	8	9	10	11	13
H_2	C_3	c-C_3H	C_5	C_5H	C_6H	CH_3C_3N	CH_3C_4H	CH_3C_5N?	HC_9N	$HC_{11}N$
AlF	C_2H	l-C_3H	C_4H	l-H_2C_4	CH_2CHCN	$HCOOCH_3$	CH_3CH_2CN	$(CH_3)_2CO$		
AlCl	C_2O	C_3N	C_4Si	C_2H_4	CH_3C_2H	CH_3COOH?	$(CH_3)_2O$	NH_2CH_2COOH?		
C_2	C_2S	C_3O	l-C_3H_2	CH_3CN	HC_5N	C_7H	CH_3CH_2OH			
CH	CH_2	C_3S	c-C_3H_2	CH_3NC	$HCOCH_3$	H_2C_6	HC_7N			
CH^+	HCN	C_2H_2	CH_2CN	CH_3OH	NH_2CH_3		C_8H			
CN	HCO	CH_2D^+?	CH_4	CH_3SH	c-C_2H_4O					
CO	HCO^+	HCCN	HC_3N	HC_3NH^+						
CO^+	HCS^+	$HCNH^+$	HC_2NC	HC_2CHO						
CP	HOC^+	HNCO	HCOOH	NH_2CHO						
CSi	H_2O	HNCS	H_2CHN	C_5N						
HCl	H_2S	$HOCO^+$	H_2C_2O							
KCl	HNC	H_2CO	H_2NCN							
NH	HNO	H_2CN	HNC_3							
NO	MgCN	H_2CS	SiH_4							
NS	MgNC	H_3O^+	H_2COH^+							
NaCl	N_2H^+	NH_3								
OH	N_2O	SiC_3								
PN	NaCN									
SO	OCS									
SO^+	SO_2									
SiN	c-SiC_2									
SiO	CO_2									
SiS	NH_2									
CS	H_3^+									
HF										

Note that observations suggest the presence of large PAHs and fullerenes in the interstellar gas (Tielens et al 1999, Foing & Ehrenfreund 1997).

Table 1: List of molecules found in the ISM. (From Ehrenfreund & Charnley, 2000)



$$CO + OH \longrightarrow COOH$$

$$CH_3 + COOH \longrightarrow CH_3COOH$$

$$NH_2 + CH_3COOH \longrightarrow NH_2CH_2COOH + H$$

Figure 3: Potential pathway to glycine

3.2 The search for amino acids

The search for amino acids in the ISM has been concentrated to finding the simplest amino acid, glycine (NH_2CH_2COOH).

In the past a number of searches have been conducted, primarily at the SiO position in Orion and Sgr B2(OH). The detection techniques used were ground based single-element telescopes (Snyder, 1997). Only upper limits for the column densities were established in these measurements conducted between 1980 and 1996.

More recently two millimeter-wavelength arrays were used in the search: the Owen Valley Millimeter Array (OVRO) and the Berkely-Illinois-Maryland-Association (BIMA) Array (Snyder, 1997). The arrays give a high angular resolution over a wide field of view and structural filtering.

As explained above the greatest abundance of large molecules is expected to be found in star-forming cores were fusion has not yet started. These sites are too young to show any outward signs of star-formation and the enhanced molecular abundance lasts only for a short time. Furthermore all parts of a cloud are not in the same evolutionary state. Consequently a successful search for large molecules requires a high resolution over a wide field of view, and this is provided by interferometric techniques.

Structural filtering allows for detection of molecules in a small region while molecules over extended region are filtered out. This is important since thee large molecules are concentrated around the contracting core while the smaller molecules are almost equally abundant over large regions of a cloud. Although interferometry greatly enhances the chance of identifying glycine no guaranteed confirmation has been made, but there are indications of the molecule in Sgr B2(LMH).

In the laboratory glycine is formed when acetic acid (CH_3COOH) is combined with NH_2. Consequently a search for acetic acid could discover locations where glycine might be found. A number of single element measurements have been conducted without success. Using the BIMA and OVRO arrays acetic acid was confirmed to exist in Sgr B2(LMH) (Mehringer et al, 1997).

3.3 UV destruction of amino acids

One explanation to the unsuccessful search for glycine could be that the abundances are too small due to the rapid photo-destruction of the molecule. Laboratory experiments have been conducted where the photostabililty of four different amino acids was tested (Ehrenfreund et al, 2001). The experiments were performed in an environment similar to the interstellar medium. The results show that amino acids are destroyed well within the lifetime of typical cloud. This is also the case when a very low intensity of 10^3 photons cm^{-2} s^{-1} is assumed. In a high UV flux region (10^8 photons cm^{-2} s^{-1}), similar to a star-forming region, the amino acids are destroyed within a few hundred years. If however the molecules are shielded by ice mantles of interstellar grains they could survive for long periods of time but this also prevents the detection of these molecules. There will however be a steady state equilibrium of gas phase amino acids if the amino acids are formed in the ice and thereafter brought into the gas, as in the case of formation in grains which are destroyed.

4 Chirality of amino acids

Most amino acids are chiral, meaning that they exist in two mirror image forms, basically one can say left- (L) or right-handed (R). On earth almost all living systems use the left-handed amino acids exclusively. However when synthesizing amino acids in the laboratory one usually gets a 50:50 mixture of right and left-handed acids. The explanation to why living things favor left-handedness has not been answered. If amino acids were formed on earth there is no reason for believing that an excess of left-handedness was created and this has led to two different approaches to how life began. The biotic theory states that life started with use of mixed amino acids and evolution for some reason favoured left handedness. The abiotic approach assumes that a tendency toward homochirality was somehow present in the chemical evolution. The latter theory could be supported if an excess of left-handed amino acids were delivered to the early earth.

Analysis of the Murchison meteorite show that the left-handed amino acids present are in excess (7% and 9.1% for two different acids) of the right handed (Cronin & Pizzarello, 1997). The amino acids are not present in biological systems on earth so the excess cannot be a result of contamination. Also, the deutrium fractionation ratio is too high for the molecules to have come from the earth.

Laboratory experiments have shown that amino acid formation under circularly polarized UV light gives amino acids with an excess of one type of handedness. This implies that if the amino acids in the Murchison meteorite were exposed to circularly polarized light during their formation, the excess of left-handedness is explained. At the present there is no other explanation to why there is an enantiomeric excess. If there was such an excess in all amino acids delivered to the early earth this would support the abiotic theory of how life started.

Circularly polarized light could have come from a neutron star (Cronin & Pizzarello, 1997) near the presolar nebula. This is however a contested theory (Rubenstein et al & Bailey, 1999). An upper limit of 0.03% circularly polarized light has been established for the crab nebula, but theorists seem to disagree on wether a neutron star actually can emit circularly polarized radiation.

5 The future

The future of astrochemistry and the knowledge of organic molecules in the interstellar medium is very much dependent on the resources available to make observations. New space-based IR telescopes will undoubtedly provide more data. The ALMA project will lead to the possibility of detection of molecules whose quantities are only one hundredth of today's detection limit. This will most certainly be employed in the search for large molecules such as amino acids. However the density of these molecules, if they exist, is impossible to predict. It may be, that the densities are so small that we never will be able to detect them even though they are present in the gas.

New projects with the aim to gain more knowledge of comets in the solar system could give new clues to what type of molecules can be formed in the ISM. However it will only give us absolute knowledge what was possible when the sun was formed and no general information regarding the rest of the universe.

Theoretical modeling and laboratory experiments could as well help solve some of the questions in the field.

References

Bernstein et al 2002 Nature 416:401-403

Cronin J. R. & Pizzarello S. 1997 Science 275:951-955

Ehrenfreund, P. & Charnley, S. 2000 Annu. Rev. Astron. Astroph. 38:427-83

Ehrenfreund et al 2001 Ap. J. 550:L95-L99

Irvine, W. M. 1998 Orig. Life Evol. Biosph. 28:365-383

Mehringer et al 1997 Ap. J. 480:L71-L74

Munoz Caro et al 2002 416:403-406

Rubenstein et al & Bailey J. 1999 Science 283:1415a

Snyder, L. E. 1997 Orig. Life Evol. Biosph. 27:115-133

Sorrell W. H. 2001 Ap. J. 555:L129-L132

Chemistry of Comets: Frozen Cosmic Preserves

Gunnar Sporrong

Göteborg University
SE-41296 Göteborg, Sweden
(upc459e@pop.tninet.se)

*

Abstract

The history of the Solar System is probably one of the most interesting parts of astronomical science, concerning the development of life on the Earth. Most of the knowledge we have of the young Earth and also about the history of life is from mineral samples in the ground. But what happened before about 3 billion years ago is impossible to study because of the movement of the tectonic plates of the earth. Therefore we have to study the Solar System itself to get clues of the conditions in the early Solar System and on the Earth. Comets are what can be called fossils from the proto solar nebula and the young Solar System.

1 What is a comet?

A comet can be seen as a dirty snowball. A normal comet has a nucleus that has a diameter of about 1 – 100 km. The comet nucleus consists of ices, mostly H_2O, CO_2 and CO. In this ice mass we also find traces of a lot of different, more complicated, molecules, for example CH_3OH, H_2CO, HCOOH. In the ices are also, frozen in, dust particles of different sizes.

Normally the comets are found in the Kuiper belt or in the Oort cloud. When the comet orbit is disturbed and a comet starts to fall down into the inner parts of the Solar System, it starts vaporizing. First more volatile compounds as for example CO start to vaporize. Later when the temperature is rising, water vapour starts to outgas from the comet. (See Fig. 1).

Soon the nucleus is surrounded with a quite dense cloud of gas. This gas is called the coma (see Fig. 2) and has typically a diameter of 10 000's of km. When the comet comes even closer the coma is drawn out into a long tail by the solar wind.

The length of the so called ion tail is about 100 million km. The coma and the tail are the parts of the comet that we usually observe, because of the normally low brightness of the nucleus when it is far from the sun.

After the short visit in the inner Solar System (a couple of month's), the comet can either go back to the Oort cloud in a hyperbolic orbit or get disturbed in the inner parts

*Hot Topics in Astrophysics 2001/2002, Alessandro B. Romeo & Eva S. Karlsson (Eds.), Chalmers University of Technology and Göteborg University, 2002.

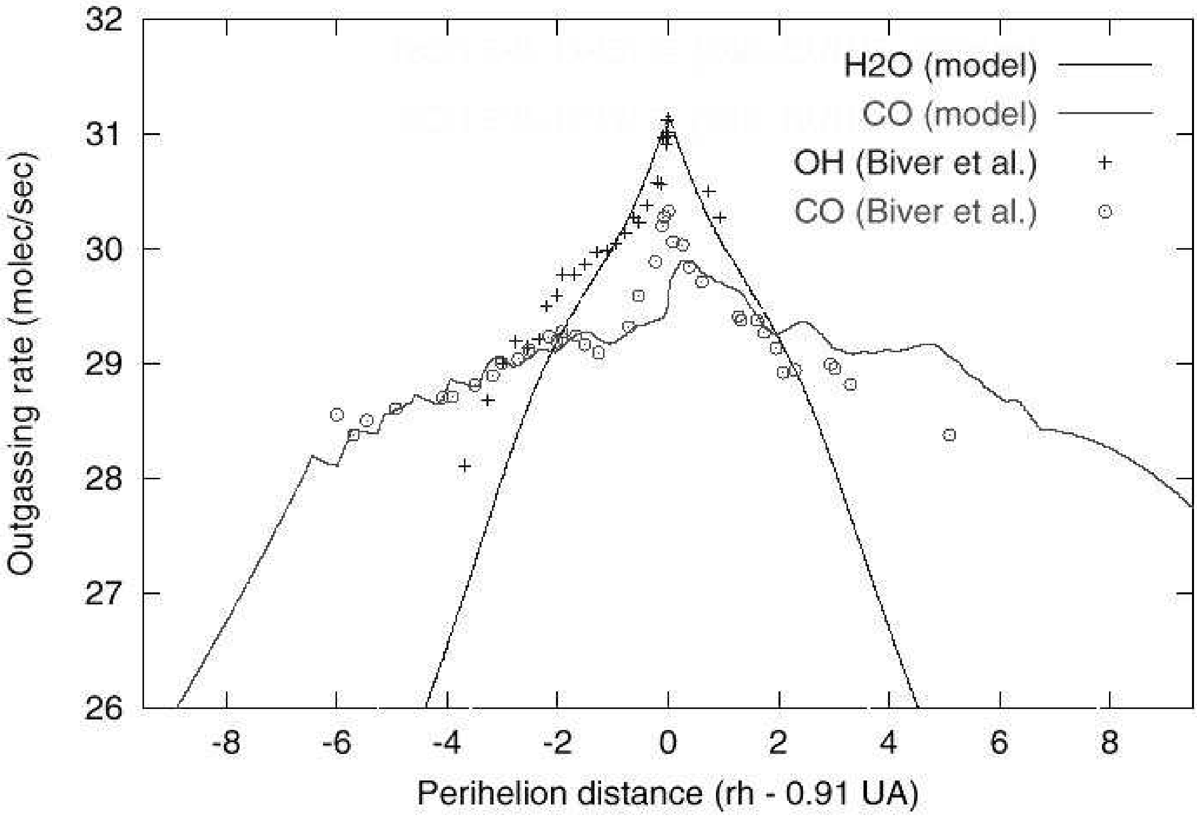


Figure 1: Evolution of the modeled CO and water outgassing rates (Enzian 1999) of comet Hale-Bopp compared with measured outgassing rates from astronomical observations (Biver et al 1999).

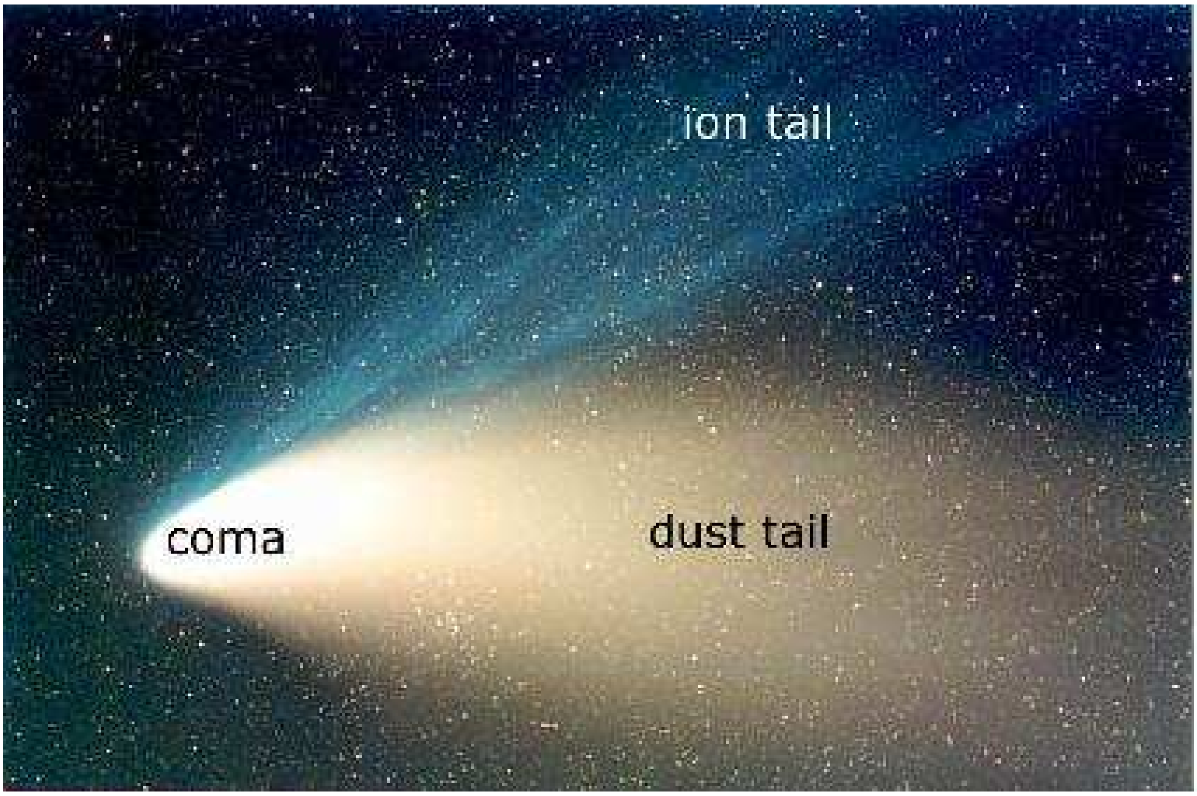


Figure 2: Typical features of a comet, (Hale-Bopp).

of the solar system to a new orbit. Either the new orbit reaches, in the outermost parts, outside Pluto and is called a long periodic comet or it gets a smaller orbit and becomes a short periodic comet (some years to a couple of hundred).

2 Were are comets formed?

In the beginning of the universe the elements present was mostly hydrogen (80%) helium (20%) and Lithium ($< 1\%$). During the years the universe has developed a lot of other elements. These elements have developed in old dying red giants and through supernova explosions. In galactical nebulas stars have been born throughout the history of the universe. The second and third generations of stars has got the newly formed elements cooperated into them. When a star is born in a cloud that contains heavier elements, there is also collected a lot of heavier elements in the accretion disc around the proto star. In this accretion disc a lot of chemical compounds from the gas cloud itself is collected.

When the proto sun did ignite the temperature in the inner parts of the disc increased. All volatiles transformed into gas and were ejected to the outer parts of the accretion disc. When they had reached a sufficiently long distance from the newly born sun, they condense and created the gas planets and also comets. These comets did probably contain most of the compounds that originally was in the nebula, in gas phase or on dust particles. By the gravity of the gas giants the comets was ejected to the Oort cloud, got captured into the planets, or sent into the Sun.

The comets we study today have spent most of their time so far away from the Sun that they have been frozen all the time. In this way they have been conserved from chemically altering radiation and reactions. Therefore we can be pretty sure that what

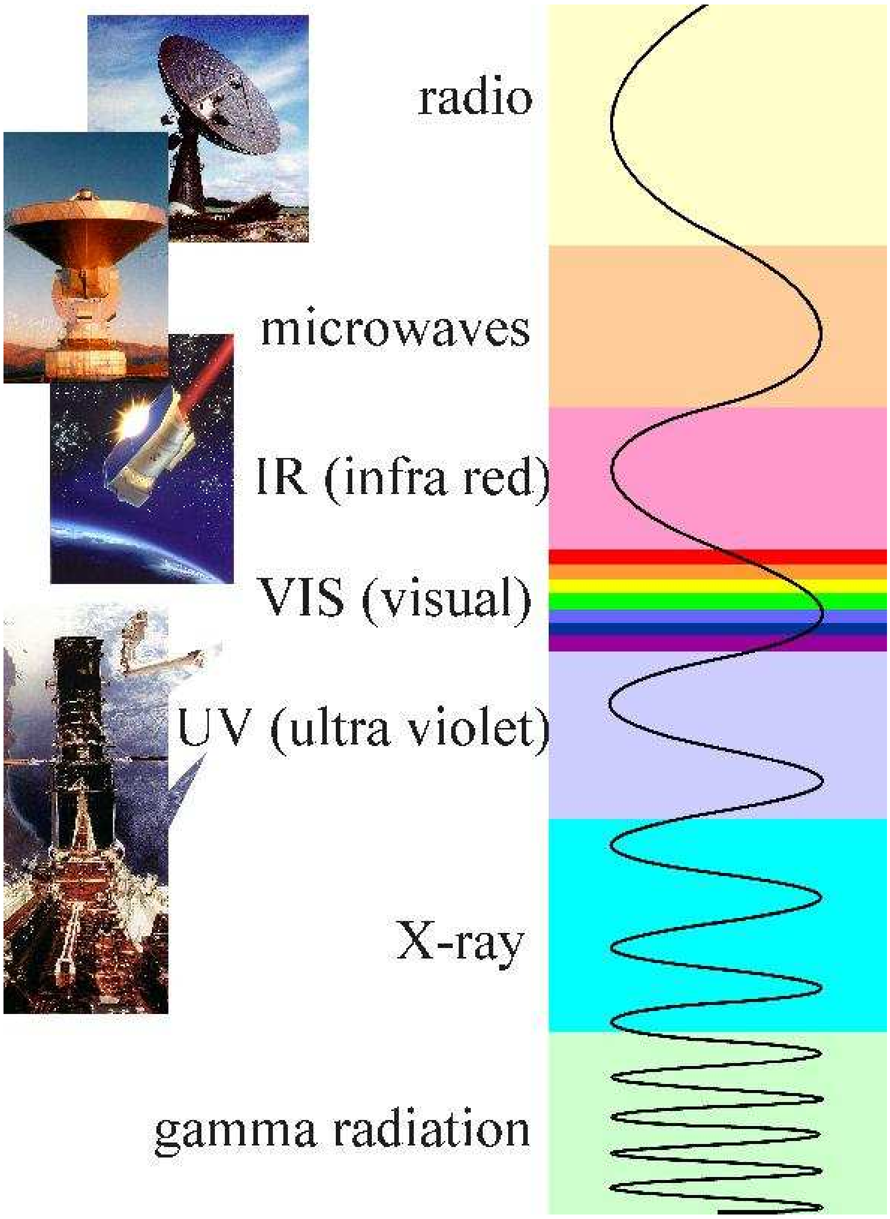


Figure 3: The electromagnetic spectrum.

we se today in comets is what was in the solar nebula 5 billion years ago.

3 The study of comets

As in many other areas of astronomy the cheapest and easiest way to study comets is by telescopes and spectrometry. When studying comets, one uses wavelengths from radio to ultra violet (UV) (see Fig. 3). Within the microwave area of the spectra we can distinguish transitions between different rotational energies of the molecules. In infrared (IR) radiation we observe vibrational transitions. In visible (VIS) and ultra violet (UV) light we study the electronic energies of the molecules. These different observational techniques need very different observational equipment which will not be explained in detail here.

When studying for example the rotational spectra of a molecule the spectra tell us a lot about the conditions were the molecules are situated. For example we can se the effects of the temperature of the molecules. In normal laboratory experiments, in gas phase, the spectra consists of lines equally spaced from each other and with continuously rising intensities from the outer part of the spectra to the middle. (That is quite symmetric.) Were the highest peak are found depends on the temperature.

When we study a cometary rotational spectrum, in VIS or UV, though, we don't find it as symmetric. The solar radiation is not as symmetric as the spectra of a light bulb. In the solar radiation we find hundreds of dips in the intensity. These dips come from the atoms in the solar atmosphere. These energy differences in quite nearby wavelengths give very differential excitations of the different rotational energies. Though the comet do move radially compared to the Sun, the Doppler effect make the solar spectra and the excitational absorption spectra for the molecules in the comet to move compared to each other. Therefore the emissionlines from the molecules rotational spectra differs with time. Also there is only a few of the lines that is representative for calculations of the temperature of the gas. At the same time the Doppler effect that comes from the relative motion between the Earth and the comet has to be accounted for. Otherwise we would not know what molecule we are observing.

When we study comets with a spectrograph, the slit often covers whole or big parts of the coma. This is necessary to get high light sensitivity. The molecules that are found with this method cannot be placed in the right part of the coma observationally. To know were the different molecules are situated we have to use models for how the vaporizing and UV decay of the molecules happens. Even though all this may seem complicated the astrochemical studies has reached quite far.

4 What have we found?

Comets contain water, carbon monoxide, carbon dioxide, methanol, hydrogen sulfide etc. This has been known for more than 100 years but most of the molecules in comets is only traces though. To find them one needs quite advanced equipment. Therefore most of the more advanced compounds have been discovered during the last 20 years, in Halley's comet in the 80's, comet Hyakutake and comet Hale-Bopp in the 90's. In table 1 below only molecules that are positively identified in the three comets and in an proto stellar object are shown.

