

OBSERVATIONAL ASTROPHYSICS

Contents

1 Basic astronomical notions	3
1.1 Introduction	3
1.2 Color index	3
1.3 Whitford law (or absorption law)	3
1.4 Metallicity indicators	4
1.4.1 X,Y and Z	4
1.4.2 The ratio [Me/H]	5
1.4.3 UV excess	5
1.4.4 Preston index	5
1.5 CCD sensors	6
1.6 The S/N ratio and the limit magnitude	8
1.7 Exposure time	10
1.8 Accuracy	11
1.9 Bolometric correction	11
1.10 Temperature and color index relation	11
1.11 From instrumental to international magnitude	13
1.12 Red-Leak effect	15
1.13 Second order interstellar reddening effect on photometry	16
1.14 Interstellar reddening maps	17
2 Measures of distance	18
2.1 Radar	19
2.2 Annual trigonometric parallax	19
2.3 Group parallax	20
2.4 Spectrophotometric parallax	20
2.5 Wilson-Bappu effect	20
2.6 Nebular parallax	20
2.7 Pulsating stars: RR Lyrae, Cepheids, W Virgins	21
2.8 Dynamic parallaxes	21
2.9 Novae and supernovae	22
3 Young stellar population	23
3.1 Color-Magnitude diagram	23
3.1.1 CMD of young population	24
3.1.2 Evolution towards the MS	25
3.1.3 Time-scale of evolution	25
3.2 The ages of young population	25
3.2.1 Two colors diagram	25
3.2.2 Other methods	26
4 Physics of planets	29
4.1 Basic equations	29
4.2 The processes of atmospheric loss	29
4.3 Atmospheric pressure distribution in hydrostatic equilibrium	31
4.4 Tidal forces	31
4.5 Electromagnetic emission, effective temperature, greenhouse effect	31
4.6 Generalities of the Solar System	34
4.6.1 Regularities and properties of the Solar System	34

4.6.2	Minor bodies	35
4.7	The formation of the Solar System	37
4.7.1	Evolution of a planet: the planet Earth	38
4.8	Dating the surface of terrestrial planets	39
5	Exoplanets	40
5.1	Research techniques	40
5.1.1	Direct imaging	40
5.1.2	Astrometric perturbation	41
5.1.3	Radial velocity	41
5.1.4	Photometric eclipses: the transit method	42
6	Supernovae remnants	44
6.1	SN remnants	44
6.1.1	Statistical results on supernovae remnants	45
6.2	Evolution of supernovae remnants	46
6.2.1	Free expansion	46
6.2.2	Adiabatic expansion	47
6.2.3	Isothermal expansion	48
6.2.4	Last phase	48
7	Maser	49
7.1	Molecular lines and maser emission in the galaxy	49
7.2	Emission of CO	49
7.3	Emission of OH	50
7.4	Maser emission	50
7.4.1	Three-levels maser mechanism	50
7.5	Stellar and interstellar masers	51
7.5.1	Distance measures with stellar maser	52
7.6	Usefulness of masers	52

Chapter 1

Basic astronomical notions

1.1 Introduction

Here are some important definitions:

- The sidereal time is defined as the hour angle of the gamma point and it grows with time since there is a difference of almost 4 minutes a day as a consequence of the motion of Earth around the Sun.
- Right ascension is the celestial equivalent of terrestrial longitude. It is measured from the position of the sun at the March equinox: the First Point of Aries which is the place where the Sun crosses the celestial equator from south to north at the March equinox and currently is located in the Pisces constellation. It is measured in hours, minutes and seconds.
- The hour angle is the angle between an observer's meridian and the hour circle on which some celestial body lies. When expressed in hours and minutes is the time elapsed since the celestial body's last transit over the observer's meridian.

The relation between sidereal time, right ascension and hour angle is:

$$ts = H + RA$$

[see notes for sections 1.2 - 1.4]

1.2 Color index

The color index is defined as:

$$c_{1,2} = m_1 - m_2$$

that is the difference in magnitude between two different passbands, that is two different filters. The calibration stars for the color index are A0V no reddened which are main sequence stars with H burning at the center and a temperature around 10000 K. Their spectrum is very similar to a black body. Moreover, at those temperatures, the Boltzmann equation tells us that the majority of chemical elements and molecules do not emit lines and the spectrum is dominated by the Balmer series so it's almost a continuum with few H lines.

There's only one color index that hasn't been calibrated on A0V: the **bolometric correction**, namely the difference between the visible band and the bolometric band (the one that includes the whole spectrum). Bolometric corrections are very important to pass from theoretical models to measurements.

1.3 Whitford lax (or absorption law)

Analyzing the spectrum of binary systems we understood that the fixed lines not subjected to the Doppler effect are caused by interstellar absorption by gas that creates very narrow lines. The main cause is the dust, made up of particles of carbon, iron, silicates of different shapes, combinations and orientations. If we look in the direction of the GC, the absorption is much bigger than if we observe perpendicularly to the galaxy plane. In particular, around the Sun the absorption is $A_V = 1$ for 1 kpc, that is 1 magnitude for 1 kpc in the direction of the GC so the intensity is reduced by a factor of 2.5 on 1 kpc in average.

The particles that cause the absorption in the interstellar medium have the same size or are bigger than the wavelength in the visible band.

In general the absorption law is:

$$A_\lambda = 1/\lambda$$

known as Whitford law valid in first approximation, from visible to infrared band. It is an empirical law and shows that absorption is more intensive at short wavelengths (more important in the blue band compared to the V one or infrared).

This law has been obtained using mainly the early-type stars because they're brighter, with less absorption lines in the spectrum and very well studied. The absorption law for other spectral type could be different.

The interstellar absorption is often estimated using the **color excess** defined as the difference between the observed color index (photometry) and the intrinsic one (spectral type) in absence of absorption. For the B,V band the color excess is:

$$E(B - V) = (B - V) - (B - V)_0$$

It is the consequence of the inverse scale of the magnitude. E stands for excess and $E(B - V) > 0$ because the absorption is bigger in the B band than in the V one and the difference between B and V increases because the reddening law is not uniform.

If $(B - V) = (B - V)_0$ that means $E(B - V) = 0$, the color index observed is equal to the intrinsic color index so there is no reddening.

Connected to this equation there is an important relation between the absorption and the color excess:

$$A_V = 3.1 E(B - V)$$

The constant 3.1 is very important but also very delicate to establish. If the number is not correct than the distance is wrong also of various orders.

The clear result of the interstellar absorption on the spectrum is the reduction of the flux of the source, respect to the continuum: an effect that is bigger at shorter wavelengths, becoming negligible from far infrared to the radio regime. A spectrum affected by absorption appears with a different slope than the original one. The color index grows and the peak of the continuum is shifted to bigger wavelengths: the spectrum is similar to the spectrum of an advanced spectral type star. However lines remain in the same position, the equivalent width remains the same and the spectral analysis allows to recognize the original spectral type of the source (and doing so, the original color index). Indeed spectral lines are quantum transitions so there is no way the lines could change their positions.

The reddening effect doesn't have to be confused with the Doppler effect: the latter causes a shift in the x-axis of the whole spectrum, lines included, to bigger wavelengths for objects that are moving away.

The absorption law does not include the corrections for the Earth atmosphere. This correction depends on the height of the object above the horizon, on the direction of observation, on the airmass and the aerosols. The average correction in the visible band is 0.15 mag per airmass.

We can say that the reddening is a consequence of the Whitford law: stars appears redder than their intrinsic color when their light passes through the interstellar medium.

1.4 Metallicity indicators

Stars contains mainly H, an important fraction of He and other metals. We can measure their metallicity in two ways:

- number of observed atoms, using spectroscopy that measures the abundance of chemical elements
- fraction of mass, theorists' preferred method for models

1.4.1 X, Y and Z

We can indicate the fractional abundance of hydrogen, helium and metals using the letters X, Y and Z. By definition:

$$X + Y + Z = 1$$

The original cosmic abundances are $X = 0.78$, $Y = 0.22$, $Z = \text{traces}$.

Fractional abundances are very important but not so easy to find because we can look only at the surface of stars,

above the photosphere, that can't be representative of the whole star. In general chemical elements are in equilibrium but the different nuclear reactions that take place during the evolution of stars could change the elements distribution. Moreover there's a particular phenomena that can occur: the **elements segregation** (or **distribution**). It consists on the separation of the heavier elements in the inner part of the star while the lighter ones remain above.

1.4.2 The ratio [Me/H]

Another important metallicity indicator is the ratio:

$$[\text{Me}/\text{H}] = \frac{\log(\text{Me}/\text{H})}{(\text{Me}/\text{H})_0}$$

It indicates the logarithm of the metal abundance (Me) with respect to H, normalized to the Sun. Usually it is used the iron like reference for the other metals because its abundance is easy to measure.

For example, if $[\text{Fe}/\text{H}] = -1$, the star has a content of metals 10 times lower (because of the logarithmic scale) than the Sun in atoms number. Otherwise if the ratio is equal to +1, the metal content is 10 times higher than the sun. If $[\text{Fe}/\text{H} = -2]$ the content is 100 times lower and so on.

In general the process is

- measure the absorption lines from the spectrum of the object
- get how much is the metal abundance
- extrapolate the total abundance
- check the results with theoretical models

However this implies an hard work for many reasons:

- it is time consuming
- you must use high resolution spectroscopy
- you need to measure the stars one by one

Moreover it is a long process and implicates the knowledge of the temperature of the star because the intensity of the lines depends on Boltzmann and Saha equations and both depends on temperature. So to know the metal abundances we need to know the precise temperature.

1.4.3 UV excess

For very faint stars for which isn't possible to make a precise spectroscopy we need to introduce another indicator: the UV excess.

Historically this indicator has been defined as the color index U - B, at fixed B - V with respect to the Hyades:

$$\Delta U = (U - B)_{\text{Hyades}} - (U - V)_{\text{obs}}$$

The Hyades are an open cluster near the Sun. They're Sun stars (population I, young in the disk) with very low reddening and same spectral type. They're reference stars with $\Delta U = 0$.