Generally one could say that the most complicated molecules are found closest to the nucleus. Due to the high intensity of ultra violet radiation a lot of photo chemical reactions probably takes plase in the outer parts of the coma. Most reactions breakes down the molecules, but some do also make new ones. That is one of the largest problem with remote studies of the comets. To get enough signal to noise ratio when taking a spectra of the comet, the slit in the spectrograph will have to cover most parts of the coma. The result is that we can find out what molecules the coma contains but not were in the coma we find them. For example, we find both the molecules SO and SO_2 but we don't know if CO is a daugter molecule to SO_2. It could also be that the SO_2 is a chemical or maby photochemical reactionproduct due to the quite high pressure in the coma that enables chemical reactions.

5 Future plans of investigation

To really answere the problem of the compositions of comets there are two ongoing projects. Satellite Stardust, that whas launched February 1999 is going to visit comet

Molecule	Elias low	Halley	Hyakutake	Hale-Bopp
H_2O	100	100	100	100
CO	5.6	15	6–30	20
CO_2	22	3	2–4	6–20
CH_4	< 1.6	0.2–1.2	0.7	0.6
CH_3OH	< 4	1–1.7	2	2
H_2CO	-	0–5	0.2–1	1
OCS	< 0.08	-	0.1	0.5
NH_3	< 9.2	0.1–2	0.5	0.7–1.8
C_2H_6	-	-	0.4	0.3
HCOOH	-	-	-	0.06
OCN^-	< 0.24	-	-	-
HCN	-	0.1	0.1	0.25
HNC	-	-	0.01	0.04
HNCO	-	-	0.07	0.06–0.1
C_2H_2	-	-	0.5	0.1
CH_3CN	-	-	0.01	0.02
$HCOOCH_3$	-	-	-	0.06
HC_3N	-	-	-	0.02
NH_2CHO	-	-	-	0.01
H_2S	-	0.04	0.8	1.5
H_2CS	-	-	-	0.02
SO	-	-	-	0.2–0.8
SO_2	0.6	-	-	0.1

Table 1: All values are normalized to the water content. The Elias 29 low-mass protostellar object is here an example of what could have been the case in our Solar System 5 billion years ago. (From Ehrenfreund and Charnley, 2000.)

Wild, fly pass it an collect samples from its coma. It will in 2006 return with the samples to Earth for analysis. The Rosetta satellite visit comet Virtanen. Comet Virtanen completes an orbit around the Sun every 5.5 years. The Rosetta satellite will take a ride on the comet (land) and make investigations during the time the comet is accelerating towards the Sun. It will reach a speed of 135 000 km/h. It will later return to the Earth and land in 2011.

Acknowledgements

I whish to thank Bengt Tegner and Theresa Wiegert for help with the editorial work and Elin Nordenmark for valuable support with the presentation.

References

Biver N. et al. 1999. Astron. J. 118(4):1850-72

Bocklée-Morvan D. et al. 2000 Astron. Astrophys. 353, 1101-1114

Ehrenfreund P. and Charnley S. B. 2000 Astron. Astrophys. 38:427-483 (2000)

Greenberg J. M. and Li A. 1998 Astron. Astrophys. 332, 374-384

Wyckoff S. et al. 2000 Astroph. Journal 535:991-999, June 1

The Search for Extra-Solar Planets

Eva S. Karlsson

Chalmers University of Technology
SE-41296 Göteborg, Sweden
(f97evka@dd.chalmers.se)

*

Abstract

Since 1995 more than 70 extra-solar planets orbiting solar type stars have been detected. Various techniques have been used of which the Doppler method have been the most successful. Since this method is biased towards detecting massive planets in small orbits, the planetary systems identified until today do not represent a statistically reliable sample to base a realistic model for planet formation on. The space missions planned for the coming 20 years would provide this.

1 Introduction

In 1995 the first planet orbiting another star than the Sun was detected (see Mayor & Queloz, 1995). Since then there have been many reported observations of extra-solar planets, of which a total of 77 have been confirmed (see Schneider, 2002). How have we been able to detect them? What are the limitations of our current detection methods? What possibilities to find and probe Earth-like planets do the future hold?

1.1 What is a planet?

The definition of a planet is a very debated subject. Planet-like objects have been discovered not only orbiting main sequence stars, but also orbiting pulsars and free-floating in young stellar clusters (see Wolszczan & Frail, 1992; Béjar et al., 2001).

Which of these are called planets? Most astronomers agree on that a planet should not be able to burn deuterium through thermonuclear fusion. Therefore the upper mass limit of a planet lies between 10–15 M_J, where M_J denotes the mass of Jupiter, depending on its chemical composition. An object that is massive enough to burn deuterium, but not hydrogen, is called a brown dwarf. (An object burning hydrogen is a star.)

There are theories, for example concerning the formation processes (see sect. 1.2), that argue for and against calling the free-floating and pulsar-orbiting objects planets. This paper will not render the complete discussion, but uses the more limiting definition that a planet is orbiting a solar-type star. The lower mass limit discerning planets from asteroids is most commonly chosen as that of Pluto.

*Hot Topics in Astrophysics 2001/2002, Alessandro B. Romeo & Eva S. Karlsson (Eds.), Chalmers University of Technology and Göteborg University, 2002.

1.2 How do planets form?

There is a number of theories of how planets are formed. The most widely accepted is the 'solar nebula theory' that has been constructed based on observations of the Solar system.

In this theory the planets are formed in a flattened disk rotating about a protostar, a "star-to-be". The material in this protoplanetary disk originates in the huge cloud of gas and dust, the 'nebula', from which the parent star is being formed. Pairwise collisions between microscopic particles in the disk eventually makes them stick together to form macroscopic objects. On a time-scale of 10^4–10^5 years the growth continues until there are solid cores with diameters in the order of a kilometer, planetesimals. In the process of growth these have been lined up in orbits divided by nearly empty gaps. Gravitationally driven collisions between planetesimals and accumulation of more disk material eventually result in full size planets. The ones that become massive enough ($M_P > 10M_\oplus$, where M_P denotes the mass of the planet and $M_\oplus$ denotes the mass of Earth) accrete gas from the disk and end up as Jovian-type planets. For a more through description of of the solar nebula theory, see Lissauer (1993).

This model for how planets form has in recent years been supported by observations of protoplanetary disks around other stars. Compared to the planets themselves the disks are easy to detect and some have for example been imaged using the Hubble Space Telescope (see O'Dell et al. 1993). The solar nebular theory is though far from complete. For example, the theory predicts planets in nearly circular orbits, then how does it come that many of the extra-solar planets have highly eccentric orbits?

1.3 Life on other planets?

When looking for extra-solar planets we are especially interested in those that we think may host forms of life. Finding extra-terrestrial life would provide us with invaluable information about the beginning and evolution of life on Earth and elsewhere.

The basic requirements for life are thought to be the occurrence of liquid water and access to the biogenic elements that are the building blocks of life. Rocky planets around other stars should contain the same elements as the Earth does, assuming that planets form from the same material as their parent stars (see sect. 1.2). Even though the proportions may be different, the important biogenic elements are there. The 'habitable zone' is defined as the region around a star where a planet hold a surface temperature such that water may exist on it in liquid form. Inside this zone the water would evaporate and outside it would freeze. The location of this zone depends on the amount of energy the star emits and variates therefore with time. The presence of liquid water also depends on the nature of the planets atmosphere.

2 Detection of extra-solar planets

A planet is much smaller and lighter than its parent star and it is not as luminous. But it does reflect the stellar light and may also radiate in the infra-red part of the spectrum. If we were able to take the spectra of an extra-solar planet, we would see whether it had an atmosphere and what it consisted of.

There are also indirect ways of detecting planetary systems. When a planet of mass M_P is orbiting a star of mass M_*, both bodies are rotating the center of mass of the

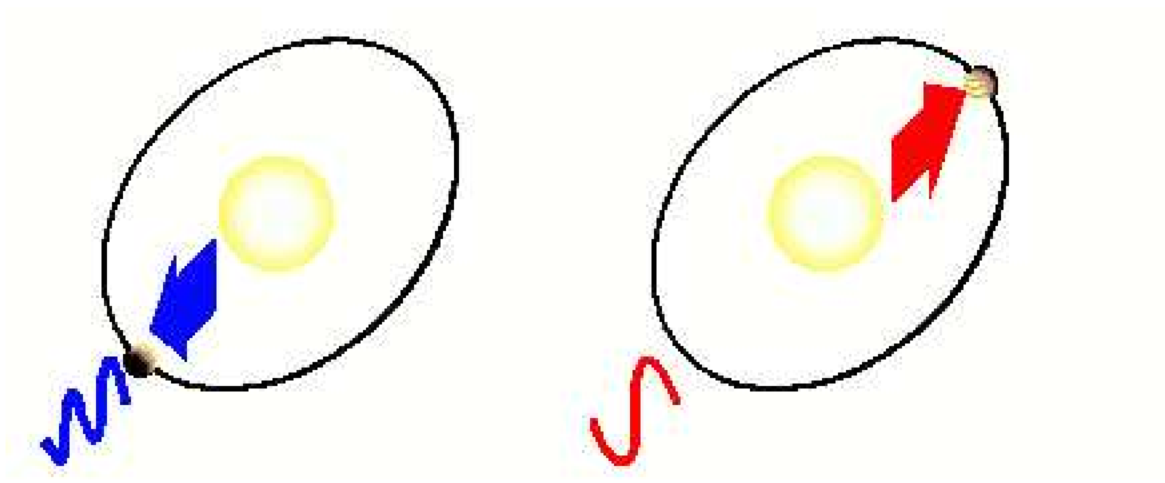


Figure 1: Illustration of the Doppler effect on the light from a star with a planet. Courtesy of G. Esquerdo and the Arizona Search for Planets Project.

system, at different radii. The rotation radius of the star, a_*, is given by:

$$a_* = a \cdot \frac{M_P}{M_*}, \tag{1}$$

where a represents the radius (or the semi major axis) of the planet's orbit. By observing this periodical movement of a star, conclusions may be drawn about its planetary companion. This is done by comparing different models of planetary systems and choose the parameters that best fit the observations.

2.1 California and Carnegie Planet Search

The California and Carnegie Planet Search (CCPS) project is a collaboration between University of California and Carnegie Institute of Washington (Department of terrestrial magnetism). The project team uses the Doppler method, measurements of the radial velocity of stars (see below), to find extra-solar planets. Telescopes and spectrometers at Lick observatory in California and Keck observatory at Hawaii are employed to perform the searches. The project has been successful with a record of over 40 confirmed detections of extra-solar planets. Measurements of radial velocity have led to the detection of most of the planets around other stars that we know of today.

When observing spectral lines from a star with a planet at high enough resolution, one can detect wavelength variations (after having compensated for Earths movement relative the stellar background). The variations are produced as the star alternately moves towards the spectrograph and away from it, pulled by the gravitation of the orbiting planet (see sect. 2 and Fig. 1).

The observed wavelengths correspond to relative velocities of the star. These velocity variations do not only depend on the star-planet mass ratio and the radius of the planetary orbit (see Eq. (1)), but also on the eccentricity e and the inclination i of the orbit. The observable velocity amplitude, K, can be expressed in terms of these variables and the orbital period P (see Cumming et al. 1999):

$$K = \left(\frac{2\pi G}{P}\right)^{1/3} \frac{M_P \; sin \; i}{(M_P + M_*)^{2/3}} \frac{1}{\sqrt{1-e^2}}. \tag{2}$$

Note that if the system is seen face on (i=0), no velocity variations can be observed. This kind of detection does not give information about the orbit inclination. The $\sin i$ dependence then implies that we are not able to determine the companion mass with this method, only a lower limit to it. Kepler's third law relates the orbital period to the orbital size:

$$P = 1\text{yr} \left(\frac{a}{1\text{AU}}\right)^{3/2} \left(\frac{M_\odot}{M_*}\right)^{1/2}. \tag{3}$$

Assuming a circular orbit and that $M_P \ll M_*$ we get the relation

$$K = 24,8\text{ms}^{-1} \left(\frac{1\text{AU}}{a}\right)^{1/2} \frac{M_P \; sin \; i}{M_J} \left(\frac{M_\odot}{M_*}\right)^{1/2}. \tag{4}$$

For reference Earth perturbates Sun to a velocity amplitude of about 0.1 $\text{m}\,\text{s}^{-1}$ and Jupiter to 12.5 $\text{m}\,\text{s}^{-1}$. The accuracy of the instruments used by CCPS allows for detection down to 3 $\text{m}\,\text{s}^{-1}$ (see Butler et al. 1996). This means that this program has no chance of discovering Earth-like planets in the habitable zone (see sect. 1.3).

2.2 Coralie Planet Search

The Coralie Planet Search is a research program at the Geneva Observatory in Switzerland. Coralie is the spectrograph, mounted at the European Southern Observatory (ESO) at La Silla in Chile, that the program is employing to measure radial velocity variations (see sect. 2.1). It surveys over 1650 stars in the southern sky and has identified several extra-solar planets this far. The highest resolution Coralie can produce is 3 $\text{m}\,\text{s}^{-1}$ (see Queloz et al. 2001).

The Geneva Observatory is running a project with a new spectrograph under construction, the High Accuracy Radial velocity Planetary Search, HARPS. HARPS will be installed at the 3.6 m telescope at La Silla. The spectrograph was designed with Coralie as raw model and is expected to achieve an accuracy of about 1 $\text{m}\,\text{s}^{-1}$ (see Queloz et al. 2001). This will be enough to detect Earth-size planets in small orbits. The limit for how accurate radial velocity measurements may be is soon reached. Bouchy et al. (2001) describe how this fundamental limit is given by photon noise.

The Doppler technique is biased towards finding massive planets in small orbits. Since it only gives a lower limit to the planet's mass, some of the detected planets may well turn out to be brown dwarfs. It have also been argued that the observed velocity variations may arise from surface magnetic activity or stellar pulsations. Therefore it is desirable to get confirmations through other techniques.

2.3 Hipparcos

In August 1989 a satellite was launched and during the more than three years it was active it collected astrometric data for 120 000 stars. This was the Hipparcos Space Astrometry Mission, the only one of its kind carried out to date. The resulting Hipparcos Catalogue contains the positions, parallaxes and proper motions of all of these stars to a precision of 1 milliarcsec (mas). These data have not led to the detection of any planetary systems, but they have given constraints to planetary masses in systems discovered by other means. How is this done?

A star with a planet companion traces an ellipse on the sky as seen from us, due to the gravitational pull from the planet (see sect. 2). Using high precision astrometry one can measure the angular semi-major axis, α, of this ellipse,

$$\alpha = \frac{a}{d} \cdot \frac{M_P}{M_*}, \tag{5}$$

where α is in arcsec when a is in AU and the distance d is in pc. Since α does not depend on the inclination of the planet's orbit, unlike the radial velocity amplitude, it gives the mass of the planet directly. Some planetary candidates discovered by the Doppler method have therefore been identified as brown dwarfs using Hipparcos data.

For reference, looking at Sun from a distance of 10 pc, Earth alone would give rise to an α of 0.3 microarcsec (μas) and the corresponding value for Jupiter is 0.5 mas (see Perryman, 2000). (Note that the presence of several planets gives rise to a more complex motion over the years.) In order to detect planets with this method sub-mas astrometry is needed. This is in principle impossible to achieve from ground due to atmospheric phase fluctuations.

2.4 GAIA

GAIA is a space astrometry mission within the scientific program of European Space Agency (ESA). The satellite is planned to be launched around 2011-2012 and it will be collecting data from more than a billion stars, in our galaxy and in the local group, during five years. One of the mission's main goals is to render a three dimensional map over these stars with a never seen precision. The accuracy of GAIA's astrometric measurements will be at least 10 μas down to apparent magnitudes of 12. Therefore the measurements of the mission will make it possible to search solar-type stars out to a distance of 200 pc for Jupiter-type planets with orbital periods of 1.5 to 9.0 years. The mission will also make it possible to detect multiple planet systems in the Solar neighborhood (see Sozzetti et al. 2001).

Since this detection method is sensitive to planets in large orbits (see Eq. (5)) and gives the mass of the planet directly, it represents a good complement to the Doppler method. Having both kinds of detections will render both the mass of the planet and the inclination of the orbit.

2.5 Microlensing Planet Search

The Microlensing Planet Search project (MPS) aims to find extra-solar planets by recording deviations in lightcurves from gravitational microlensing events. Microlensing occurs when a nearby star passes in front of a background light source. The potential of the star then acts as a gravitational lens that bends and magnifies the background light (for a more through description of the theory behind gravitational lensing, see Schneider et al. 1992). As the lensing star passes in front of the source, the background light is first amplified and then attenuated. If the star has a planet, the result is a multiple lens that may give a detectable perturbation to the otherwise smooth light curve (see Fig. 2).

MPS monitors stars towards the galactic bulge using the 1,9 m telescope at the Mount Stromlo Observatory in Australia. This way microlensing events are recorded and examined for anomalities that could result from the presence of a planet. The photometric microlensing technique makes it possible to detect planets down to Earth-size, but exact alignment is required to have a detectable brightening. The probability for a detectable microlensing event towards the galactic bulge is very small, less than one in a million. The signature of a planet should though be relatively frequently seen when detecting

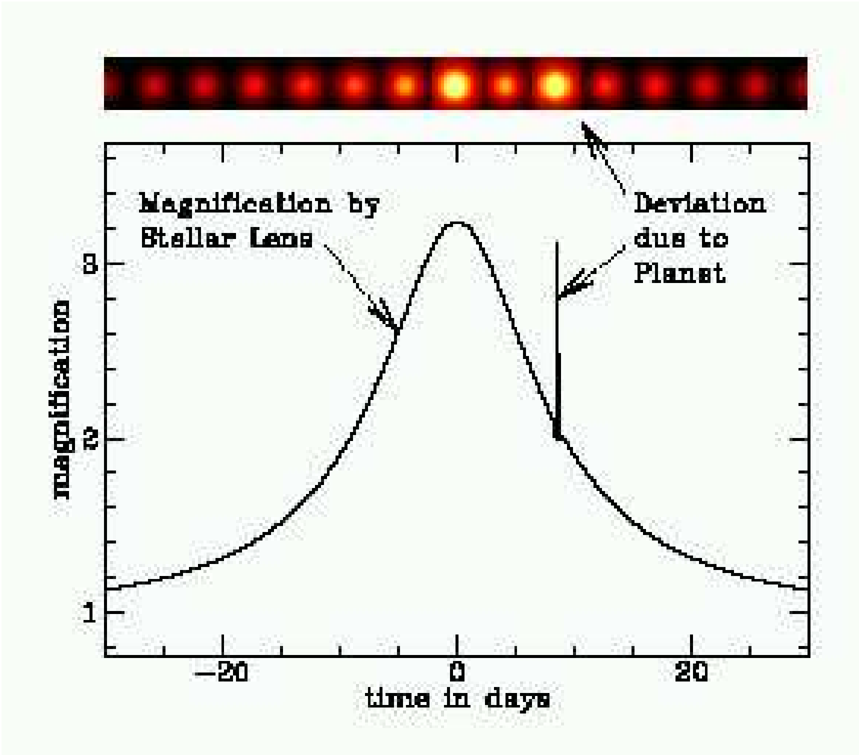


Figure 2: A theoretical microlensing light curve from a star with a planet. (From Bennett & Rhie, 1996).

a microlensing event of a planetary system (see Perryman, 2000). The MPS team has found some planetary system candidates and is waiting for opportunities to confirm these detections by other means.

2.6 STARE

STARE (STellar Astrophysics & Research on Exoplanets) is a photometric telescope and a project initiated by the High Altitude Observatory (HAO) in Colorado. The project is regularly monitoring over 24 000 stars in search for signs of planetary transits. STARE is mounted on the island of Tenerife in the Canaries, mainly watching a 5.7 degree square field of the Milky Way.

Planetary transits show as periodic drops in the stellar luminosity L_* during eclipses of a star by a planet in a nearly edge-on orbit. The observed drop in luminosity is:

$$\Delta L \simeq L_* \cdot \left(\frac{R_P}{R_*}\right)^2, \tag{6}$$

where the R's represent the physical radii of the planet and the star. Equation (6) holds assuming that the surface brightness of the star is constant. In reality this is not true and the brightness varies with stellar radius due to 'limb darkening' (the closer the edge of the star, the thicker is the atmosphere the light has to pass) but using the maximum luminosity drop, it gives a good estimate of the planet's radius. From the duration of the drop in luminosity, together with the observed orbital period, the orbital inclination may be derived (see Perryman, 2000).

In principle, drops in a star's luminosity may rise from other phenomena than transits, such as stellar surface activity or a binary companion. In order to be sure of the presence

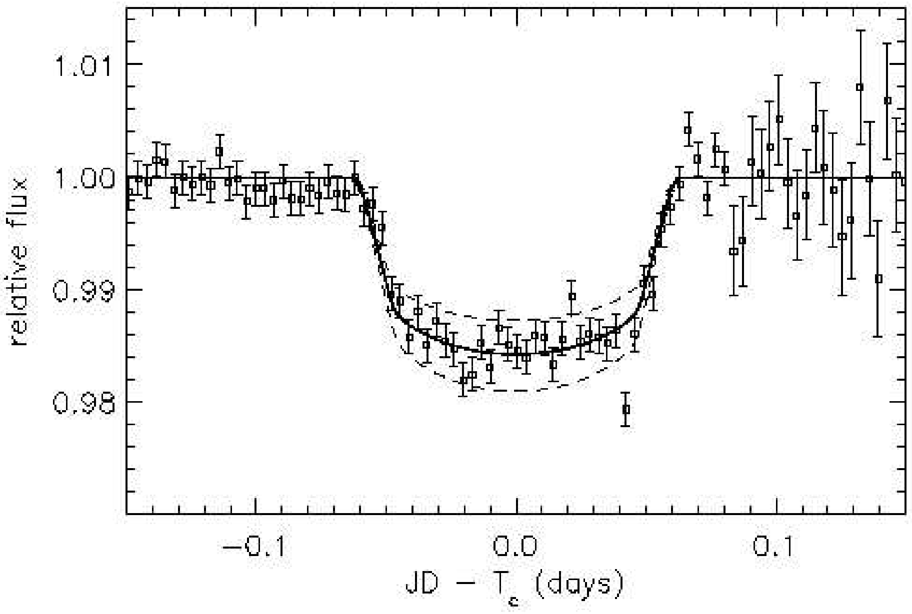


Figure 3: Superposed lightcurves of star HD 209458 showing transits occurring on 9 and 16 September, 1999. (From Charbonneau et al. 2000)

of a planet one should detect the same form of luminosity drop several times, recording a constant period. Two planetary transits were detected by STARE in September 1999 (see Fig. 3). These transits reveal a planet of 0.6 M_J orbiting the star HD 209458 (see Charbonneau et al. 2000).

The probability of observing transits for a randomly orientated system is very small. For an orbit with a certain inclination this probability depends on the size of the orbit and the stellar radius. If a planet has been detected by other means and it is believed to have an edge-on orbit, transit observations could be used to determine the planets mass. These could also render information about the planet's atmosphere since the stellar light is passing through this atmosphere on its way to us.

It is more likely to detect a transit from a space-based photometry telescope than from ground. The reason is that the atmosphere puts limits on the accuracy of the measurements. In addition a space-based telescope may make very long uninterrupted observations. Project of this kind are planned both in US and in Europe.

2.7 Darwin

Darwin is a European Space Agency (ESA) project for direct imaging of extra-solar planets from space. The project both aims at detecting the point-like sources of reflected starlight that planets produce and performing spectral analysis of the planets' atmospheres. This is a long term project, launch is planned to around 2014, and the exact design and technology of the mission have yet to be defined. Only the concepts for achieving its goals are worked out.

To image a planet is like trying to discern a candlelight next to a lighthouse from a distance of 1000 km. The luminosity-ratio between planet and star is:

$$\frac{L_P}{L_*} \propto p(\lambda) \cdot \left(\frac{R_P}{a}\right)^2, \tag{7}$$

where the function $p(\lambda)$ depends on the observed wavelength, the albedo of the planet and the angle between the star and the observer as seen from the planet (see Perryman, 2000). In the visible part of the electromagnetic spectrum this ratio is about one in a billion for a Jupiter-type planet. In the infra-red it is "only" one in a million and this is one of the reasons for Darwin to operate in this part of the spectrum.

Choosing an appropriate wavelength is not enough to obtain an image of a planet, the light from the star has to be blocked somehow. Darwin will obtain this using a technique called nulling interferometry. This is when the light from a star (and a planet) is simultaneously registered by two (or more) telescopes. The signals from the two telescopes are combined, but before that one of them is delayed by half a wavelength. If the star has a planet companion and the two are separated by an angle, this means that while the signals from the star are cancelled out, the signals of the planet may be reinforced (see Angel & Woolf, 1997).

In order to obtain a resolution high enough to resolve planetary companions a telescope with a large diameter is needed. Since this is not a plausible solution for a space based telescope, the Darwin project intends to use interferometry between several (five or six) smaller telescopes. Then the combination of signals from them synthesize a telescope with diameter as large as the baseline between them.

The Darwin interferometer is planned to be placed at L2, the Lagrange point 2, of the Sun-Earth two-body gravitational system. This gravitationally stable point is situated about four moon orbital radii out from Earth, on the opposite side with respect to Sun. Here the telescopes will be able to point away from Sun, Earth and Moon at all times in order to avoid their radiation. There are two basic proposals for the configuration of the telescopes. Either they will be connected by a fix structure or they will fly in formation and thereby be able to change their positions and get more data-points during the same exposure.

Calculations have shown that Darwin will be able to detect an Earth at a distance of 3 pc. The possible atmosphere of the planets that are found will go through spectral analysis. Since one of life's signatures is that it changes the chemical balance in its surroundings, this analysis may show signs of living organisms.

3 Discussion

The planetary systems that we have identified until today do not represent a statistically reliable sample to base a good model for planet formation on. Massive, Jupiter-sized planets in small orbits are overrepresented and the smallest planet detected has a mass 50 times that of Earth! The here presented projects that are still in their initial phase will hopefully, together with others, change that over the coming 10–20 years. In the same time span we will also be able to collect signs of extra-terrestrial life, assuming that life is a fairly common occurrence in our universe.

Space based missions are needed in order to receive better resolution no matter what detection technique is preferred. Unfortunately these missions tend to be costly and do not benefit from the economic situation of the western world today. In fact both NASA and ESA are reviewing their research programs and the funding for GAIA and Darwin, amongst others, are unsure at the moment. Collaboration between the space agencies have started to be considered and would be a solution seeing to the interests of both politicians and scientists. For example, NASA is running a project similar to Darwin called Terrestrial Planet Finder (TPF). Why not merge the missions to one, get access to the expertise from both agencies and cut the total expanses in half?

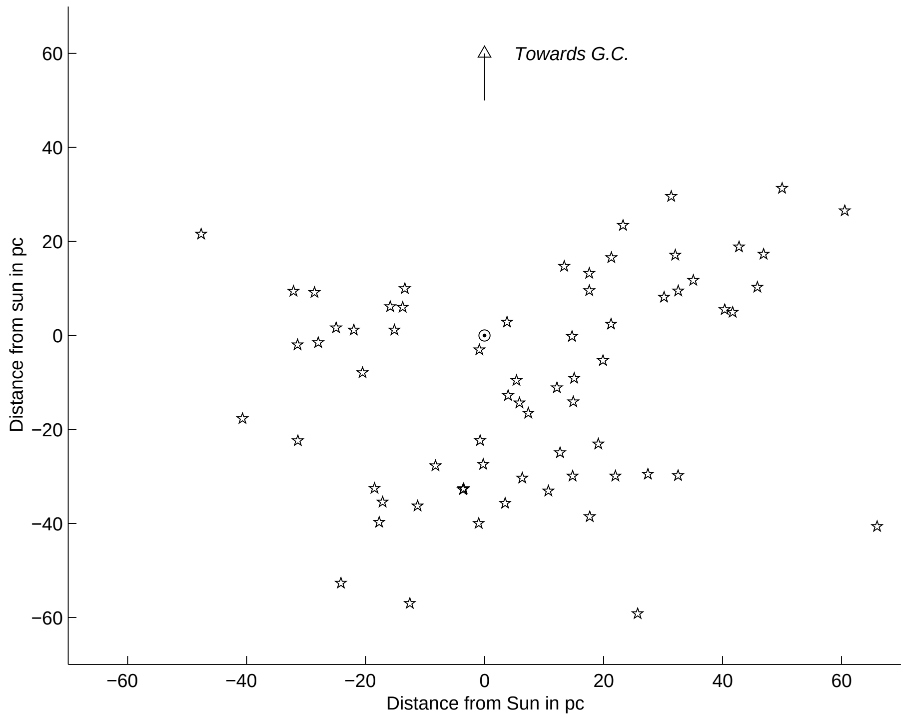


Figure 4: The distribution of known planetary systems in the galactic plane.

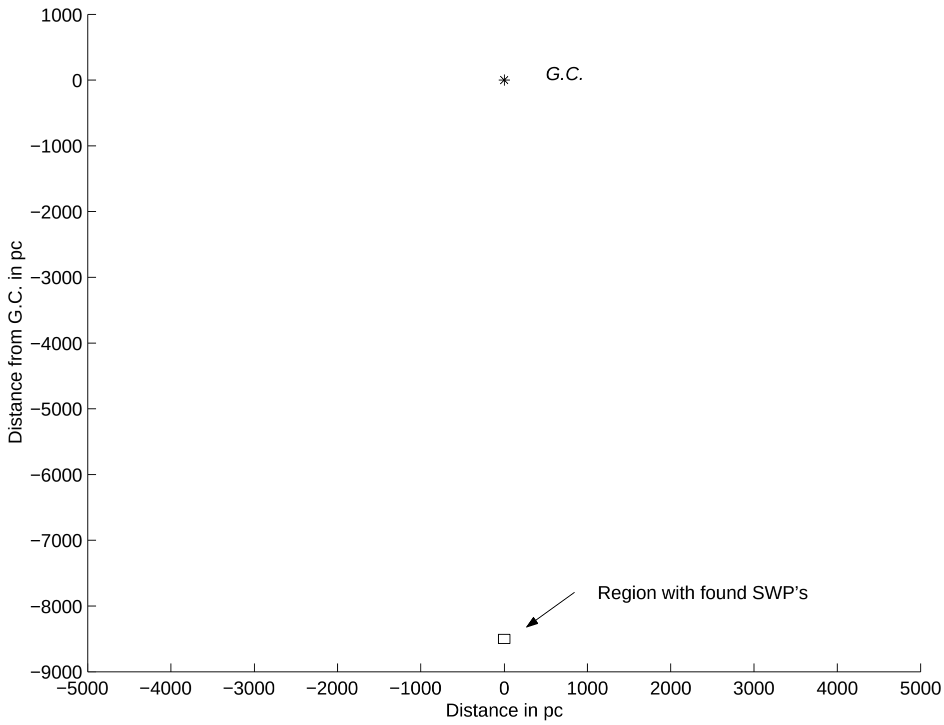


Figure 5: The region with known planetary systems on a galactic scale.

Using data accessible from Schneider (2002) I have produced figure 4 that shows the distribution in the galactic plane of the planetary systems we know of today. For a reference, the size of the region where they are found is compared to the distance to the Galactic center in figure 5. With this comparison in mind, the search for extra-solar planets may continue a long time yet.

References

Angel, J.R.P. & Woolf, N.J. 1997, ApJ, 475, 373.

Béjar, V.J.S. et al. 2001, ApJ, 556, 830.

Bennett, D.P. & Rhie, S.H. 1996, ApJ, 472, 660.

Bouchy, F., Pepe, F., & Queloz, D. 2001, A&A, 374, 733.

Butler, R.P., Marcy, G.W., Williams, E., McCarthy, C., Dosanjh, P., & Vogt, S.S. 1996, PASP, 108, 500.

Charbonneau, D., Brown, T.M., Latham, D.W., & Mayor, M. 2000, ApJ, 529, L45.

Cumming, A., Marcy, G.W., & Butler, R.P. 1999, ApJ, 526, 890.

Lissauer, J.J. 1993, ARA&A, 31, 129.

Mayor, M. & Queloz, D. 1995, Nature, 378, 355.

O'dell, C.R., Wen, Z. & Hu, X. 1993, ApJ, 410, 696.

Perryman, M.A.C. 2000, Reports of Progress in Physics, 63, 1209.

Queloz, D. et al. 2001, The Messenger, 105, 1.

Schneider, J. 2002: *The Extrasolar Planets Encyclopaedia*, http://www.obspm.fr/encycl/encycl.html.

Schneider, P., Ehlers, J. & Falco, E.E. 1992, *Gravitational Lenses*, XIV, Springer-Verlag Berlin.

Sozzetti, A., Casertano, S., Lattanzi, M.G., & Spagna, A. 2001, A&A, 373, L21.

Wolszczan, A. & Frail, D.A. 1992, Nature, 355, 145.

Supernova Explosions

Jan Lindahl

Göteborg University
SE-41296 Göteborg, Sweden
(gu98jali@dd.chalmers.se)

*

Abstract

Supernova explosions Type II and Type Ia are believed to have different progenitors. Though they have been studied in detail since the 1930s the question of how they explode has not yet been answered. Observations suggests that the Type I progenitor is a white dwarf that accretes material from a binary companion, and Type II is a massive star that collapses and leave neutron stars.

1 Introduction

A supernova is an exploding star. They are almost as bright as the whole galaxy where they exist. Since there are billions of stars in a galaxy, yet altogether they are emitting almost as much light as the single brilliant star. Supernovae represent the end point of normal stellar evolution, so by studying them and the evolutionary pathways that lead to them, we can gain insights into the whole course of stellar evolution. In the evolution and subsequent explosion, supernovae forge the majority of the heavy elements and hence drive the chemical evolution of the whole galaxies. In addition, the energetic input of supernovae may influence the birth of new stars by compressing interstellar gas. Because supernovae are so bright they can be observed at great distances and can serve as distance indicators to measure the very scale of the universe. The kinetic energy of the explosion is of the order $10^{51} ergs$, and the velocity of the ejected material is about a hundredth of the speed of light.

Several supernovae are described in ancient literature, in China they observed the star which exploded in 1054, now the Crab nebula. Other famous supernovae are the ones observed by Tycho in 1572 and Kepler in 1604, Keplers supernova is the last one observed in our galaxy. So how often do this events happen? Are they rare events or are they fairly common? It is estimated that our galaxy experiences one supernova every 40 years (Tamman 1982). The supernova that appeared 1885 in the Andromeda was the first one which was observed spectroscopically, because the technique was not developed before that time. About 50 years later the Swiss astronomer Fritz Zwicky and the German-American astronomer Walter Baade proposed the name supernovae (1934). Zwicky also proposed

*Hot Topics in Astrophysics 2001/2002, Alessandro B. Romeo & Eva S. Karlsson (Eds.), Chalmers University of Technology and Göteborg University, 2002.

that special observing patrols must be organized to look for supernovae. The idea behind this was that by systematically monitoring many hundreds of galaxies we might hope to observe one or two annually. The first supernova patrol was instituted in 1933 by Zwicky, using a very modest 10–inch telescope. During the period 1936 – 1939 he observed 12 supernovae in different galaxies. Zwicky reached the conclusion from his observations that a galaxy will experience on avarage one outburst every 360 years. The supernovae discovered were investigated both photometrically and spectroscopically. Lightcurves as well as spectra were obtained for these supernovae (see Shklovskii 1978). The spectra of one type were similar from event to event. The lightcurve (luminosity as a function of time) had a predictable regularity. It rose to peak brightness in about two weeks and then after declining for two more weeks, it faded in an almost exponential fashion. Other events were dimmer at maximum by a factor of a few in luminosity and had lightcurves which rose and faded in a manner that varied from supernova to supernova. The most important aspect of these events, which differentiated them as a separate class, was that the spectrum contained the common optical lines of hydrogen. These two different kinds of supernova were differentiated by Minkowski as Type I and Type II, respectively. Further differentiation of Type II are based on the shape of their lightcurve (see Petschek 1990):

- Sn II-P (Plateau)
- SN II-L (Linear)
- SN 1987A
- SN 1987K

Type I are subclassified according to features of their spectra:

- SN Ia (Si)
- SN Ib (no Si, He)
- SN Ic (no Si, no He)

2 Supernovae Type II

They are not seen in ellipticals and when they occur in spirals they tend to lie in the spiral arms, where massive starformation occurs. Because of this and that the Type II supernovae do have hydrogen in their spectra, their progenitors are belived to be massive stars which eventually collapse. It is belived that stars with masses less than about 8 $M_\odot$ shed their mantles and become white dwarfs, and that stars with a mass greater than 8 $M_\odot$ becomes supernovae Type II and leave neutron stars. The best studied examples are those who develop an iron core. The dense iron core, $\rho \sim 10^8\ g\ cm^{-3}$, can be approximated with a hot relativistic white dwarf where the pressure is dominated by $\gamma = \frac{4}{3}$ electrons, which is generated by the motions of the degenerate electrons. The iron core is incapable of generating energy and since iron has the highest binding energy of all the elements, it can only undergo endoergic reactions. The iron core, steadily being fed more and more mass will eventually reach the Chandrasekhar mass M_{CH} where the electron pressure no longer can resist the gravitational pressure. Since degenarate cores become unstable at roughly the same M_{CH}, the mass of the core is independent of the total mass of the star which could range from $\sim 8\ M_\odot$ to $\sim 60\ M_\odot$ (see Petschek 1990). The outer shells are important for nucleosynthetis and observational consequenses such as the total energy of the explosion and its lightcurve.

2.1 Core collapse

There are two processes that makes the collapse (see Petschek 1990):

Photodisintegration: $^{56}Fe \longrightarrow 13\alpha + 4n$

Electroncapture: $e^- + p \longrightarrow n + \nu_e$

The photodisintegration reactions take place when the center, as in the previous burning stages , contracts trying to ignite a new fuel, raising the temperature with only a small increase in pressure. This reaction is endothermic and costs $\sim 124 Mev$. This energy must come from somewhere, and it comes from the electrons. In this process 56 nucleons marching in step go on 17 different paths and thus the nucleonic entropy must increase. Since the total energy cannot change, because the core contraction is an adiabatic process, entropy must be transfered from the relativistic electrons to the nucleons. In this process no electrons are destroyed, but thermal energy is transfered to the nucleons (see Petschek 1990).

When the gravitational pressure increases, electrons and protons fuse together through the weak interactions to neutrons and neutrinos. When pressure support is lost through iron thermal disintegration and electron capture on protons, the core collapses on freefall timescales ($0.1s$). When the core shrinks to the point when the nuclei begin to touch each other, a uniform sea of nuclear matter begins to form and the repulsive part of the nuclear force comes into play. The near incompressible nature of uniform nuclear matter is responsible for the halting of the collapse. The collapsing sphere of neutrons halts and bounces at nuclear densities, ($\rho = 1.3 \cdot 10^{14}\ g\ cm^{-3}$). A promt hydrodynamical shock forms at the moment of core bounce. The shock starts at the radius whitch encloses about $0.7\ M_\odot$, exit the outer core and travel into the hydrogen-helium envelope. The shock must be energetic enough to overcome loss of energy to the thermal decomposition of iron as it passes through the shock wave. Colgate & White (see Janka et al. 2001; Colgate & White 1966) showed that the shock is not able to reach the outer layers. Wilson (see Janka et al. 2001; Wilson 1985) discovered that the stalled promt shock can be revived by neutrino heating in the dissociated postshock medium.

2.2 Delayed mechanism

While immediately after shock breakout neutrino emission extracts energy from the shocked gas , energy transfer from neutrinos to the postshock medium is favored hundreds of milliseconds later. If the shock has expanded to a larger radius and the postshock temperature has decreased, energetic neutrinos can deposit a small fraction of their energy in a gain layer behind the shock (see Janka et al. 2001; Bethe 1985). The neutrino heating mechanism can yield the energy of about $10^{51}\ ergs$ for the explosion of a massive star. 99 % of the energy is released to neutrinos and 1 % to kinetic energy. A neutron star with $15 - 20\ km$ radius is formed.The neutron pressure then prevents the star from becoming a black hole.

Hydrodynamical instabilities and mixing processes on large scales play an important role within the core as well as in the outer layers of the exploding star. Multidimensional calculations are therefore necessary to understand why and how supernovae explode. Spherically symmetric simulations, Newtonian and general relativistic, with the most advanced treatment of neutrino transport by solving the Boltzmann equation, do not produce explosions (see Janka et al. 2001).

3 Supernovae Type Ia

Type Ia are seen in all galaxies and, in spirals, they tend to lie between spiral arms or in the halo where we find the older stars. They do not show any evidence of hydrogen in their spectra. So the progenitor of Type Ia supernovae are belived to be a white dwarf composed of carbon and oxygen, near the Chandrasekhar mass, that accretes material from a binary companion, reaching M_{CH} and exploding when the burning of carbon triggers the runaway.

3.1 Sub M_{CH} helium ignitors

A carbon-oxygen white dwarf accumulates a helium layer. Detonations in the accreted helium layer were suggested to drive a strong enough shock into the C-O core to initiate a carbon detonation. This might explain the class of subluminous explosion but these models disagree both photometrically and spectroscopically with observations (see Niemyer et al. 2002).

3.2 Merging white dwarfs

If the accretion rate of C+O onto the remaining white dwarf is larger than a few times $10^{-6}\ M_{\odot}\ yr^{-1}$, the most likely outcome is off-center carbon ignition, leading to an inward propagating flame that converts the star into O+Ne+Mg. This configuration is unstable to electron capture onto ^{24}Mg and will undergo accretion-induced collapse to form a neutron star. Dimensional analysis of the expected turbulent viscosity suggest that it is very difficult to avoid such high accretion rates. This model can explain the absense of H and He in spectra (see Niemyer et al. 2002).

3.3 Favored model

The favored model is a white dwarf that accretes matter from a binary companion and explodes.

The strong temperature dependence of the nuclear reaction rates $\dot{S} \sim T^{12}$ as $T \simeq 10^{10}\ K$ (see Niemyer et al. 2002), confines the nuclear burning to microscopically thin layers that propagate either conductively as subsonic deflagration (flames) or by shock compression as supersonic detonations. Both modes are hydrodynamically unstable to spatial perturbations. In the nonlinear regime, the burning fronts are either stabilized by forming a cellular structure or become fully turbulent, the total burning rate increases as a result of flame surface growth.

The first hydrodymamical simulation of an M_{CH} white dwarf (see Niemeyer 2002; Arnett 1969) assumed that thermonuclear combustion commences as a detonation wave, consuming the entire star at the speed of sound. Given no time to expand prior to being burned the C+O material in this scenario is transformed almost completely into iron-peaked nuclei and thus fails to produce intermediate mass elements, in contradiction to observations. It is for this reason that the explosions are belived to begin in deflagration mode. The explosion starts as a deflagration near the center of the star. Rayleigh-Taylor unstable blobs of hot burnt material are thought to rise and lead to shear-induced turbulence at their interface with the unburnt gas. This turbulence increases the effective area of the flamelets and, thereby the rate of fuel consumption (see Hillebrandt et al. 2000). Agreement with observations is best achived if the burning front propagates at a substantial fraction of the sound speed. A slower speed does not make a supernova. Best agreement is achived if a little over 1/2 of the mass of the white dwarf is turned into iron before expansion gradually quenches the burning.

Neither flames nor detonations can be resolved in explosion simulations on stellar scale and therefore have to be represented by numerical models. The length scales of physical processes range from 10^{-3} *cm* – 10^{7} *cm* (see Hillebrandt et al. 2000).

Conclusions

Supernova explosions of massive stars are important for applying nuclear and particle physics, in particular neutrino physics. The neutrino-heating mechanism, the favored explanation for the explosion, is still controversial. Definitive answers concerning the progenitors of Type Ia can be obtained from observations in: x-rays, radio, and high resolution optical spectroscopy.

References

Arnett W.D., 1969, Astrophys. Space Sci. 5, 168

Bethe H.A., Wilson J.R., 1985, Astrophys. J. 295, 14

Colgate S.A, White R.H., 1966, Astrophys. J. 143, 626

Hillebrandt W., Niemeyer J.C., Reinecke R., 2000, arXiv:astro-ph/0005584

Janka H.T., Konstantinos K., Rampp M., 2001, arXiv:astro-ph/0103015

Livio M., 2000, arXiv:astro-ph/0005344

Niemeyer J.C., 1998, arXiv:astro-ph/9802182

Niemeyer J.C., Reinecke M., Hillebrandt W., 2002, arXiv:astro-ph/0203369

Shklovskii I.S., 1978, Stars: Their birth, life, and death. W.H. Freeman and Company, San Francisco

Petschek A.G., 1990, Supernovae. Springer Verlag, New York

Wilson J.R., 1985, Supernovae and Post-Collapse Behavior. In: Numerical Astrophysics, ed. by J.M. Centrella, J.M. LeBlanc, R.L. Bowers, J.A. Wheeler (Jones and Bartlett, Boston) p. 422

Gravastars:
Can Anything Be Stranger than a Black Hole?

Bengt Tegner

Göteborg University
SE-41296 Göteborg, Sweden
(gu97bete@dd.chalmers.se)

*

Abstract

A possible new solution for the endpoint of the collapse of a massive star is reviewed as an alternative to black holes. This new object, called a *gravastar* solves some of the paradoxes plaguing black holes.

1 Introduction

In 1783, Reverend John Michell noted that if one assumed a corpuscular theory of light with a finite velocity combined with Newton's theory of gravity one could produce 'dark stars', that is, stars so massive light would not be able to escape their gravitational pull. Michell calculated a critical radius r_{crit} determined by the mass of the star. The derivation is quite straightforward. The potential energy of Newton's gravity for a point mass is given by

$$U = -\frac{GMm}{r} \tag{1}$$

and the kinetic energy is given by

$$T = \frac{1}{2}mv^2. \tag{2}$$

Setting the energies equal, we arrive at the formula for escape velocity

$$v_{esc} = \sqrt{\frac{2GM}{r}}. \tag{3}$$

Now, setting v_{esc} equal to the speed of light c, we finally get the formula for r_{crit}, given by

$$r_{crit} = \frac{2GM}{c^2}, \tag{4}$$

where c is the speed of light and M is the mass of the star. The formula for kinetic energy, however, is not correct for photons. Michell did not know that.

*Hot Topics in Astrophysics 2001/2002, Alessandro B. Romeo & Eva S. Karlsson (Eds.), Chalmers University of Technology and Göteborg University, 2002.

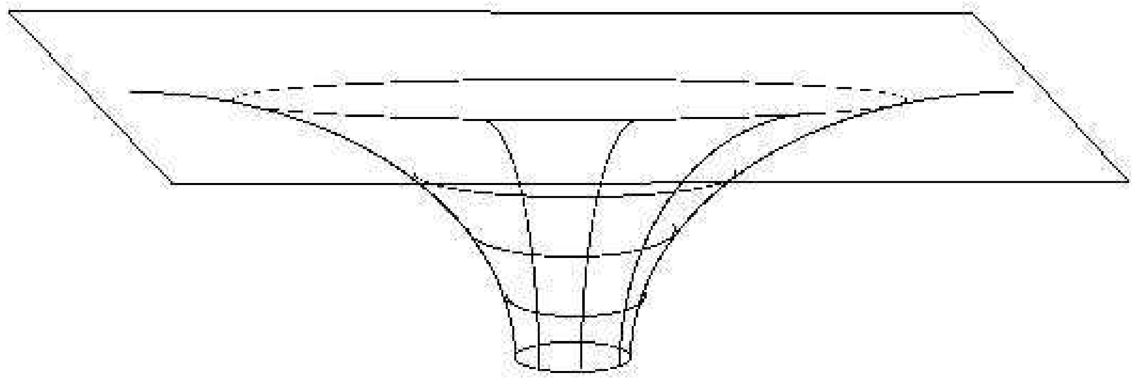


Figure 1: Embedding diagram of a black hole. This figure shows the geometry of space in the equatorial plane ($\theta = \pi/2$). The 'throat' represents the Schwarzschild radius.

2 General Relativity

Albert Einstein presented his General Theory of Relativity in 1915, which describes gravity as a curvature of space-time due to the presence of energy and matter. The Einstein Field Equations look like the following

$$G_{\mu\nu} = 8\pi G T_{\mu\nu} \tag{5}$$

in symbolic (tensor) form, where the left hand side describes the curvature of space-time and the right hand side describes the distribution of energy and matter. In reality Einstein's field equations are quite horrible. They are 10 coupled, non-linear partial differential equations. The corresponding equation in Newton's theory is

$$\nabla^2 \phi = 4\pi G \rho \tag{6}$$

where the left hand side describes the derivatives of the gravitational field and the right hand side describes the distribution of matter.

2.1 The Schwarzschild solution

In 1916 the German astronomer Karl Schwarzschild produced the first solutions to Einstein's Field Equations (Eq. (5)) for the space-time around a static, spherically symmetric star of mass M. They can be written as

$$ds^2 = -dt^2\left(1 - \frac{2GM}{r}\right) + dr^2\left(1 - \frac{2GM}{r}\right)^{-1} + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right), \tag{7}$$

in units where $c = 1$, with the signature $(-1, 1, 1, 1)$, which will be used throughout this paper. The expression above is the so called *line element* which describes the invariant distance in space-time between two events. One important consequence is that the solution (the metric) is singular at the radius $r = 2GM$. This radius became known as the Schwarzschild radius r_S, given by

$$r_S = \frac{2GM}{c^2}, \tag{8}$$

in ordinary units. Interestingly, this is the same expression as Eq. (4) above. Moreover, the singularity at $r = 2GM/c^2$ can be removed by a suitable change of coordinates, since space-time is well-behaved at that radius. The "true" singularity lies at $r = 0$, where the solution breaks down. However, the Schwarzschild radius also defines the *event horizon*,where no information about the events inside can escape to the outside world. But it should be stressed that there is nothing *particular* about the Schwarzschild radius apart from this fact. One way to visualize the Schwarzschild geometry is to use what is called an embedding diagram. An example is shown in figure 1. The radial coordinate goes outward and the azimuthal angle goes counter-clockwise, as usual. The 'throat' represents the Schwarzschild radius.

3 Entropy

Entropy can be seen as a measure of the amount of disorder within a system. The modern view of entropy was introduced by the Austrian physicist Ludwig Boltzmann and can be expressed as

$$S = k_B \log_e W, \tag{9}$$

where W is the total number of micro states of the system. In this view, a disordered system has a large number of micro states and therefore a large entropy. The constant k_B in front of the logarithm is of course called Boltzmann's constant. It has the value of 1.38×10^{-23} J/K.

The second law of thermodynamics which governs the change of entropy within a system states that

$$\mathrm{d}S \geq 0,$$

i.e. the entropy must increase or remain constant.

4 Black Holes

The name 'black holes' was coined in 1968 by John Wheeler. They where previously called "Schwarzschild singularities", a somewhat complicated name. Einstein himself didn't believe they could exist. However, X-ray observations provided the first clues that Einstein might be wrong by detecting several bright X-ray sources in the sky. The generation of these X-rays where thought to be gravitational accretion onto a black hole. One famous example is the X-ray source Cygnus X-1. (See Thorne 1994).

4.1 Hawking Radiation

Stephen Hawking discovered in 1974 that black holes are not completely black, but can radiate with a Planck spectrum corresponding to a temperature given by (Hawking 1974)

$$T_H = \frac{\hbar c^3}{8\pi k_B G M}, \tag{10}$$

with the associated entropy (Bekenstein 1973)

$$S_{BH} = \frac{4\pi k_B G M^2}{\hbar c}, \tag{11}$$

for a Schwarzschild black hole. This entropy is simply enormous, of the order

$$S_{BH} \approx 10^{77} K_B (M/M_\odot)^2.$$

This entropy increases as the black hole swallows matter so the second law of thermodynamics is obeyed but in is puzzling how a black hole can be both very simple yet have such an enormous entropy, which would indicate that it is also very complex.

4.2 Black Hole Information Paradox

The thermal spectrum of the Hawking radiation from a black hole does not reveal any details of the matter falling past the event horizon into the black hole. Information concerning the matter seems to be destroyed. This is a violation of information theory which states that information can not be destroyed. A black hole can be completely described by its mass M, angular momentum J and charge Q. Nothing more! This is indeed very puzzling. Two answers to this question can be found. One is that the information indeed is lost to us. The other answer is that the information somehow is encoded in the Hawking radiation.

5 Gravastars

The gravastar is a kind of "bubble" consisting of a thin shell surrounding a strange kind of springy vacuum. It has the following properties:

- No singularity
- No horizon
- No information paradox
- Ultra-cold and completely dark

This new solution was proposed by Pavel Mazur and Emil Mottola (Mazur and Mottola 2002) as an alternative endpoint to the collapse of a massive star to massive to become a neutron star.

5.1 The New Solution

The gravastar solution assumes that gravity undergoes a vacuum rearrangement phase near $r = r_S$ to form a new kind of gravitational Bose-Einstein condensate.

A Bose-Einstein condensate is a property of bosons, which are particles who obey Bose-Einstein statistics. Protons and electrons are fermions, which means they obey Fermi-Dirac statistics and the Pauli exclusion principle. This means that if you constrict the the number of states a system of fermions can occupy, (by lowering the temperature, for example) the fermions will resist being squeezed together to tightly.

Bosons, however, to have this problem. When the temperature is low enough all the bosons will condense into a single state. This means all the particles can be described a a single particle, using a single wave-function ψ. The bosons behave as a kind of 'super atom'. This is a Bose-Einstein condensate. The experimental realization of this was awarded the Nobel Prize in Physics in 2001.

In order to solve Einstein's Field Equations, we begin by defining the equations of state for the gravastar.

$$\text{I. Interior: } 0 \le r \le r_1, \quad \rho = -p, \tag{12}$$
$$\text{II. Shell: } r_1 \le r \le r_2, \quad \rho = +p, \tag{13}$$
$$\text{III. Exterior: } r_2 \le r, \quad \rho = p = 0. \tag{14}$$

Density in the interior region is a constant due to conservation. Let us call this constant $p_v = 3H_0^2/8\pi G$. This, along with the requirement of no mass singularity at $r = 0$ yields the following metrics (Mazur and Mottola 2002)

$$\text{I. Interior: } ds^2 = -dt^2\left(C\left(1 - H_0^2 r^2\right)\right) + dr^2\left(1 - H_0^2 r^2\right)^{-1} + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right), \tag{15}$$
$$\text{III. Exterior: } ds^2 = -dt^2\left(1 - \frac{2GM}{r}\right) + dr^2\left(1 - \frac{2GM}{r}\right)^{-1} + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right). \tag{16}$$

in units where $c = 1$. The integration constants C, H_0 and M are arbitrary. See (Mazur and Mottola 2002) for details including a solution for region II. Also, a suggestion on a theory producing this interior and the $p = \rho$ shell can be found in (Mazur and Mottola 2001). An embedding diagram of the exterior solution would look exactly like figure 1. The interior solution looks like a four-dimensional hyperboloid embedded in a five-dimensional space.

5.2 Gravastar Entropy

The entropy of a solar mass gravastar S_{GS} is of the order

$$S_{GS} \approx 10^{38} k_B(\ell/L_{Pl}),$$

where ℓ is the thickness of the shell, which is comparable to L_{Pl}. This entropy is some 38 orders of magnitude smaller that the Bekenstein-Hawking entropy for the same mass. It is also some 20 orders of magnitude than a typical stellar progenitor, which have an entropy around $10^{58} k_B$ for $M/m_N \sim 10^{57}$ nucleons.

5.3 Gravastar Creation

A violent shedding of entropy, such as a supernova explosion, is needed to produce a gravastar. Details of the phase transition and the collapse process are not known at present.

6 Observations: Black Holes versus Gravastars

Gravastars might be distinguishable from black holes observationally in the following ways:

- The stiff shell could produce shock fronts when struck by in-falling matter.
- Gravitational waves from a stuck gravastar should reveal its fundamental modes of vibration.

In most other aspects, the gravastar would look exactly like a black hole of the same mass.

7 Conclusions

A collapsing star to massive to become a neutron star may become a black hole or a gravastar. Further observational and theoretical investigations are needed to settle this issue. However, the maximally stiff shell with $p = \rho$ requires the creation of some very exotic matter. Furthermore, it is not clear if the gravastar is dynamically stable.

Acknowledgements

I wish to take this opportunity to express my thanks to Alessandro Romeo for his encouragement and to my fellow students for their support.

References

Bekenstein J.D., 1973, Phys. Rev. D 7, 2333

Hawking S.W., 1974, Nature 248, 30

Mazur P.O. and Mottola E., 2002, arXiv:gr-qc/0109035

Mazur P.O. and Mottola E., 2001, Phys. Rev. D 64, 104022

Thorne K.S., 1994, Black Holes and Time Warps: Einstein's Outrageous Legacy. W.W. Norton & Co., New York

Large-Scale Structure of the Universe: the Formation of Galaxies

Jens Andersson

Chalmers University of Technology
SE-41296 Göteborg, Sweden
(f98jean@dd.chalmers.se)

*

Abstract

In this paper I give a review of the current research on galaxy formation from a cosmological perspective. There has been two competing theories explaining the formation of galaxies for quite some time. Measurements and simulations now seem to point in the same direction, namely that large scale structure form by hierarchical clustering. This favors the ΛCDM model for galaxy formation.

1 Introduction

We seem to have some kind of understanding in how stars form and evolve (at least in the present universe). The Big Bang theory is also widely accepted and has produced the correct amounts of light elements according to observations, thus the birth of the universe is at least described. But how do the structures between the entire universe and separate stars form?

The understanding of formation of galaxies have been called the holy grail of cosmology. The research on the subject has been going on since the 50-60's and a lot of theories have been proposed, and some of them rejected. Observational techniques have improved, and since then, have given support to some theories, but falsifying others. From the first tries of understanding galaxy formation by collapse of spherical systems (Sandage et al, 1962) to todays advanced numerical simulations using semianalytical galaxy formation in a universe filled with dark matter (Cole et al, 1994), astrophysicists have been trying to understand the complex cosmology and physics that lie behind the birth of galaxies, clusters and superclusters.

In this paper I will give an introduction to cosmology to familiarize the reader with the concepts used in theories of galaxy formation. I will then concentrate upon describing the various theories, also introducing dark matter as a solution to some of the problems arising when trying to model the process. Current research in the field is of both observational and theoretical character, where the link between the two is given by the numerical

*Hot Topics in Astrophysics 2001/2002, Alessandro B. Romeo & Eva S. Karlsson (Eds.), Chalmers University of Technology and Göteborg University, 2002.

simulations performed on supercomputers. The current research have been excellently summarized by various authors (Ellis, 2001; Silk, 2002), this paper gives a review of the research, and should be readable by the interested layman, as well as by the professional astronomer.

The results from various observations and simulations seem to point to a solution of the problem, but there are still problems that need to be solved before we can say that the secrets behind galaxy formation have been unraveled.

2 The early universe, before the birth of galaxies

Much of the following material have been influenced by the text book in modern cosmology, *Galaxy Formation* by Malcolm S. Longair, 1998. If you have a good understanding of basic cosmology, you may skip this section and go on to section 3, A first approach to galaxy formation.

2.1 Timeline of the early universe

In Big Bang theory, the universe comes into existence at some distinct time, about 15 billion years ago (the exact figure is not yet determined). In the beginning the universe was extremely small and dense, all the energy in the universe was contained in an infinitesimal volume. The very first 10^{-43} seconds is called the Planck time (t_p), we don't really know anything about this era, since the physics we use is not applicable at the extreme energies found in the infinitesimal universe. Then for some reason the universe exploded and started to expand. It is important to realize that the universe didn't start to expand into a surrounding medium, the universe really is everything that exists. Figure 1 illustrates some of the important events in the early universe

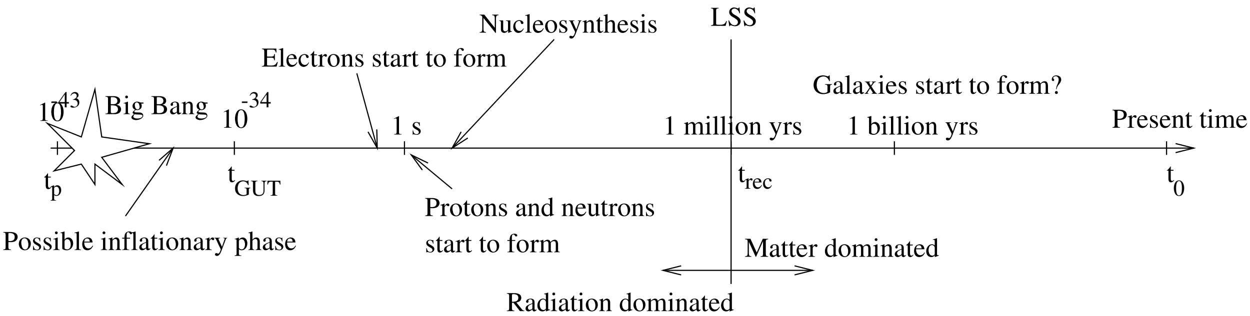


Figure 1: Timeline of the history of the Universe (t_0 denotes the present time)

In the Planck era, the energy density (or the temperature) of the universe was extremely high and all the four fundamental forces (gravity, electromagnetism, the weak and the strong nuclear forces) were all coupled to each other. Physics was then determined by one single force (called quantum gravity). As the universe started to expand the energy density decreased and gravity decoupled from the other three forces. This is the GUT (Grand Unified Theory) era, quantum field theory can be used to describe this period. During this epoch the universe might have had a period of exponential expansion (inflation). At $t_{GUT} \sim 10^{-34}s$ the energy had decreased to the level where the strong nuclear force decoupled from the elektroweak force.

In an universe with high energy density matter and radiation is really the same thing, since photons can become matter and the other way around. At some time after t_{GUT}

matter started to decouple from the radiation and electrons, protons and neutrons started to form, also building up the light elements (this is roughly at the time $t \sim 1s$, where the temperature was low enough for the reactions to take place). However since the radiation was very energetic the dynamics of the universe was entirely radiation-dominated. The expansion continues and after quite some time (which I will call t_{rec} for reasons which will soon be apparent) the temperature had decreased to values where the universe became matter-dominated. This happens after about one million years.

At about the same time electrons started to combine with protons to form hydrogen (the energy density had decreased below the ionization potential of hydrogen), this is called the recombination era (hence t_{rec}). When trying to observe the early universe we run into trouble when trying to look back through the recombination era. This is because of something that is called the Last Scattering Surface (LSS). Before recombination the electrons scattered the photons very efficiently (by Thomson scattering) and therefore we can't see through the LSS.

2.2 Cosmological concepts needed to model galaxy formation

Some basic knowledge of concepts used in cosmology is needed to understand the following sections. The first is one is the Hubble constant, H_0. Galaxies far away move away from us, the further away, the faster they move. This is due to the expansion of the universe, H_0 defines the current expansion rate of the universe:

$$v = H_0 D = \left(\frac{\dot{R}}{R}\right)_{t_0}, \tag{1}$$

where D is the proper distance to the galaxy. R is the scalefactor of the universe and is a measure on the size of the universe. Usually $R(t_0) = R_0$ is defined to be one (t_0 of course denoting present time).

The general theory of relativity is the theory that lies behind the cosmology. From general relativity comes the concept of time dilation. A time interval in the dense early universe is observed to be longer in our reference frame than in its own. This means that a photon emitted at some earlier time is *redshifted* when it reaches us. This is in fact what we measure when we look at the velocities of distant galaxies. The redshift, z, is related to the scale factor and the observed/emitted wavelengths (λ_0/λ_e) by the following equation:

$$1 + z = \frac{1}{R(t)} = \frac{\lambda_0}{\lambda_e}, \tag{2}$$

so if we observe photons emitted at time t, they will be redshifted by an amount z when we observe them.

If we assume that particles can only interact by causal communication (they can't communicate faster than light), we get the so called *particle horizon*, r_H.

$$r_H = 3ct, \tag{3}$$

where the factor 3 comes from the fact that the particles were closer to each other in the past. The particle horizon sets a limit to how large structures can be and still interact. Structures that are larger than the horizon is therefore stable against gravitational collapse.

2.3 Quantum fluctuations: the seed for galaxies?

When doing cosmology a couple of basic assumptions are made to get a theory. One of those is the assumption that the universe is isotropic and homogeneous on large scales. This means that the amount of energy (both radiation and matter) is evenly distributed on large scales, this is also confirmed by observations. Thus the early universe must be isotropic. Where do the structure we observe on smaller scale come from? Figure 2 shows the anisotropy of the cosmic microwave background. In the pre-recombination

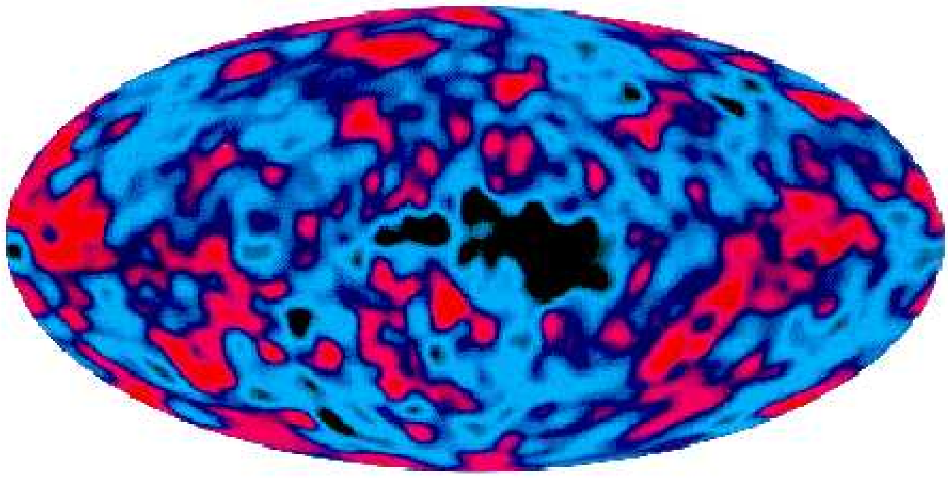


Figure 2: Anisotropy of CMBR, the red (dark) parts are hotter than the blue (pale) (courtesy of COBE, NASA)

universe small instabilities form in the otherwise isotropic radiation, perhaps as quantum fluctuations, perhaps they were remains of processes in the planck era. The spectrum of fluctuations, if there are more fluctuations on the large scale than on the small, will have a large influence on the formation of structure. However both theory and observations point to that fluctuations have the same probability at all scales, the universe is *self-similar*.

When the particle horizon has grown to encompass the entire fluctuation things can start to happen, this will be described in the next section. We will see that the small fluctuations will in fact be responsible for the formation of all large scale structure, in one way or the other.

3 A first approach to galaxy formation

3.1 Growth of instabilities and gravitational collapse

The first thing we need to understand is how the fluctuations can become so dense that stars and galaxies can form. A perturbation will collapse if the gravitational forces that try to contract it are greater than the outwards pressure. When collapsing the density of the perturbation will increase a lot. Assume that the perturbation has a density $\rho = \rho_0 + \delta\rho$ and that the mean density of the universe is given by ρ_0. We can then derive a wave equation for the density contrast of the perturbation, assuming that the perturbations are adiabatic. Let us define $\Delta = \delta\rho/\rho_0$ as the density contrast. The wave equation then becomes (see Longair 1998 for a complete derivation):

$$\frac{d^2\Delta}{dt^2} + 2\left(\frac{\dot{R}}{R}\right)\frac{d\Delta}{dt} = \Delta(4\pi G\rho_0 - k^2 c_s^2), \tag{4}$$

where G is Newtons constant of gravity, $k = 2\pi/\lambda$ is a wave vector for a perturbation of the scale λ and c_s is the soundspeed of the medium. If the right hand side of the wave equation (4) is positive then the perturbations are oscillatory and will keep their current density. But if it is negative we have an unstable situation and the density contrast will start to grow exponentially. This happens if the scale of perturbation is larger than the critical *Jeans' wavelength*, λ_J defined by:

$$\lambda_J = c_s \left(\frac{\pi}{G\rho}\right)^{1/2} \tag{5}$$

This unstable situation leads to gravitational collapse, high-density regions can be created thus enabling stars to be born. When combining the wave equation with the fact that the universe is expanding we find that the collapse of the initial fluctuations is not exponential, but linear in the scale factor (the expansions halts the collapse somewhat), and that gravitational collapse cannot begin until the particle horizon has grown to encompass the perturbation.

The *Jeans' mass*, M_J is defined to be the mass contained in a region of diameter λ_J. This mass then sets a limit on how small masses that will collapse. Perturbations with masses less than M_J will not collapse.

The first theories of gravitational collapse of large structure into gravitationally bound objects was given by Sandage et al, 1962. They measured the radial velocity of old stars in the Galaxy, and showed the first stars seemed to have been formed during gravitational collapse. They then drew the conclusion that elliptical galaxies might have formed in this way. This simplified picture of galaxy formation is called *monolithic collapse.*

3.2 The first theories: the adiabatic and isothermal models

In the analysis of the previous section we assumed that the perturbations were adiabatic, which means the density perturbations are coupled to the pressure by an adiabatic relation (this relation is different for radiation and matter). In the matter dominated universe we find that the Jeans' mass before recombination is roughly:

$$M_J \sim 10^{15} M_\odot \tag{6}$$

So perturbations with masses higher than this collapses gravitationally, perturbations of lesser mass are stable. This mass corresponds to masses of galaxy clusters. The Jeans' mass changes at the recombination era due to the fact that the outwards pressure is no longer given by radiation, but by thermal pressure of matter. It then becomes:

$$M_J \sim 10^5 M_\odot. \tag{7}$$

Perturbations of mass greater than this will undergo gravitational collapse and form bound systems. This mass is roughly the size of globular clusters.

We also have to worry about the non-growing perturbations before recombination. They might be damped out before the Jeans' mass decreases after recombination. If we take into account that radiation can diffuse out of the perturbation we find that this is indeed the case. The stable oscillatory solution of the wave equation (4) is damped out during recombination (so called Silk damping). This means that the only structure left after recombination is the large scale structure corresponding to the Jeans' mass in this model. Then we have to accept that small scale structure must form from the large scale, yielding a *top-down* scheme of galaxy formation.

The perturbations might not be adiabatic, they can also be isothermal fluctuations in the matter density. This means that they cause no fluctuations in the background radiation temperature. The growth of isothermal perturbations are not hindered by pressure and fluctuations on all scales should collapse. In reality there are effects that hinder the collapse before recombination, but after recombination we have that, in the isothermal model, perturbations on all scales are left and can start to collapse. This yields a *bottom-up* scheme for the formation of structure. Galaxies are then built up in a process called *hierarchical clustering*, the theory behind this can be found in an important paper by Press and Schechter, 1974.

3.3 Why do these theories fail?

Unfortunately both these theories fail. One of the problems is that if we want inflation to work, the universe has to be flat (geometry of space-time). If the Big Bang theory is correct the matter (or baryons) produced (and measured by abundance ratios) are too few to make the universe flat. One way of circumventing this problem is by introducing a non-baryonic component of matter, that we call dark matter. If we have dark matter in the universe the process described above is not correct.

Another problem for the theories is that in both theories structure on the scale of galaxy clusters and up begins to collapse after recombination. If the matter is purely baryonic these structures should leave big imprints on the CMBR. In measurements of CMBR we find that the the observed fluctuations is smaller than what is needed for the adiabatic and isothermal non-dark matter models.

4 Current models of galaxy formation

4.1 Dark matter

The theories without a dark matter component fail to produce the observed CMBR map. To solve the problems a non-interacting component of matter was introduced, the so called dark matter. The term dark matter normally includes some baryonic matter like brown dwarfs and black holes, but in this paper I will use the term to denote the non-baryonic part. Baryons are normal matter like protons and neutrons. Dark matter can be divided into two parts considering the effects on formation of structure, Hot dark matter and Cold dark matter. What the dark matter is comprised of is treated in a lot of papers, one good review is given by Moore (1999).

The hot dark matter is thought to be in the form of neutrinos from the Big Bang. They decoupled from the thermal equilibrium sometime before recombination when they were still relativistic. If the neutrinos have a mass of $\sim$ 10 eV they will be massive enough to make the universe flat. Cold dark matter can be in the form of exotic particles that have not been observed yet like WIMP's (Weakly Interacting Massive Particles) or axions. These particles didn't decouple from the equilibrium until after they became non-relativistic.

4.2 The Hot Dark Matter model (HDM)

In this model it is assumed that the dark matter is in the form of neutrinos with a rest mass of 10 eV. In the very early universe there were perturbations in the neutrino densities on all scales (as for the radiation and matter). These perturbations grew in the same way as baryonic perturbation (see Sect. 3.1) but small perturbations were damped out because of neutrino diffusion. At the recombination phase the perturbations left were those with masses greater than $\sim 10^{15} M_{\odot}$. After recombination the baryonic matter collapsed into the neutrino perturbations and produced large structures. These large structures then fragmented into smaller scales like galaxies and clusters. HDM thus yields a top-down scenario of the formation.

The problem with the HDM model is that galaxies form late in this scenario, fragmenting out of the larger structures. This causes problems with how the early interstellar gas was heated and how heavy elements could have been formed at early times (Longair, 1998). Another problem is that neutrinos might not be as massive as needed. In recent observations of the Sun, Ahmad et al (2000) have measured the neutrino mass to be between 0.05-8.4 eV, more observations will hopefully yield a more exact value.

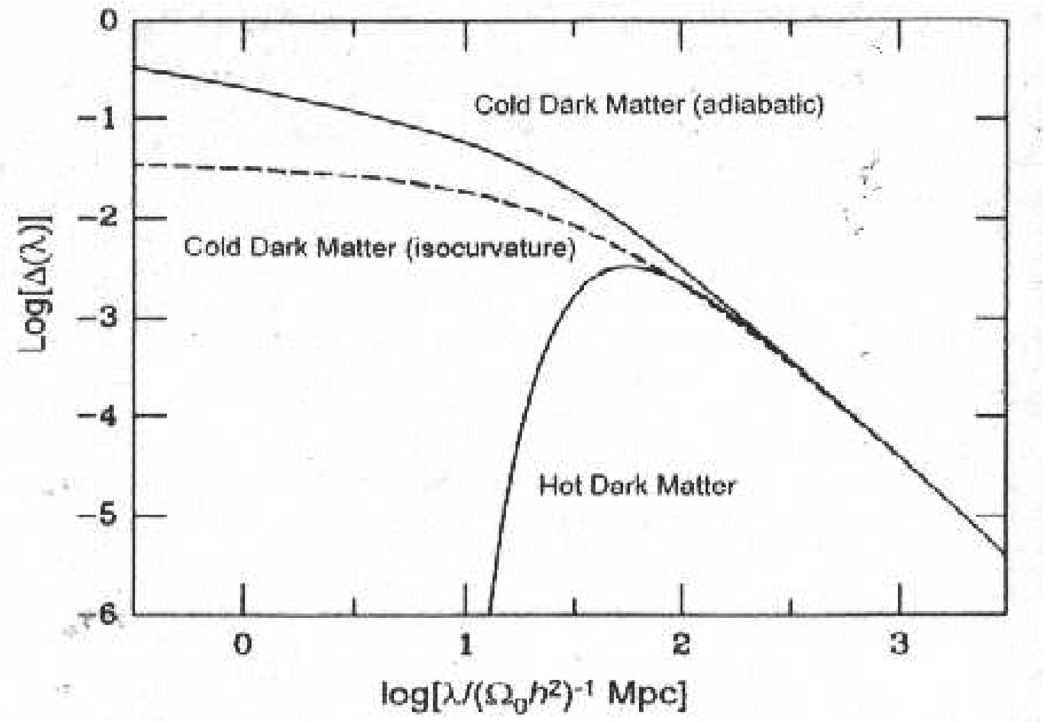


Figure 3: Spectrum of fluctuations for different models, λ is the scale of the perturbation. (From Kolb and Turner, 1990)

4.3 The Cold Dark Matter model(CDM)

The difference between HDM and CDM is that the cold dark matter decoupled when it had became non-relativistic. Non-relativistic particles couldn't diffuse out of the perturbations, so the small perturbations were not damped out. After decoupling from the thermal equilibrium the dark matter perturbations started to grow independently of the radiation and matter fluctuations. The baryonic perturbations were damped by Silk damping, thus leaving only large scale structure.

The presence of dark matter also affects the way instabilities grow in the normal matter because the density in the derivation of Jeans' mass includes dark matter. This causes the perturbations in the baryon and radiation plasma to become smaller, thus solving the problem of too large fluctuations in the CMBR. The dark matter perturbations were

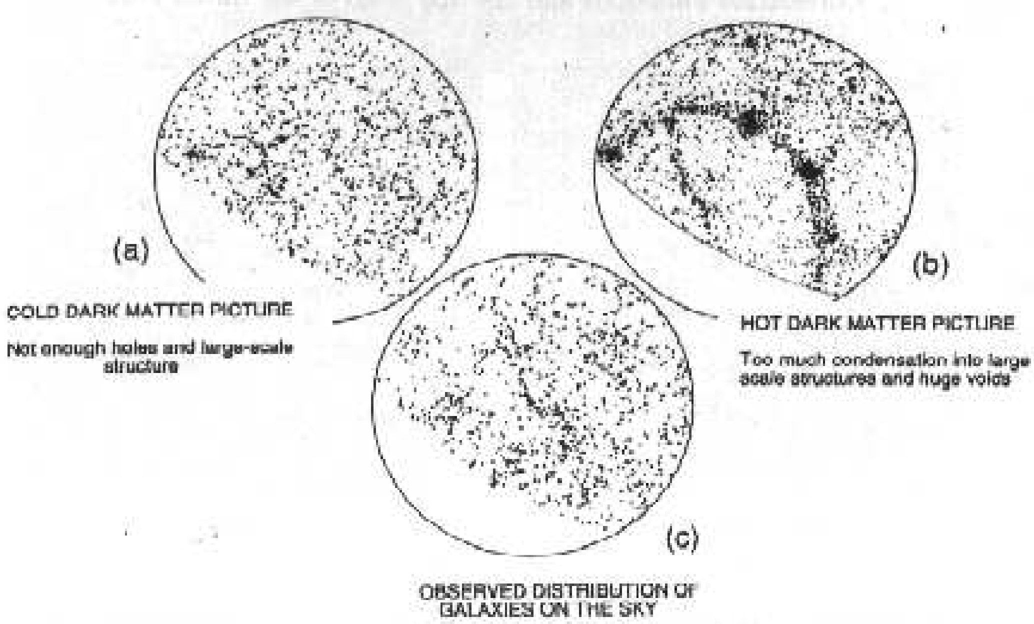


Figure 4: A comparison of the structure produced by the models in simulations (a and b) with the observed structure (c). (From Longair, 1998)

left on all scales after recombination. The baryonic matter now started to interact with the dark matter structures, collapsing into them. The effective Jeans' mass by that time for the baryons is given by equation 7. So baryons are drawn into the high dark matter densities, and when they become more massive than the Jeans' mass they collapse and form bound gravitational system (see for example Cole et al, 1994 or Sandage et al, 1962). The structures formed have masses of the same order as the Jeans' mass. This mass is roughly the mass of globular clusters.

In the CDM model larger scale structure is then built up by the process of *hierarchical clustering* (see Press and Schechter, 1974). This process causes structure to grow when the separate particles collide and stick together to form larger particles. In the advanced models we have to take into account how the dynamics of the dark matter halos surrounding the protogalaxies interact with the gas (Cole et al, 1994). The gas is cooled and stars start to form. The formation of stars then influence how matter and radiation is distributed in the galaxies. Different processes can form disk galaxies and spherical systems (Silk, 2002). In the hierarchical clustering picture merging of galaxies is also important. Some galaxies are created as mergers of two galaxies, some (or all) of the elliptical galaxies might have been formed in this way. CDM thus produces a bottom-up model of galaxy formation. Large scale structure is produced relatively late (at low z) in the model. This was the main problem of CDM models, as seen in figure 4. In figure 3 the spectrum of fluctuations after recombination is shown for the CDM and HDM theories.

This problem is solved by modifying the CDM model, but without altering the central idea. One of these models is the ΛCDM model with a non-zero cosmological constant (a term in the relativistic equations describing the expansion of universe that accelerates the expansion), which stretches time so that more large scale structure has time to be built.

This model is of great interest, since other observations seem to give the cosmological constant a non-zero value (see Fig. 5).

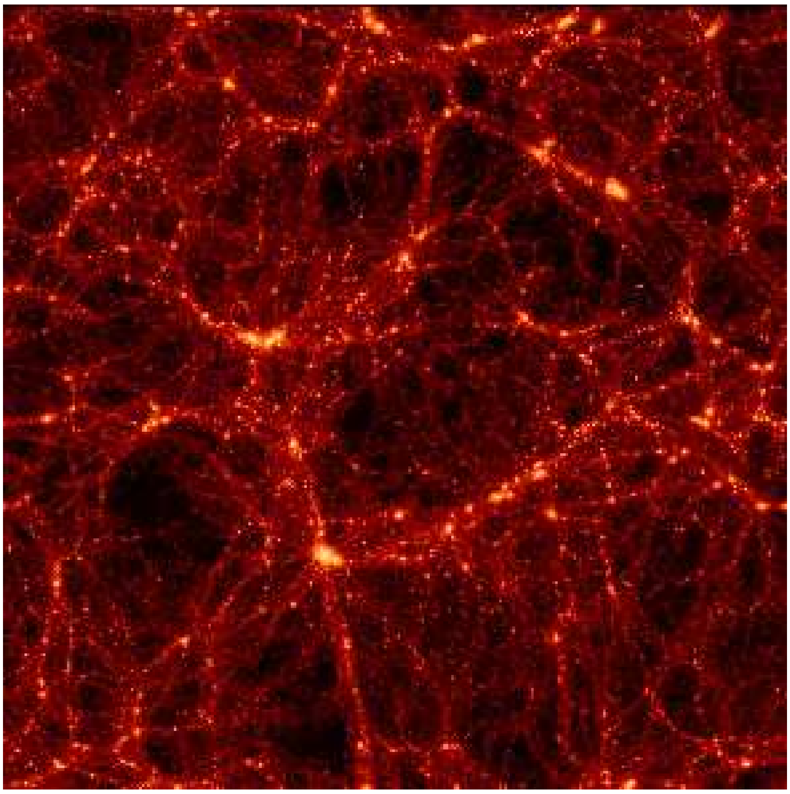


Figure 5: N-body simulation of a ΛCDM universe. (From http://www.nbody.net)

4.4 Numerical methods

So how can we check our theories? We can't observe the early universe in detail, the sources are too faint for observing internal dynamics, also the LSS sets a limit to how far back we can observe. So instead we use the theories to construct numerical simulations where the early times of the universe is played out on a supercomputer. For these simulations to be good we need to include some detail (but not too much) of the physics behind formation (Cole et al, 1994). We need to model how the dark matter perturbations collapse and merge, but also how it couples to the baryonic gas through shocks, cooling and heating. To be able to produce the observations we need to put in theories of stellar evolution into our numerical model. Finally we need to include the effects of mergers and gravitational interaction between systems.

One numerical method is Semianalytical galaxy formation, first introduced by Cole et al, 1994. They use Monte Carlo simulations to model the above processes for density perturbations. They obtain characteristics of the model galaxies which are observable, so their theory can be tested. Another simulational technique is N-body simulations where a spectrum of perturbations is given as initial condition, and the program lets them evolve (see Fig. 6).

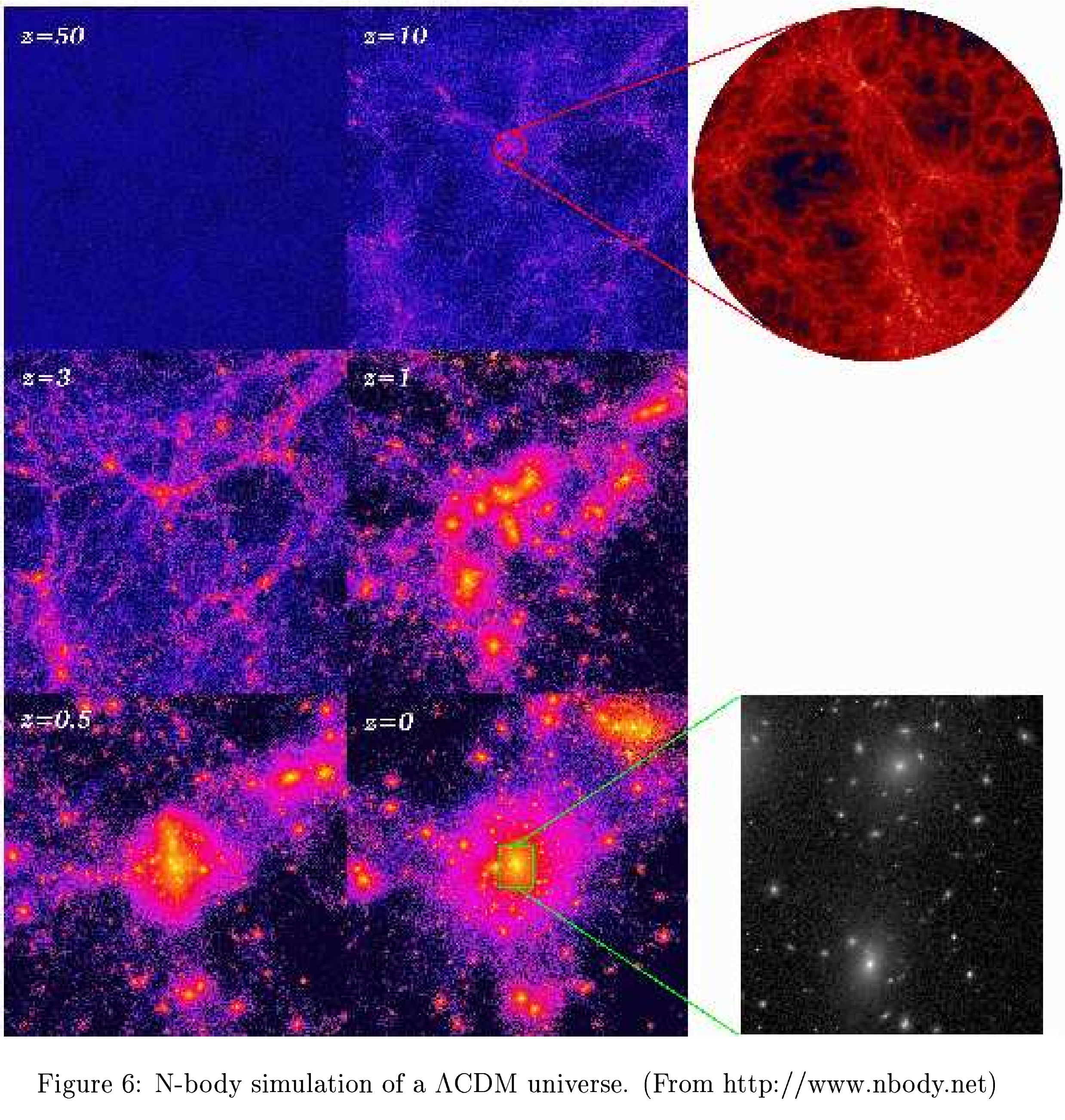


Figure 6: N-body simulation of a ΛCDM universe. (From http://www.nbody.net)

5 Observations of the early universe

In order to get our simulated models to fit with the observed structures we need to be careful. For example: the stellar evolution theory is an input in the simulations, so we need to be sure that it isn't something wrong with this theory when observations are in conflict with theoretical predictions. The cosmological parameters are also coupled to the observations and stellar evolution in various ways (Ellis, 2001).

5.1 Number counts of high-z galaxies

This is the observational method that is mostly used. By counting the number of observed sources in the early universe (at different redshift) and comparing with simulations we can test the theories. When doing this we have to remember that the cosmological parameters affect the number counts. The volume element that we can observe is dependent (because of general relativity) on how curved the universe is (flat, closed, open, accelerating). Cosmological effects also enter in the theories for hierarchical clustering in a complex way (Ellis, 2001). The Hubble deep field (obtained in measurements with the Hubble Space Telescope) can be studied using a technique called photometric redshifts (Lanzetta et al, 2002).

Kauffmann et al (1994) have studied the problem that faint blue galaxies are observed to more numerous than expected in the CDM model. They find that if they assume that these faint objects (which often are peculiar galaxies) are merging into other galaxies, the model can account for the observed number. Other studies have come to the same conclusion, (see Ellis, 2001 or Brinchmann & Ellis, 2000), using other methods.

5.2 Quasars at high redshift

High redshift quasars might challenge the hierarchical clustering model, because the number of observed quasars are too many in a CDM universe where most of the galaxies form late. Katz et al (1994) studied this problem and did numerical simulations that showed that the CDM model could explain the observed number of quasars. The most distant quasars can also give us a limit on how early stars were formed. One recent analysis of the most distant quasar (at z$\sim$6.28) indicate that stars have been formed as early as z$\sim$8 (see Pentericci et al, 2002).

5.3 Lyman break galaxies

Lyman break galaxies are high-redshift galaxies (around $z \sim$ 3-4 that produces UV-continuum photons (from star formation or an active galactic nuclei)(see Steidel et al, 1998). They can be detected using the Lyman break technique where the ionization of hydrogen give rise to a distinct spectral feature that enables the detection of distant galaxies. These primordial starforming galaxies can then be used to study how clustering of the galaxies takes place. Baugh et al (1998) used semianalytical galaxy formation to model the Lyman Break galaxies, and found that these objects are among the first starforming objects.

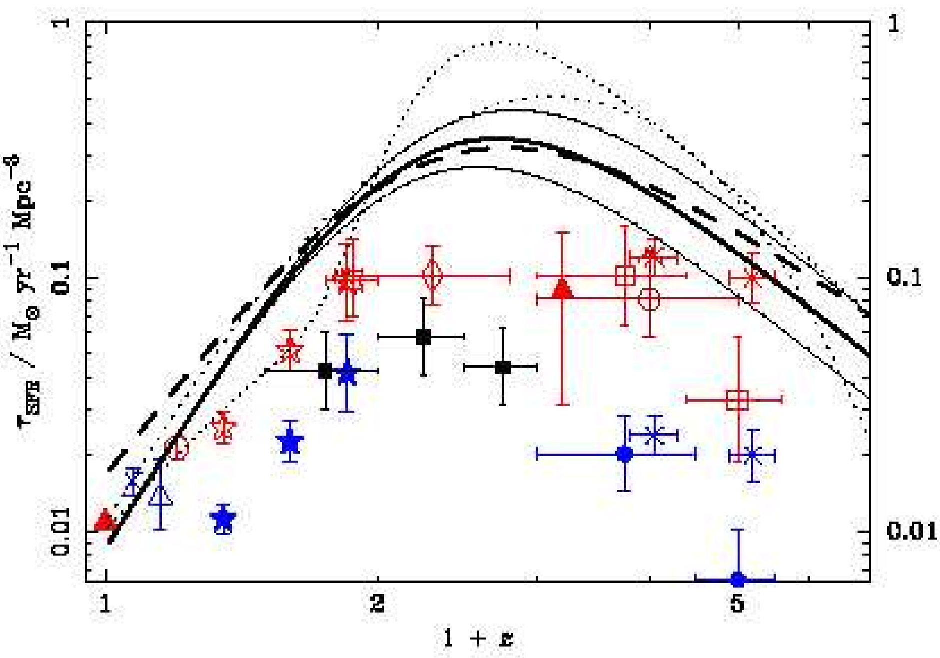


Figure 7: The history of star formation up to z∼5 with different fits and datapoints (from Ellis, 2001)

5.4 Cosmic star formation histories

How and when stars form is closely coupled to which model of galaxy formation we use. In the hierarchical picture stars are formed continuously during the period when galaxies start to form (at z∼8+) to z∼ 1. After this period the starforming is expected to decrease, since the universe has expanded and mergers become more rare. In the monolithic collapse scenario all stars are formed at roughly the same time, simultaneous to the formation of galaxies.

Using a compilation of various observations Ellis (2001) construct a diagram of the cosmic star formation history (shown in figure 7). In this diagram we see that stars form continuous over a wide range in z. This seem to fit with the hierarchical models. Even more recent observations indicate that the decrease in star formation after z∼3-4 (see figure 7) does not exist. Star formation increases monotonically with redshift up to very high redshifts (Lanzetta et al, 2002).

6 Summary and discussion

The subject of galaxy formation is a very large one, involving astrophysics from the large cosmological scales to small scales like stellar evolution. The problem is complex and the number of papers on the subject reflects this, but it is not unsolvable. During the last 20 years or so the observations seem to have caught up with and even surpassed theory, and this has caused some of the theories to stand out as more probable.

The CDM model for galaxy formation (or, more exactly, the ΛCDM model) seem to be the prime candidate for explaining how structure on the scale of galaxies and up form in the universe, but this is not yet a certain fact. The monolithic collapse scenario have not yet been completely abandoned since there are really no strong evidence that it isn't possible, but observations seem to point in the direction of hierarchical clustering.

The most probable process of galaxy formation is then given by the ΛCDM model. Selfsimilar fluctuations were set up before the recombination era, the only ones left after recombination were the dark matter perturbations. The baryonic matter then interacted gravitationally with the dark matter and this caused small scale baryonic structure to form in the early universe. Then larger structure was built up by hierarchical clustering and various kinds of galaxies were formed by merging and gravitational collapse. At the same time the gas in the protogalaxies was coupled to dark matter and could become cool enough for stars to form. At the redshift z∼1 most of the galaxies had been formed. After this the galaxies evolved by themselves, in some cases interacting, until today.

One problem of the hierarchical clustering model is that we would expect too see red galaxies (meaning that they have a old stellar population) with high redshift, corresponding to the first starforming galaxies (Ellis, 2001). In observations these red objects seem to be absent. This could be explained by stating that in the early universe the galaxies were gasrich. Even the old galaxies would then have some star formation, which would cause the galaxies to be blue, despite the fact that they really are old.

To conclude I would like too say that more research on the subjects needed. With better observations (or by using the observations we got more extensively, like HST deep field (Ellis, 2001) and with more refined theories (perhaps implementing more detailed models of the astrophysics (Silk, 2002)) we should be able to finally solve the problem of galaxy formation.

Acknowledgements

I would like to thank Alessandro Romeo for his support and the encouragement he has given during my work with the subject. I would also like to thank Duilia de Mello for some helpful comments on the subject and a lecture on presentation techniques. At last I would like to thank the other students of the course *Hot Topics in Astrophysics* for comments and advice.

References

Ahmad, Q. R. et al. 2001, Physical Review Letters, 87,1301

Baugh, C. M., Cole, S., Frenk, C. S., & Lacey, C. G., 1998, ApJ, 498, 504

Brinchmann, J. & Ellis, R. S., 2000, ApJL, 536, L77

Cole, S., Aragon-Salamanca, A., Frenk, C. S., Navarro, J. F., & Zepf, S. E., 1994, MNRAS, 271, 781

Eggen, O. J., Lynden-Bell, D., & Sandage, A. R., 1962, ApJ, 136, 748

Ellis, Richard, 2001, arXiv:astro-ph/0102056

Katz, N., Quinn, T., Bertschinger, E., & Gelb, J. M., 1994, MNRAS, 270, L71

Kauffmann, G., Guiderdoni, B., & White, S. D. M., 1994, MNRAS, 267, 981

Kolb, E.W. & Turner, M.S. 1990, *The Early Universe*, Addison-Wesley Publishing Co., Redwood City: California

Lanzetta, K. M., Yahata, N., Pascarelle, S., Chen, H., & Fernández-Soto, A., 2002, ApJ, 570, 492

Longair, Malcolm S., 1998, *Galaxy Formation*, Springer, Berlin

Moore, B 1999, Phil. Trans. R. Soc. Lon. (A) (obtained from http://www.nbody.net)

Press, W. H. & Schechter, P.,1974, ApJ, 187, 425

Silk, Joseph, 2001, arXiv:astro-ph/0112295

Steidel, C. C., Adelberger, K. L., Dickinson, M., Giavalisco, M. & Pettini, 1998, arXiv:astro-ph/9812167

Firestorm of Starbirth in the Early Universe

Theresa Wiegert

Chalmers University of Technology
SE-41296 Göteborg, Sweden
(f95thwi@dd.chalmers.se)

*

Abstract

New Analyses of the Hubble Deep Field survey show that the star formation in the early Universe was much more massive than earlier measurements imply. This paper looks into the subject, what is seen and how it was obtained.

1 Introduction

A few hundred million years after the Big Bang, the black space of the universe was suddenly lit up by a firestorm of star formation. Today's star formation in the galaxies is very small compared to this big one, when a large part of the stars of the universe were formed. Using the Hubble Space Telescope, astronomers have been able to look a long time back, and the further one looks, one can see that the rate of starbirth is continually escalating.

Analysing the two Hubble Deep Field images, from 1995 (towards the celestial north pole) and 1998 (towards the celestial south pole), it has been found that the farthest objects in the deep fields are 'the tip of the iceberg', i. e. due to dimming effects, what we see are only the very brightest objects. And that implies that there should be a lot more light. About 90 percent of the light is missing in the pictures. One popular metaphor for this is that "it's like seeing only the lights on a distant Christmas tree and inferring the presence of the hole tree". This missing light would reveal a period of starbirth unlike anything the universe will see ever again.

The farthest objects in the deep field images seem to be extremely bright and hot newborn stars in primordial galaxies. These findings are entirely different from earlier beliefs, which stated that the star birth rate was more continous and not so massive.

In this paper, I have looked at the background of the star formation rate intensity in the Universe, how these new results of the early massive star formation were obtained, and how accurate they might be. Another interesting question is: what happens next? These results are rather recent, and a couple of years ago we did not know much about the subject. What will happen in say ten years?

*Hot Topics in Astrophysics 2001/2002, Alessandro B. Romeo & Eva S. Karlsson (Eds.), Chalmers University of Technology and Göteborg University, 2002.

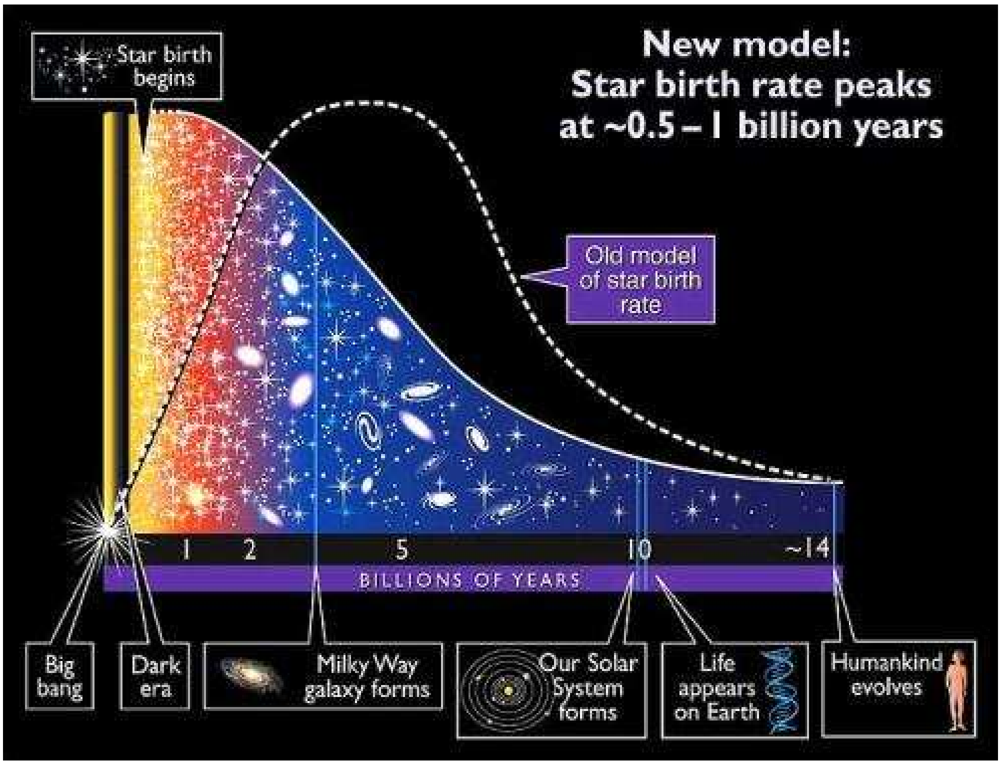


Figure 1: A schematic picture of the timeline of the universe, showing the new model of the star birth explosion, versus the old one of more continuous star birth. (From A. Feild (STScI))

The first parts of the paper, section 2 and 3, are concentrated on explaining different concepts giving a perspective important for understanding, while the rest of the article gives information of the star formation rate intensity distribution function and its implications for being able to reach results of this kind. I will also touch the subject of reliability and the futuristic aspects in the last parts.

2 The young Universe

2.1 How long ago?

First of all, we have to be able to look back in time a long way. The universe is proposed to be somewhere between 10 and 20 billion years old. It's hard to measure it better than that, although there are stars measured to be 14 billion years old, which would imply that the universe is at least older than that.

According to the theory, after the Big Bang, the universe was composed of glowing plasma. For the next half a billion years this plasma expanded and cooled down. Then the plasma recombined (or combined in this case, since it was the first time it happened) into atoms of neutral gas - hydrogen and some helium. The glow from this era is left in our time as the cosmic microwave background radiation, the small temperature of 2.7 K.

2.2 The dark ages

The next nearly one to two billion years after the recombination is called the Dark Ages. That's when the cold gas began to gather together into the first galaxies. When these galaxies and quasars that were formed reionized and changed the character of the surrounding neutral gas, the dark ages ended. The period after the dark ages, is the period we are interested in - the period of the earliest stars in the universe.

2.3 Redshift and distances

When looking at great distances, it is no longer possible to use lightyears or parsecs or other units of distance. The reason is that the distances are far too great, and there is no good way of measuring them. Redshift is a way of measuring how fast an astronomical object is moving away from us (most galaxies are receding from us). It is measured directly from the spectral line shifts.

Even though redshift is a measurement of velocity, it is often used as a convenient way of measuring cosmological distances, as the redshift depends on the distance of the object - the further away a galaxy is from us, the larger its redshift is (and the larger its recession velocity is). Out to distances of 4000 Mpc, equivalent to a redshift of $z = 1$, every galaxy whose distance can be measured obeys the following Hubble's redshift relation:

$$v = H_0 \cdot D \tag{1}$$

where v is the velocity of the object, H_0 is the Hubble parameter, which is not very accurately determined yet, and D the distance. The relationship between redshift and velocity looks like this (relativistically):

$$z = \sqrt{\frac{(1 + \frac{v}{c})}{(1 - \frac{v}{c})}} - 1 \tag{2}$$

One of the assumptions we have to make is that the original light coming to us from thousands of millions of years ago was emitted at the same wavelengths as it would today in equivalent stellar processes.

Redshift is good way of measuring distance, since it is a directly observable quantity and can be measured accurately from spectral line shifts.

2.4 Cosmological surface brightness dimming and dust

Space is not entirely empty but contains high quantities of dust, interfering with the measurements we make. Cosmological surface brightness dimming is another effect of the distances that makes galaxies far away not only small but faint and dim.

For objects of redshift $z > 2$, the cosmological surface brightness dimming effects plays a dominant role. Now, how much will the dimming affect measurements? Do we need to take it into account? The answer is yes. About 90 % of the light disappears because of the great distances, especially from the weaker areas.

To find out the effect of the cosmological surface brightness dimming, we take a look at the distribtion of x as a function of redshift z, where x is the intensity of unobscured star formation rate, i. e. the star formation rate density inferred directly from the rest frame ultraviolet light that is not obscured by dust.

3 The Hubble Deep Field Images

600 km above the surface of the Earth, and thus able to avoid the disturbances of the atmosphere, a spacebound telescope orbits the Earth. It is called the Hubble Space

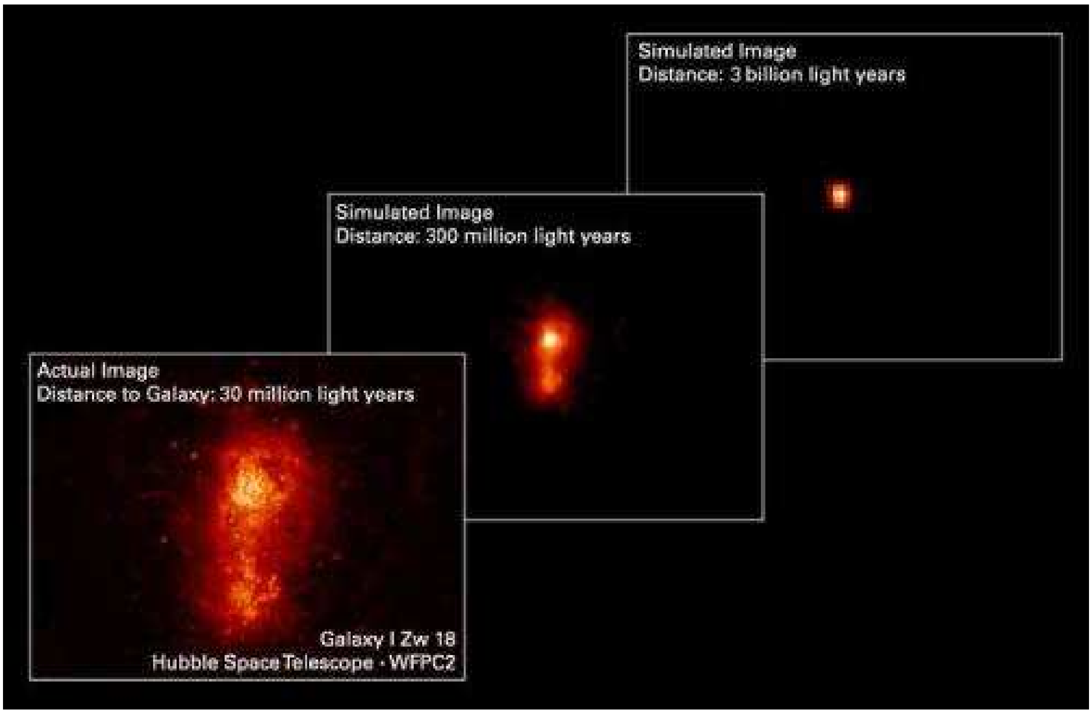


Figure 2: This is a galaxy not so far from us, which is giving birth to lots of stars. To illustrate the dimming effects of the distance, the picture has been simulated as to how it would look at greater distances. (Archival Image: D. Hunter (Lowell Observ.) and A. Aloisi (JHU). Simulated Images: A. Fruchter and Z. Levay (STScI)

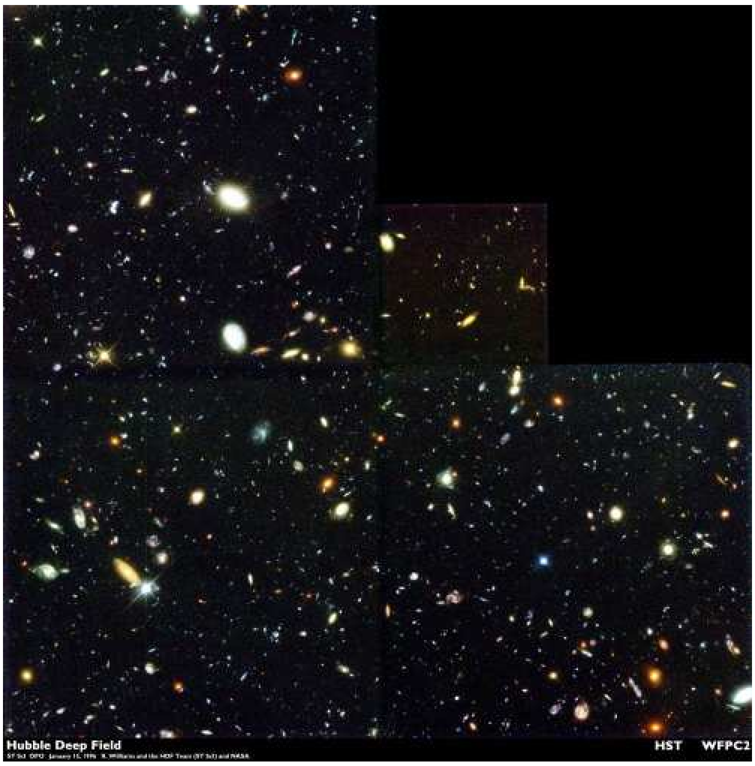


Figure 3: The Hubble Deep Field North - taken in 1995, in a spot just above the Big Dipper in the Great Bear constellation. (R.Williams and the HDF Team (ST ScI) and NASA)

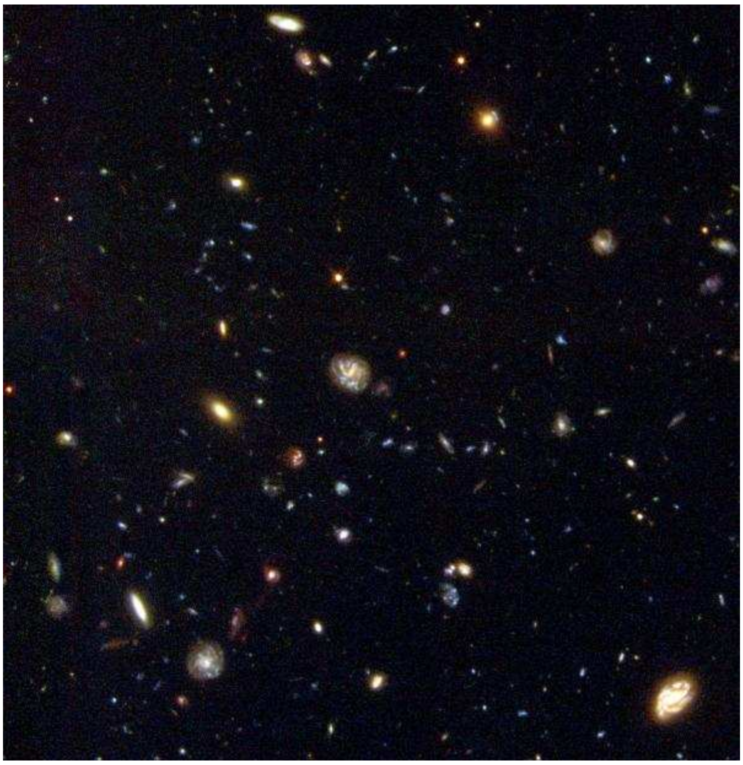


Figure 4: Hubble Deep Field South - HDF-S. A deep field image taken from the southern hemisphere, in the constellation Tucana, near the celestial south pole. (R. Williams (STScI), the HDF-S Team, and NASA)

Telescope (HST), named after the famous astronomer Edwin Hubble, the same Hubble responsible for the Hubble distance law mentioned in section 2.3. The HST was launched in 1990, and after some adjustments of the intruments, it has been producing lots of images loaded with valuable information.

In december 1995, the Hubble telescope was set to observe a dark spot in the northern sky, just above the Big Dipper in the Great Bear. During 10 days (150 orbits around the Earth) it collected photons from this spot, and the result was an astonishing picture of about 3000 galaxies, lying at extreme distances from us. It is called the Hubble Deep Field image (HDF). In 1998, another survey just like this one was done, this time of a spot in the sky of the southern hemisphere, in the constellation Tucana, near the celestial south pole. Consequently, the result is called Hubble Deep Field - South, or HDF-S for short. This spot was chosen because it lies near an interesting quasar (quasi-stellar object, for example a bright galaxy at great distance) of redshift $z = 2.2$.

Both pictures are quite alike, showing galaxies of the 30th magnitude or about four-billion times fainter than can be seen by the human eye. So far away, that some of them might be as old as the first galaxies formed in the universe. In fact, the pictures show an about 12 billion light year long tunnel of space. In the foreground we can even see some stray stars from our own galaxy.

These Hubble Deep Field measurements have lots of information to examine, which give insight on the formation of galaxies and the structure and evolution of the universe. They were used by Lanzetta et al (2001) for the subject discussed in this article of the explosion of early star formation.

4 The survey of the HDF-galaxies

By processing all possible available data from spacebound and earthbound images, optical and infrared (from HDF and HDF-S both NICMOS and WFPC2-fields, images from Kitt Peak national Observatory KPNO 4 m telescope, European Southern Observatory (ESO), New Technology Telescope (NTT) and VLT (Very Large Telescope) a survey of a large amount of the galaxies was produced.

This survey consists of photometry measurements and photometric redshift measurements for about 5000 galaxies. They are collected in a catalogue, called the Stony Brook Faint Galaxy Redshift Survey (Stony Brook is the name of a university in New York)

4.1 Photometry measurements

Photometry is the measurement of optical light, i e the wavelength interval our own eyes are sensitive to (360 - 830 nanometers). In the images, the objects were detected by starting at low wavelengths, and going on towards higher wavelengths. After having made the lowest wavelength-measurements, the measured areas were 'removed' by masking them, and then measurements were done again, thus not measuring again the already measured areas. Bandpasses of short wavelength were used to measure low and moderate-redshift galaxies, and high wavelength bandpasses were used for the high-redshift galaxies. Short wavelength bandpasses have better resolution and sensitivity, but do not work well at high-redshift galaxies, whereas high wavelength bandpasses are the exact opposite - low sensitivity and resolution, but suitable for high-redshift galaxy measurements.

4.2 Photometric redshift measurements

As well as the photometry measurements, photometric redshift and spectral types of the objects were measured. For this, astronomers (Lanzetta and his associates) used six spectrophotometric templates for different galaxies: E/S0 (elliptical and lenticular), Sbc (barred spirals) Scd, Irr (irregular) and low and high extinction starburst galaxies (SB1 and SB2). Another interesting feature to look at is the effect the neutral hydrogen has, by intrinsic and intervening absorption. For galaxies of redshift z < 6, it was possible to see how good an accuracy and reliabilty the measurements have, by comparing them with with spectroscpic measurements of some galaxies from the HDF-images. And the result was rather satisfying. The RMS relative dispersion with respect to spectroscpic redshift measurements has a value of $\frac{\Delta z}{(1+z)} < 0.065$, and there are no known examples of photometric redshift measurements that differ more from spectroscopic redshift measurements than some times the RMS relative dispersion.

4.3 Results of the survey

In the catalogues that became the results of this survey, there are some 5000 faint galaxies, and about 1000 of those have redshift more than $z = 22$. More than 50 have a redshift of more than 5, all the way up to and past $z = 10$. The measurements done of photometry and photometric redshift, as well as angular area versus depth relations (see Lanzetta et al, 2001) are thus collected in a galaxy redshift survey, containing galaxies of highest z and faintest limits that are available now. The survey is called the Stony Brook faint galaxy survey.

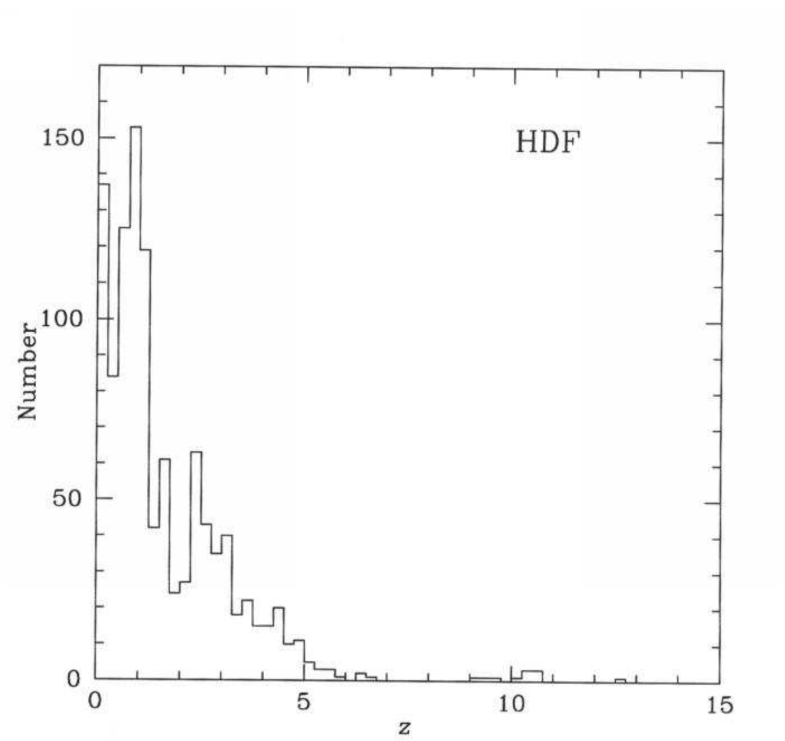


Figure 5: The distribution of galaxies from HDF of different redshifts (Lanzetta et al 2001)

Fig. 5–7 show the redshift-measurement distribution of all galaxies identified in the HDF-pictures (one HDF-north and two HDS-south, the NICMOS and the WFPC2). There are som differences in the pictures, which might not be expected, since universe is supposed to be similar in all directions on a large scale - isotropic and homogenuous. The HDF-S has a larger peak around $z = 2$. This could be explained by the fact that the spot for the HDF-S images was not chosen randomly, but because of its vicinity to a quasar of redshift $z = 2.2$ - we might be looking directly into a galaxy cluster, and thus the bias towards redshift $z = 2$.

5 The star formation rate intensity distribution function

In all this, the rest-frame ultra violet luminosity density is an important concept - this ultra violet light is produced in hot massive young stars, and is thus a good tracer of the history of cosmic star formation.

Just some years ago (the end of the 1990's), it was believed that the UV-luminosity density got higher with redshift, to the redshift $z = 1 - 2$, and then it would sink or remain constant as redshift went higher. That would mean that the stars of the universe were formed gradually over all of the 'cosmic time'.

But - an important feature that was not taken into account then is the effect of cosmological surface brightness dimming, which very likely is important, especially at high redshifts. The surface brightness diminishes as $\frac{1}{(1+z)^3}$, due to the expansion of the universe.

This shows when you look at far-away galaxies. The ones not so far away (lower

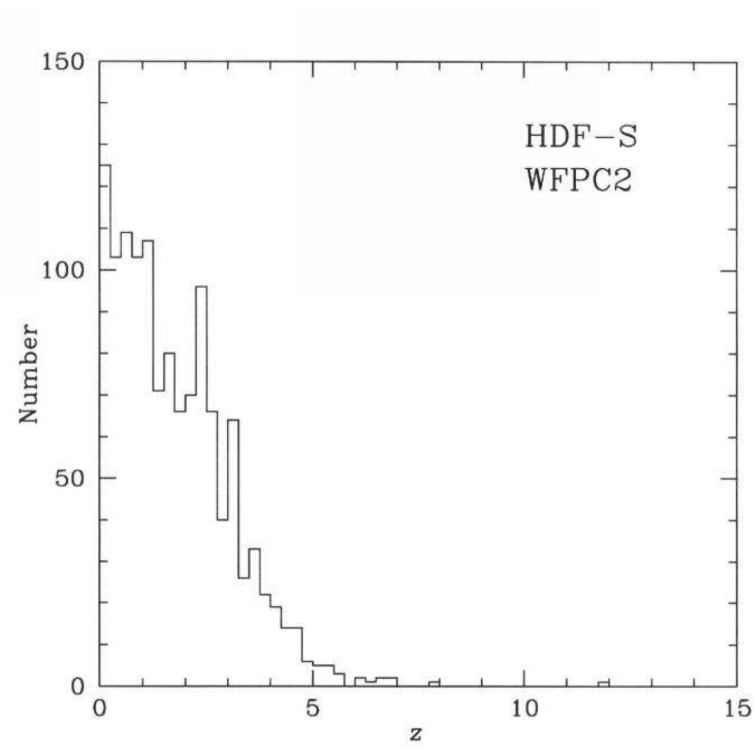


Figure 6: Distribution of galaxies of different redshifts, from the HDF-S, WFPC2 (Lanzetta et al, 2001)

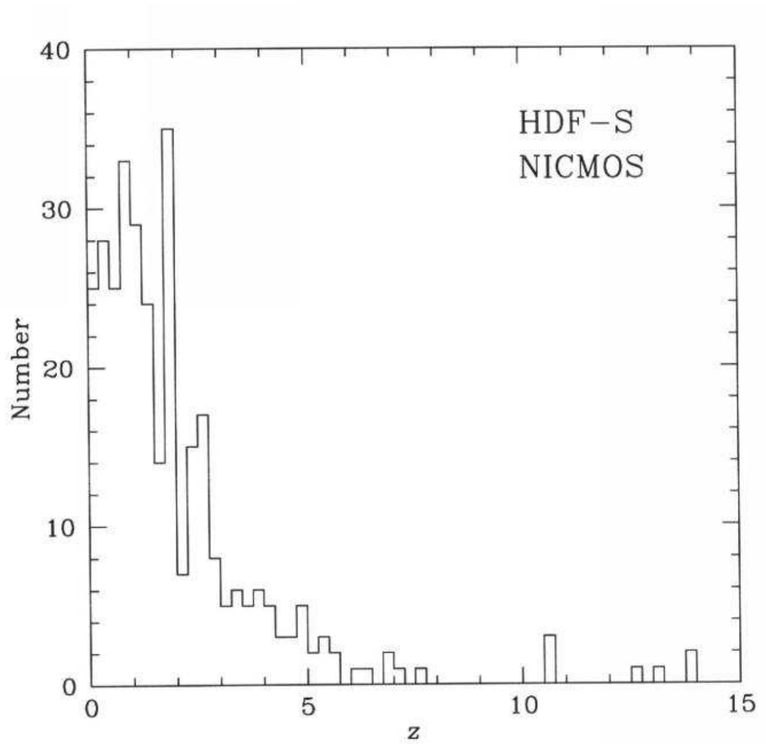


Figure 7: Distribution of galaxies of different redshifts, from the HDF-S, NICMOS (Lanzetta et al, 2001)

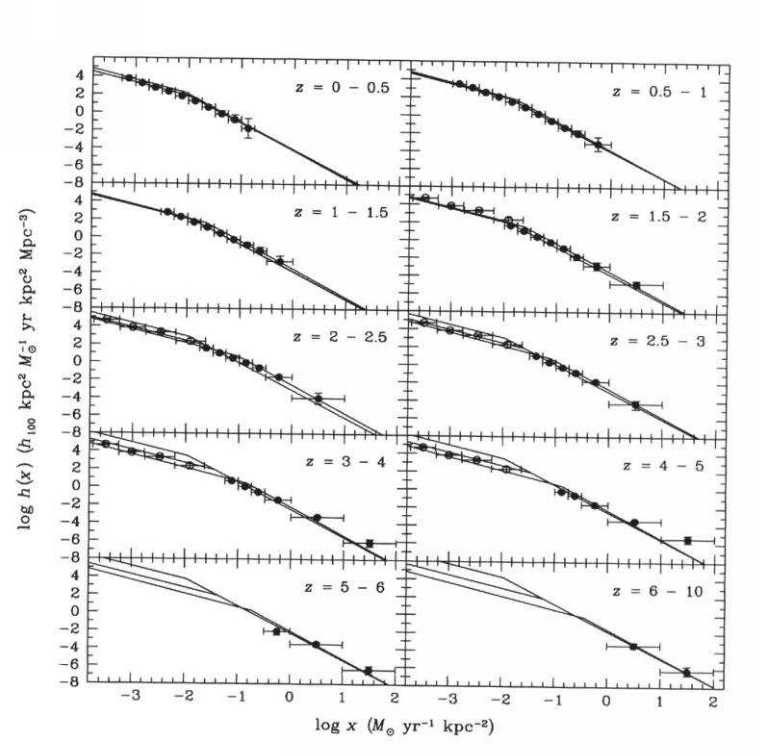


Figure 8: The star formation rate intensity distribution function $h(x)$ plotted against the star formation rate intensity x at different redshifts. (Lanzetta et al, 2001)

redshift) show both bright and faint areas, but the ones with high redshift only show the bright fields. The faint fields in these ones can not be seen because of the background noise. All measurements are thus losing a big part of the light from high-redshift galaxies.

There is still a great deal to do. One can examine the part of the star formation rate intensity that is visible (unobscured), versus redshift - that is, the star formation rate intensity that is directly inferred from the observed rest-frame ultraviolet light, not obscured by dust.

In all, you can directly estimate how much the previous measurements have underestimated the ultraviolet luminosity density at high redshifts. The measurements do not take into account the obscuration by dust, so everything is a lower limit to the total star formation rate density.

One comment about the dust effects would be appropriate here: Dust is of a limited size, so when looking at higher wavelengths than the optical, more information should be able to get through, and the dust would not be such a big problem. We can directly see this by looking at figures 5 through 7. Fig 7 show measurements taken with the NICMOS camera (Near Infrared Camera Multi-Object Spectrometer), that is, measurements of the infrared wavelength. Comparing with the other HDF-S graph WFPC2 (Wide Field Planetary Camera), it is obvious that the NICMOS measurements have more data.

6 The results

Obtaining the star formation rate intensity distribution function is not an easy task, neither to do nor to explain. For a more thorough explanation, see Lanzetta et al (2001). Here I will only sketch the outlines, to try to give some perspective of it.

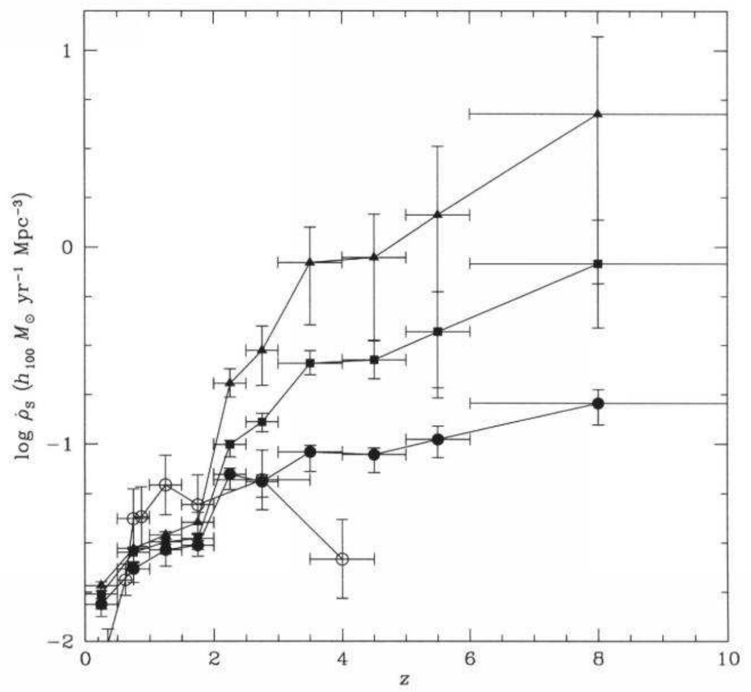


Figure 9: The unobscured star formation rate densities against redshift z. (Lanzetta et al, 2001)

The deep field images show the galaxies in a rather good resolution, so good that it is possible to divide the different galaxies into pixels, and measure the star formation rate in every pixel. To determine the star formation rate intensity distribution function, you have to consider all pixels in all galaxies in all fields. The first thing to do is measuring the luminosity of each pixel (using energy flux of the pixel, the redshift (same as the redshift of the galaxy the pixel is a part of), empirical K-correction of pixel, and the cosmological model). Having the luminosity, next step is to determine the star formation rate from it, by using a Salpeter stellar initial mass function. One more thing is needed: the proper area of each pixel (using redshift, angular plate scale and the cosmological model again). Those are the ingredients needed for the star formation rate intensity x of each pixel - you divide the star formation rate by the proper area.

We also need the star formation rate intensity distribution function $h(x)$. That is achieved by using a series of steps concerning proper areas, comoving volume and angular area versus depth relation. I again refer to the paper by Lanzetta et al (2001) for an complete explanation. Star formation rate intensity distribution function $h(x)$ versus the star formation rate intensity x at different redshifts is shown in figure 8.

Applying a model to $h(x)$ versus x, (a broken double power law model) and integrating over the models, the star formation rate intensity versus the redshift is determined (the models are shown as the lines in figure 8).

The final plot (Fig. 9) is the one of the unobscured star formation rate intensity plotted against the redshift z. And this is from where our conclusion is drawn: The luminosity density (ultraviolet) increases with redshift. The ultraviolet luminosity density is equivalent to the unobscured star formation rate density, which thus also increases with redshift. Figure 9 also shows the old measurements of star formation (the unfilled dots), from 1996 and 1998 (Madau et al)

6.1 Summary

At redshifts higher than $z = 2$, the dimming effects are important. Not taking the dimming into account, a big part of the luminosity would be missed, as has been in earlier measurements. Taking the dimming effects into account, and using the survey of galaxies from the Hubble deep field images, it is shown that the star formation rate is increasing with redshift. That is, when we look back in time (distance) far enough, we can see that the formation of stars was enormous, much greater than we have believed before. Compared to the not so very significant star formation of today, the earlier star formation can be called a firestorm of starbirth in the early universe.

7 Reliability of the measurements

How reliable are the measurements? What facts can we rely on, and what might make it somewhat shaky? There are a number of advantages (Barkana et al, 2002) like the fact that we are looking at a really large number of galaxies from the Lanzetta galaxy survey, imaged with a high resolution with HST and with measured photometric redshifts. Since the galaxies are of good resolution, one can divide them into pixels and measure the star formation rate in every pixel. The result of that is that the effect of the surface brightness dimming is easier to understand than in the case of measuring the galaxy luminosity function.

There are some possible drawbacks as well. Although the observed fields are deep, they are also small in angular size, and they were not randomly selected. Cosmic variance could thus be of importance. One example is the biasing of objects of redshift around $z = 2$ in the HDF-S images. One other unsure aspect is the measuring of the spectroscopic redshifts. To do this, spectral templates were used, for different kinds of galaxies. The technique of the templates has been tested, but only for redshifts of z=0-1.2 and z=2.2-3.5. There might still be a risk that the templates do not work for higher redshifts.

8 What happens next?

As I mentioned in the introduction, one interesting aspect of these new findings is that they are very new. We have not been able before to look so far away in the universe, as we have done now with the Hubble Deep Field images. So what will happen next?

On the 1st of March this year (2002), the space shuttle Columbia took off on a servicing mission to the Hubble Space Telescope. During the mission, the astronauts installed a new device, The Advanced Camera for Surveys (ACS). With this camera, we are able to look even further out into the universe. The first images from the ADS show great promise. Lanzetta will use the Advanced Camera for Surveys to try to directly verify some of the missing light, and apart from that, he will look for distant supernovae, as an alternate measure of star formation. Supernovae are point sources of light, as compared to galaxies that are extended sources of light, so they are not subject to the same cosmological brightness dimming effects.

After that, I can only speculate what will happen. But astronomical research is in a very exciting period of time now, as telescopes are improving in every way, and new ones are being built (the ALMA for instance). One thing is sure - the research of this particular area has just begun, and in the following years, new results are sure to come, as our equipment gets better. In short - it is all very exciting!

References

Carroll B.W. & Ostlie D. A., 1996, An Introduction to Modern Astrophysics. Addison-Wesley Pub Co

Stellar 'Fireworks Finale' Came First in the Young Universe: `http://oposite.stsci.edu/pubinfo/PR/2002/02/`

Lanzetta K. et al, 2001, The Star Formation Rate Intensity Distribution Function - Implications for the Cosmic Star Formation rate History of the Universe, arXiv:astro-ph/011129

Barkana R., 2002, The Star Formation Rate Intensity Distribution Function - Comparison of Observations with Hierarchical Galaxy Formation, arXiv:astro-ph/0201492

Madau P., 1999, Cosmic Star Formation History and the Brightness of the Night Sky, arXiv:astro-ph/9907268

Madau P., 1999, Starlight in the Universe, arXiv:astro-ph/9902228

HDF and HDF-S: `http://www.stsci.edu/resources/`

HDF-S: `http://www.stsci.edu/ftp/science/hdfsouth/hdfs.html`

ACS: http://acs.pha.jhu.edu/

ACS: `http://sites.stsci.edu/pubinfo/pr/2002/11/pr.html`

Dark Ages: `http://www.news.harvard.edu/gazette/1997/01.09/DarkAgeoftheUni.html`

Cosmological and Astrophysical N-Body Simulations

Anders Hedenrud

Chalmers University of Technology
SE-41296 Göteborg, Sweden
(f97anhe@dd.chalmers.se)

*

Abstract

The topic Cosmological and Astrophysical N-Body Simulations has many sides. In this paper several of the basic aspects, both from a theoretical and from a practical point of view, are reviewed.

1 Introduction

N-body simulations are an essential part of both cosmological and astrophysical research. It helps us to understand and to reach conclusions about the theories and models that are formulated in the field. Since this paper is a complement to a seminar held for the course "Hot Topics in Astrophysics" it feels necessary to say that some graphics and animations of simulation results can be found on the internet with the homepage for Max-Planck Institute for Astrophysics as a good starting point. See references.

First we should ask the relevant question why we need N-body simulations. Let us first think about the N-body part of this expression. Whenever we have two bodies interacting, calculations about what this interaction will look like can be done. When a third body or more bodies are let loose in the system chaos is introduced and complicates things. Analytical solutions to the problems are now impossible to find and an N-body problem stand at hand. It is also true that practically all systems that we find in real life are of this nature and it is more of a question how many bodies do we need - in other words: how large will our N have to be, to accurately enough characterise our system.

If we then turn to the simulation part there are several reasons why we should use simulations. One is that since astronomical timescales are very large generally, simulations can be the only way to test models and can often at least sort out the ones that are wrong. Other models and theories can get support from simulations, just as simulations can get support from reality and collected data. Simulations are also a valuable tool when it comes to tuning in parameters in models to make them agree with reality. Simulations can also often give valuable insight in how systems work and produce visualisations of complex systems.

*Hot Topics in Astrophysics 2001/2002, Alessandro B. Romeo & Eva S. Karlsson (Eds.), Chalmers University of Technology and Göteborg University, 2002.

The paper is organized as follows. A brief history and background are given and some important objects of interest are mentioned: formation and development of the universe, globular clusters and galaxies. In addition three groups of simulation methods are mentioned. These are the Particle-Particle, Particle-Mesh and Tree code methods. In the end information about related hardware and software is treated. The whole paper is meant to give an insight into all these fields and to give an understanding of what happens therein today and in the future.

2 History

The history of astrophysical simulations like simulations in many other fields, more or less walks hand in hand with computer development. Of course one can always simulate galaxy formation by perturbing a cup of coffee, pour in some cream and observe how white spiral structures are formed. However, since it is hard to control initial conditions as well as read out any results, I believe that computer models and simulations are preferable.

It all begun in the early 60s and 70s when basic models were formulated and important but not to advanced simulations were made. For example, contractions of particle clouds consisting of a few hundred particles were simulated by raw summation of two-body gravitational forces. As time went by, the size of N increased together with computational power. Also new more intricate models were introduced that again lowered the number of particles, but also made simulations applicable on more problems. While in the 60s 300 point masses were used in the 70s 700 particles with different masses could be used. The 80s were a golden age for both new and refined models as well as for hardware development, but there was still a hunger for more. Also in the 90s models were formulated and further developed, but I believe that the main progress has been the hardware development with dedicated computer models or supercomputers that has made problems previously impossible to solve solvable. Depending on the problem at hand perhaps hundreds of thousands or millions of particles can now be used.

3 N-body systems

What kind of N-body systems are out there? One of the most complex and uncertain areas are cosmological studies and large scale structure formation of the universe. Without simulations we can use observational data from deep field views, but these can merely give hints to solutions and in the end it is needed to test models and try out parameters in other ways. Here simulations are an essential tool. What is simulated is how dark matter is contracted to the structures we see today from an initial system of evenly distributed matter due to the supposed inflation right after the proposed Big-Bang. Dark matter is the by us still undetectable matter that due to the formations we see is supposed to contribute to the bigger part of the mass in the universe. In the most condensed parts of this dark matter galaxies or clusters of galaxies, depending on scale, are supposed to dwell. A 3D visualisation of such a simulation can be found in figure 1.

N-body systems can be found everywhere in nature. What is so special with astronomical systems then? The one simple answer is: gravity. Since mass is quite sparsely distributed in the universe it is quite reasonable to say that pure gravity is the single most important factor to take into account when dealing with astronomical simulations. However, in certain problems, like in colliding galaxies or when large time spans are used, methods can be borrowed from gas dynamics and hydrodynamics to give results closer to

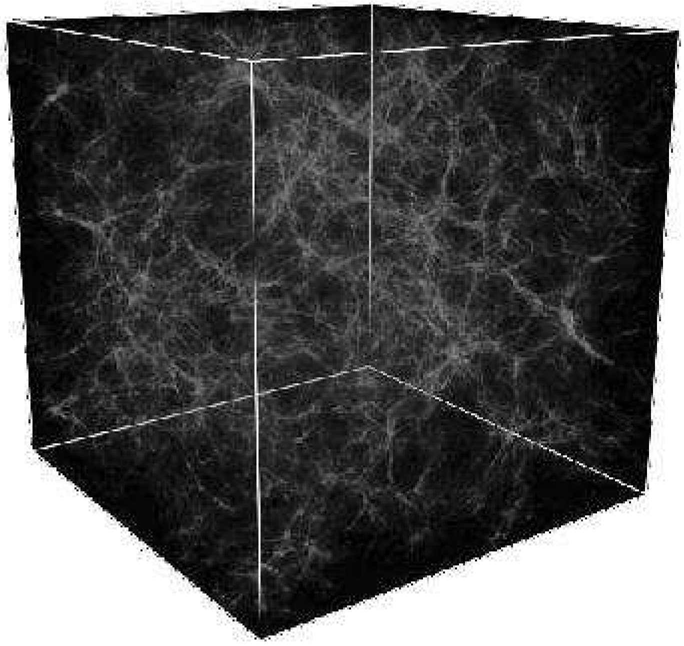


Figure 1: 3D visualisation of cosmological dark matter simulation from NCSA by Greg L. Bryan and Michael L. Norman, University of Illinois

reality. Approximations can always be used to reduce CPU usage and perhaps enable the use of more particles or use of other time consuming calculations. The art is to know or find what is needed to get meaningful and usable output. Both exaggerated precision - perhaps even in wrong parameters, and coarse approximations can give problems.

4 Methods

At this point I would like to bring up some methods used and knowing that there are practically as many varieties of methods as there are science groups, I will only mention three main groups and their main characteristics respectively. These three methods are Particle-Particle, Particle-Mesh and Tree code. (see Klypin 2000; Spurzem 2002) Let us start with Particle-Particle.

4.1 Particle-Particle method

The Particle-Particle method can be said to be the most fundamental of the methods. It was the first, but is refined through the decades. In the Particle-Particle case you simply accumulate forces by every particle for every particle, so that each particle feel the gravitational interaction of every other particle. With this in hand, equations of motion are integrated for each particle. New positions and properties are given as a result and the process can start from the first particle again. The basics are fairly simple here, but the integration part is far from simple. It can also be clarified that this method can give "exact" results while the following two make certain approximations to save CPU time.

A good example of application for this method could be a globular cluster, where the compact structure of the system makes interactions from all the other components important.

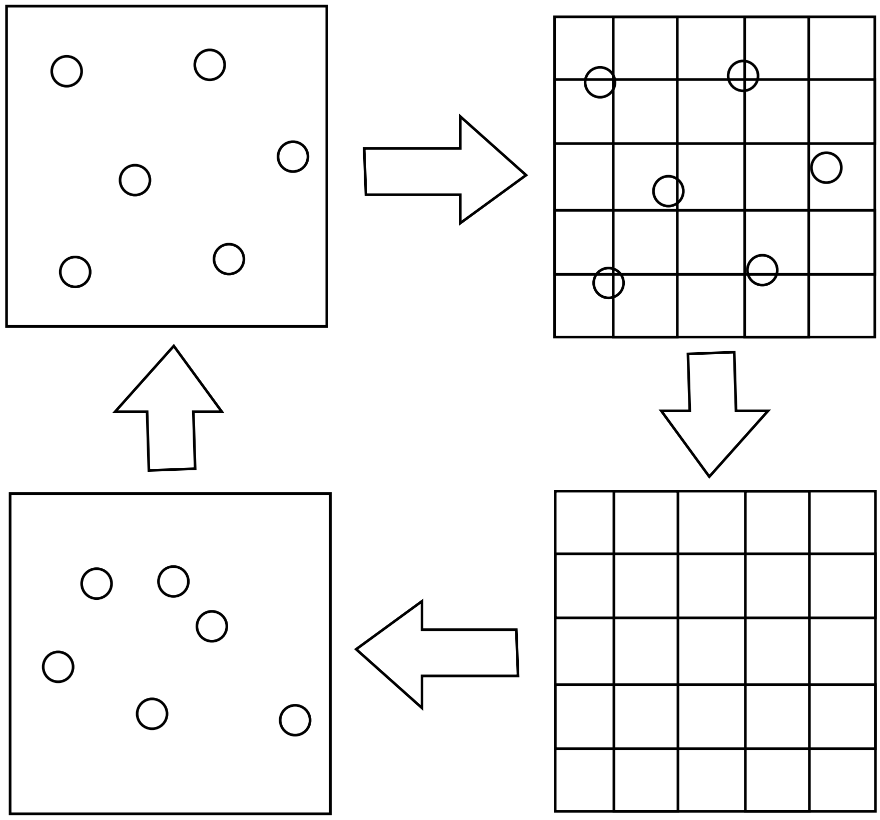


Figure 2: The Particle-Mesh method. Data from particles are interpolated on a mesh. Equations are integrated and applied on the particles to give new values.

4.2 Particle-Mesh method

The Particle-Mesh methods do these approximations by quantifying forces from the particles used as a field on a mesh. Potentials and forces for each particle are interpolated. After this integrations of motions can be made and new particle positions and a new grid-density constructed. The resolution is determined by the size of the mesh used. A visualisation of the method is shown in figure 2.

An application for the Particle-Mesh method could be the dynamical model of a galaxy, where a field grid is enough to simulate the force interaction and more focus can be set on other factors.

4.3 Tree code method

The third method was the Tree code. It uses the fact that clusters of particles far away from a single particle could be treated as one mass rather than several when gravitational interactions are calculated. So by building a tree with a hierarchy from large scale structures down to singular particles, lots of calculations can be omitted in the far off cases. Each particle gets its own cell in the tree grid as part of the greater whole. The factor that determines the resolution of this method is a factor called *theta*. It is taken as the angle necessary to include a certain size of a grid structure from a position. In other words: A large angle results in that systems of particles close to the point of view are dealt as one mass. This will bring us a bad resolution. A small angle brings us closer to a situation where every single particle counts as a single mass, that is the Particle-Particle case but with extra calculations being made, and gives better and better resolution. A

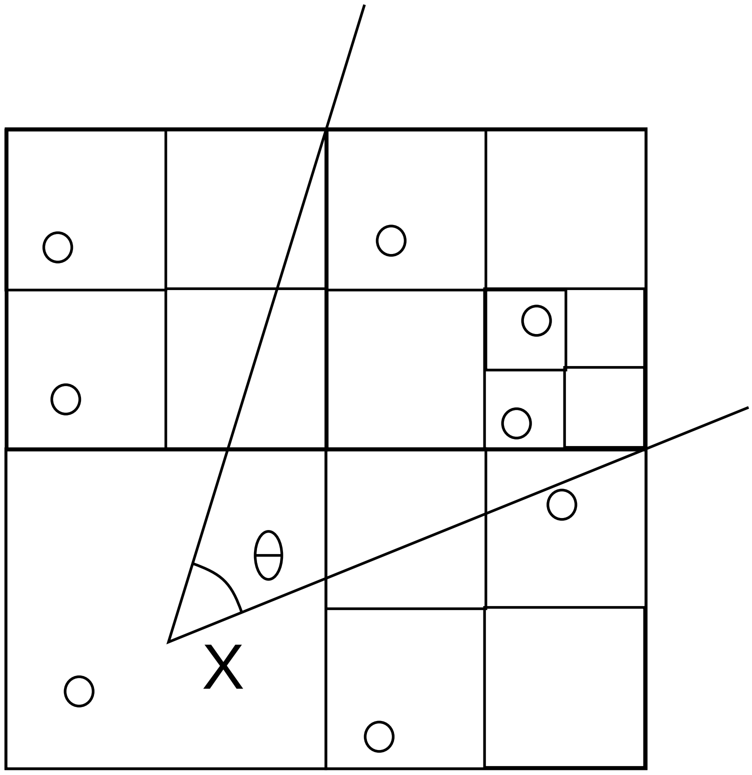


Figure 3: The Tree code method. Structures are divided into subunits and resolution from the point X is determined by the angle between its field of sight.

visualisation of the method is shown in figure 3.

This method is well suited for spread out systems with humps of particles far away from each other. Cosmological calculations like the one presented in figure 1 is a good example of this.

4.4 Hydrodynamical effects

In all these cases non gravitational forces can be introduced by adding hydrodynamical models like SPH, Smoothed Particle Hydrodynamics, to the calculations.

5 Hardware and software

5.1 Can I do this at home?

We have now come to the hardware and software part of the paper and immediately the question arises: Can I do this on my computer at home? The answer to this question is both yes and no and goes back to the fact how we define N, what kind of results we wish to achieve and for how long we are prepared to wait. In early days what then counted as supercomputers had far less computing capacity than an ordinary workstation of today. Harder tasks would need some hundred years to be completed, but would be possible. For more educational use there are several programs and applets that one can play around with, without loosing the scientific value of the results.

5.2 Scientific use

5.2.1 Hardware

When it comes to science of today the goal is always to use the best system or supercomputer available to be able to increase the number of particles and factors relevant to give a good resolution in the results without introducing to many errors. These systems range from parallel supercomputers constructed for general purposes, and well suited for these tasks as well, to highly specialised systems like the model series GRAPE that is built and maintained in Tokyo, Japan. (see Spurzem, Kugel 2002) Now in 2002 this series has been upgraded and the new GRAPE 6 has a calculation capacity of 48 Tflops. This is however a highly specialised calculation process and can not be used for other purposes than for computing the gravitational interaction. As a comparison the, at the moment by far, fastest supercomputer of 40 Tflops (vector based), also situated in Japan demands a large two floor building, while the GRAPE 6 can be fitted in a large closet more or less.

5.2.2 Software

The use of supercomputers reflects in the scientifically used software. Programs are often customized for certain processor architecture. To be able to exchange results and datasets between groups with different equipment it has become more and more important to set standards. The programs used can be made for very specific purposes to make everything effective. There are also program packages available with a core program and combined with that several different packages that solves a range of problems. It is also possible to create your own plug-ins to meat your specific demands. While this is flexible it might decrease efficiency somewhat. One of these platforms is Starlab, which beside the calculation packages wants to put a focus of how the received results can be used to visualise the data in different ways. This can be made in tree dimensions in for example a planetarium and bring about a virtual observatory that can be used both for show and understanding. (see Hut 2002; Teuben 2002)

6 Final thoughts

Simulations are suited for a wide range of problems and methods used are numerous. I hope that this paper has given some understanding of this and would like to conclude with a few final statements. The future of cosmological and astrophysical simulations is a bright one, since there is always another particle to add to a simulation and always a new set of parameters to be tested. The bottleneck will in the end always be on the hardware side and as the rest of society continuously strives towards smaller circuits and faster systems science will benefit. I really look forward to what the future holds in store for us.

Acknowledgements

Since this work is a result of a seminar workshop there are many persons that deserve attention. A good audience is worth a lot and during the progress of the work the other participants in the course have helped to bring a joy about what has been produced. There is however one person that really deserves to be mentioned by name. Alessandro Romeo, the organiser of the course "Hot Topics in Astrophysics" that this work is a part of. As a coordinator and supervisor he has with energy and determination given all the support and encouragement needed. I want to direct a sincere "thank you" to him for his great efforts and enthusiasm.

References

Hut P., 2002, in Astrophysical Supercomputing Using Particles, IAU Symposium 208, 2001 (arXiv:astro-ph/0204431)

Klypin A., 2000, arXiv:astro-ph/0005502

Max-Planck Institute: `http://www.mpa-garching.mpg.de/`

National Center for Supercomputing Applications (NCSA): `http://www.ncsa.uiuc.edu/`

Spurzem R., Kugel A., 2002, arXiv:astro-ph/9906155

Spurzem R., 2002, arXiv:astro-ph/9906154

Teuben P., 2002, in Astrophysical Supercomputing Using Particles, IAU Symposium 209, 2001 (arXiv:astro-ph/0204470)

Stability and Dynamical Evolution in Galactic Discs

Fredrik Janson

Chalmers University of Technology
SE-41296 Göteborg, Sweden
(f94frja@dd.chalmers.se)

*

Abstract

This paper gives an introduction to disc galaxies and their properties, particularly on stability issues. Spiral structure, and a new idea in secular heating in galactic discs are also described. Thermo-dynamical properties of open, far from equilibrium, many particle systems are also discussed in relation to the idea.

1 Introduction to disc galaxies

When one look upon the galaxy, one finds that galaxies, like so many other things in nature, come in different sizes and characteristics. But, there are a few number that specifies a middle ground for the collective of galaxies. An ordinary galaxy contains a total visible mass of about about one hundred thousand billions solar masses, or 10^{11} $M_\odot$. (1 $A_\odot$ means that that A has the same value as the sun for A) It has a radius of 10 to 20 *kpc* and shines with a luminosity, or power, of about 10^{11} $L_\odot$.

A galaxy is thought to consist mainly of three forms of matter

- Stars
- Gas
- Dust

If course there are other things like different forms of large solids like planets and asteroids, but compared to the other three they are often insignificant.

*Hot Topics in Astrophysics 2001/2002, Alessandro B. Romeo & Eva S. Karlsson (Eds.), Chalmers University of Technology and Göteborg University, 2002.

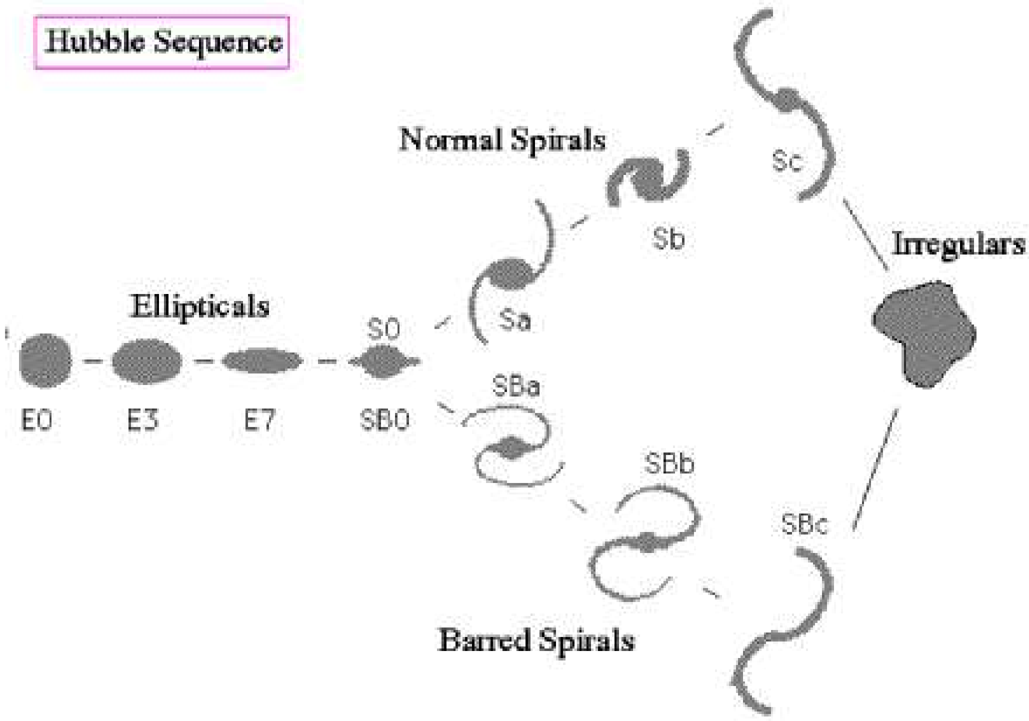


Figure 1: The hubble classification scheme.
(Figure taken from www.dur.ac.uk/Physics/students/handout3.html)

1.1 Hubble classification galaxies

There have been many different attempts to classify galaxies, but perhaps the most well-known is the one originally proposed by the astronomer Edwin Hubble. Hubble looked on the morphological features of galaxies. As seen in figure 1.1 the Hubble scheme has 3 distinct classes.

- Ellipticals

  Unlike indicated the ellipticals are 3 dimensional ellipsoids. the reason for this can be understood when looking at a photographic plate with an elliptical on, realising it is hard to find the the features that implies that it is really a 3 dimensional object. An elliptical galaxy i designated like E7 , where E gives that it is elliptical and 7 in a measure of the eccentricity.

- Irregulars

  The Irregulars are galaxies that can not be included in any other class, and are designated Irr.

- Spirals

  These galaxies are divided into two subclasses; Spirals and Bared Spirals. They consist mainly of a nearly 2 dimensional disc, often showing a spiral pattern. A bared spiral can be said to have had the central part of the spiral elongated. The Spirals are designated S and the Bars SB.

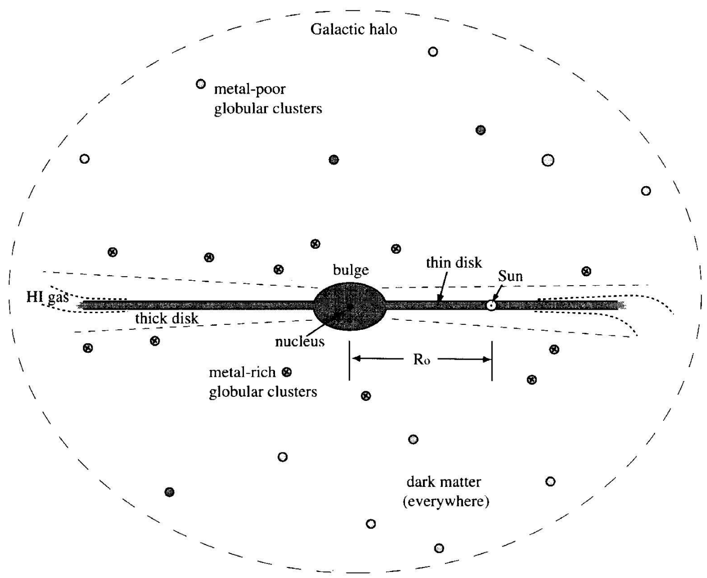


Figure 2: A picture of a disc galaxy.

Figure 3: Figure is taken from Sparke et al.

1.2 Structure of disc galaxies

The mass of a disc galaxy is mainly distributed in three large parts, the disc, the bulge and the halo.

1.2.1 Scale-lengths

Unlike countries, a galaxy does not have a sharp, well-defined boundary. Instead, the concept of scale-lengths is used. A galactic disc is usually brightest at the centre and asymptotically falls of exponentially, the scale-length, r_d is introduced to model this like:

$$I = I_0 e^{-\frac{r}{r_d}} \tag{1}$$

1.3 The bulge

This component of the galaxy is located in the centre of the galactic disc as seen in picture 1.4. It shares many of the same characteristics as with an elliptical galaxy and unlike the galactic disc, it's spherical.It has a a scale-length of about $1-2$ *kpc*.

1.4 The disc

This part of the has an average scale-length of about $3-5$ *kpc*. Unlike the galactic bulge, this part is nearly 2 dimensional with a scale-height of $300-400$ *pc* for the thin disk, which consist of most of the stars and all the young stars in the galaxy. The gas and dust lies in an even thinner disc of about 100 *pc*. There also exist a thick disc, containing older, metal poor stars, with a scale-height of about 1 to 1.5 kpc.

1.5 The galactic halo

This mysterious part of the galaxy, is nearly alway made up of dark matter and has consequently never been observed, except in a few cases. The reason to believe in them, comes from observing the velocities of the orbits of star in galaxies, which can not be sustained without this part which lies outside 10 kpc and is often assumed to lie in a nearly spherical halo. Research about the dark halo is subject to intensive research today.

1.6 Orbital velocities and the epicyclic approximation

The stars in the galactic disk have been predicted and observed to be moving on nearly circular orbits. However, it takes an average disc star a long time interval for one revolution around it's galaxy. Our sun, for instance, which lies about 8 kpc from the centre of the Milky Way, has a velocity of about 200 km/s but will not come back to this neighbourhood again for about 250 Myr.

Although the orbits are nearly circular, there exist a certain random motion to the orbits in question. To model this, the astrophysicists use one of the most used tools in physics, the harmonics oscillator. The star will oscillate around it's orbit with a frequency called the *epicyclic frequency* κ. For the approximation to hold the amplitude must not be too large. The epicyclic frequency can be used in several ways, (see Binney & Tremaine, 1987; Sparke, 2000).

2 The collisionlessness of galaxies

To model the dynamics of galactic disc one can choose to model the stars of galactic disc as either particles or as a fluid. While the later is done more easily, it can not predict everything that a particle treatment can. A full treatment is a lengthy process and would not fit into this paper, and the interested reader must instead be referred to other texts. (See Sparke (2000) for an introduction and Binney (1987), Bertin (1995), (2000) and Shu (1992) for details).

It can however be easily understood that the medium of galactic disc is to a good approximation collision-less. By doing a standard calculation of how often a star collides with other stars, it is found out that the time between strong near collisions, $t_s \sim 10^{15} - 10^{16}$ years. More refined collisions for weak distant collisions gives $t_d \sim 10^{13}$ years but this is at least comparable to the age of the universe which in modern cosmology is thought to be about $10^{10} - 10^{11}$ years. One conclusion that can be drawn from this is that the motion of stars in galaxies can to good approximation be calculated with a smoothened potential.

By using the results above, the time for a galaxy to reach thermal equilibrium can be calculated. This time, the *relaxation time*, is calculated to be $\sim 10^{13}$ years if reasonable values is used in the calculations. This is longer than most estimates of the age of the universe and we must conclude that at least most disc galaxies is not in thermal equilibrium, only virial.

2.1 Virial equilibrium

When a system has reached thermal equilibrium, a balance between the different energies has been reached. In virial equilibrium, there is instead a balance between the forces.

This is a result from the virial theorem, one of the most important result in dynamics, The virial theorem states that

$$\left\langle \frac{\partial^2 I}{\partial t^2} \right\rangle = 2\langle K \rangle + \langle U \rangle \tag{2}$$

Here I is the moment of inertia, U the potential energy and K the kinetic energy. The $\langle\rangle$ brackets means that it is the time average of the quantity which is in question. When the left hand of the equation becomes zero, virial equilibrium has been reached (see Binney & Tremaine (1987) for details).

3 Negative heat capacity of gravitational systems

Another striking feature of not only galaxies, but all gravitational systems is that they have a negative heat capacity. When there exists a steady state we have virial equilibrium. From the previous section we have that the kinetic potential and gravitational potential will be related as

$$2K - W = 0 \tag{3}$$

Then, the energy will be $E = -K$ and since K is made up from positive factors only the heat capacity must be negative. This will result in a seemingly paradoxical conclusion: *By losing energy, the system becomes hotter.*

4 Theories of spiral structure

There have been many theories of spiral structure in galactic discs. At first there was suggested that the arms were material arms. But unless these arms are rigid, they will wind themselves up in a short while, due to differential rotation. This is known as the winding problem.

The more modern is that stars are *traffic jams of stars*, however due to the winding problem, which effects has lessened, it is not completely understood how the spiral structures can be maintained for long enough periods. The most commonly used ways to understand spiral structure at present is the WKBJ approximation and the Lin-Shu theory. Both have been successful in explaining some properties of spiral structure but the complete and accepted explanation has not yet been proposed.

- WKBJ

  The WKBJ approximation is a perturbation method that has it's origin in quantum physics. To use this method the spiral arms must be tightly winded, meaning that the angle between the spiral arm and a circle around the galactic centre must be small.

- Lin-Shu

  This theory proposes that the spiral arms are made up of a quasi-stationary density wave that is stable over many rotations. The wave is maintained by amplification and feedback to strengthen the wave and dissipation to weaken it.

4.1 The role of resonances

Since there the stars have differential rotation and spirals are maintained, there is at one location , the *Co-rotation radii*, where the angular velocity of is equal to the angular velocity of the spiral pattern, Ω_p. This angular velocity is called the *pattern speed*. Standard derivations show (see Binney & Tremaine, 1987) that at this location there exist a resonance between the wave and the stars.

There also exist other resonances, the so-called *Lindblad Resonances*. These occur at radii where $\Omega = \Omega_p \pm m\kappa/2$, Where m is natural number. If the sign is negative the resonances are called *Inner Lindblad Resonances, ILR,* and if it is positive we have *Outer Lindblad Resonances, OLR,*. The density wave will treat these resonances as boundaries for the stars but not the gas. This is due since the stars are collision-less and can lie longer on the resonances while the gas collides and lies in the resonances very shortly, and then do not feel as large an effect as do the stars.

5 Stability

To determine if an equilibrium state maintains a spiral or bar structure for a long time, one has to look at its stability.

5.1 Dispersion relations

Since we are dealing with waves, it is natural to look at the dispersion relation. Such an equation determines the relation between the wavenumber and the frequency.

Depending on if one uses the fluid or particle description and other approximations, the dispersion relation will look a bit different, but asymptotically it will look like

$$(\omega - m\Omega)^2 = \kappa^2 + k^2\sigma^2 - 2\pi G\Sigma|k| \tag{4}$$

Here ω is the frequency, κ the epicyclic frequency, k is the wavenumber, σ is the velocity dispersion of the medium and G is Newton's constant of gravity. The term $m\Omega$ is due to an eventual periodic disruption in the medium like spiral arms or bars.Ω will be the angular velocity and m will be a natural number of how many times per revolution the medium will be disturbed, ie for a two armed spiral, $m = 2$. As can be observed the two first terms on the right hand side, which both are due to rotation and pressure, tend to stabilise, while self-gravity tends to destabilise the system. Also note that the famous Jeans length can be calculated if the pressure and rotation terms are set to zero.

5.2 The Toomre parameter

It is convenient to have a set of parameters of that by describes the state of the system. For the stability of the galactic disc, this parameter is the so-called Toomre parameter Q:

$$Q = \frac{c\kappa}{\pi G\Sigma} \tag{5}$$

The Toomre parameter also acts as a measure of the 'temperature' of the galactic disc, since c is the velocity dispersion. This can be understood if we remember that in kinetic gas theory, a large movement of the gas particles is the same as a large temperature.

5.3 Spiral structure as a faster way to thermal equilibrium

Introductory thermodynamics and statistical physics usually deals with isolated systems that reaches thermal equilibrium with an evolution in entropy that leads to an increasing degree of macroscopic uniformity. However, for systems like disc galaxies that lies far from equilibrium, this path is unfavoured and unstable and instead other solution branches presents itself. Here large structures that are highly organised temporally. Since these branches in certain aspect are similar to equilibrium phase transitions, these structures are called *dissipative structures* since dissipation play a large role in their maintainance, and their formations for *non-equilibrium phase transitions.*

The formation of the structures are induced by amplifications of fluctuations and feedback. In a realistic system there exists a fluctuation around the normal thermodynamical branch. For a system a certain parameter (in the case of galaxies the Toomre parameter) will have the role of determining if global amplification/feedback or local dissipation is occurring, giving rise to a stabilisation of the fluctuations.

The structures also have another function, namely to greatly increase the evolution on the entropy towards reaching thermal equilibrium. The coherent structure, having a low entropy, will maintain a constant entropy by exporting the produced entropy to it's surrounding environment. See Zhang for a further discussion and references.

6 Secular evolution of galactic discs

While the the galactic medium is often treated as a collision-less fluid, collision do more occur frequently than one could think. That here exist a thermal distribution that varies with the age of the stars in the galactic disc is proof of that there indeed must occur some collisions. Due to the nature of the gravitational force, each star "collides" with the rest of the stars all the time. When collisions occur, circular velocity is transformed into random velocity. If one then remembers our previous analogy between heat and movement, one can say that a dynamical friction has been included.

There exist several ways for stars to increase their heat. Stars most often collide with other stars or with giant molecular clouds of gas and dust, but these are all individual collisions, do there exist any mechanisms for collective effects, between the individual stars and the density wave.

Previously it was believed that angular momentum exchange between stars and the density waves can not occur, except at resonances. There exists several such mechanisms; At The Lindblad resonances there exists Landau-dampening processes and at co-rotation We have the WASER and the SWING mechanisms. These process are sometimes quite intricate and the precise description lies outside this paper (see Binney & Tremaine, 1987; Bertin, 1995, 2000 or Shu, 1992 for details).

However new results indicates that there can exist a collective effect that is dissipative and turns orbital energy into heat. This is such a process that has been described in section 5.3. Recently rediscovered solutions provides such a mechanism. It turns out that for galactic discs there exist a phase-shift between the mass and the potential distribution and interaction between the distributions provides the mechanism. For further descriptions see Zhang (1996), (1998), (1999).

References

Bertin G. & Lin C.C., 1995, *Spiral Structure in Galaxies*, MIT Press, Cambridge, Massachusetts London England.

Bertin G., 2000,*Dynamics of Galaxies*, Cambridge University Press, Cambridge.

Binney J. & Tremaine S., 1987,*Galactic Dynamics*, Princeton University Press, Princeton.

Romeo A., 1994, A&A 286, 799.

Sparke L. & Gallagher J., 2000,*Galaxies in The Universe*, Cambridge University Press, Cambridge.

Shu F., 1992,*The Physics of Astrophysics, Vol II, Gas Dynamics* University Science Books, Sausalito, California.

Zhang X., 1996, ApJ 457, 125.

Zhang X., 1998, A&A 499, 93.

Zhang X., 1999, ApJ 518, 613.