This indicator can be interpreted as the **blanketing effect** due to the presence of metals: the abundance of some metals produce a reduction in the U band, thereby U magnitude and color excess increase their values. In other words for metal poor stars (like population II) ΔU is higher. In general, for metal poor stars, $\Delta U > 0$.

1.4.4 Preston index

There is another index calibrated for variables RR Lyrae, stars of F type with well defined metal lines. It is defined as 10 times the difference of the spectral classes obtained from the lines of the Balmer series compared to that obtained from the lines H and K of CaII.

$$\Delta S = 10[SP(\text{H}) - SP(\text{CaII})]$$

This is valid only for spectral classes A5 - F5 ((B - V)=0.15 - 0.45).

RR Lyrae have been chosen because they're very bright, easily identified, variable with small periods. Due to their nature they're visible even at big distances so they're also perfect standard candles to measure distances.

For fainter stars, for which we have low resolution spectra, it is valid:

$$[\text{Fe}/\text{H}] = -0.23 - 0.16\Delta S$$

In general it is important to have a spectral classification, for which you just need low resolution spectra, easily obtained also for fainter stars while for abundance analysis you need high resolution spectra.

Application of UV excess

The UV excess has been used to discover the population II stars. As you can see from Fig.1.1 in the first graphic,

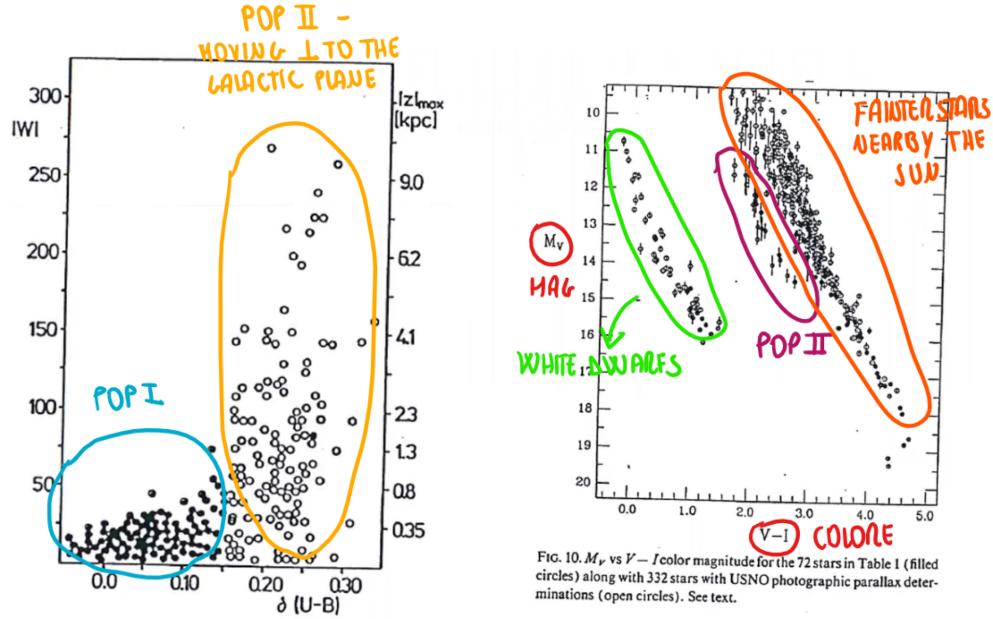


Figure 1.1: Evidences of population II stars

population II stars are metal poor and their cinematic index indicates that they're moving perpendicular to the galactic plane. In the second graphic, that is a CDM, we can observe different stars:

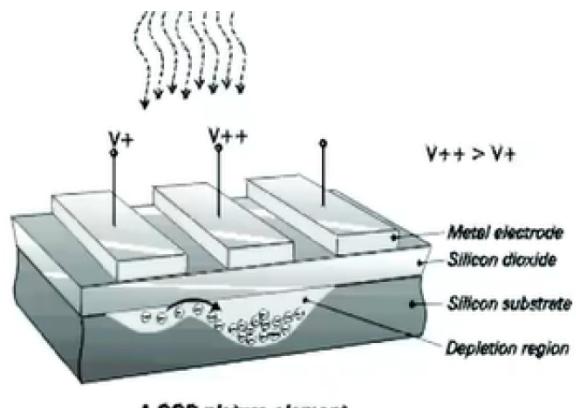
- the white dwarf group: last phase of the evolution of stars, often rich in carbon and which are cooling down
- on the upper right we have some faint stars, very common nearby the Sun, also called "normal stars" because they're burning H in the inner part. Their distance d has been measured using the parallax method from ground
- the branch in the middle is the branch where we can find population II stars. They're "intruders" in the galactic disk and they're a minority. They're main sequence stars moving perpendicular to the galactic plane. Because of their lower metallicity, they follow the main sequence phase parallel to the fainter stars, so they appear more blue. This type of stars are 100 time less rich in metals with respect to the Sun and they're shifted for 2/2.3 in color index (x-axis) so they have higher temperature

1.5 CCD sensors

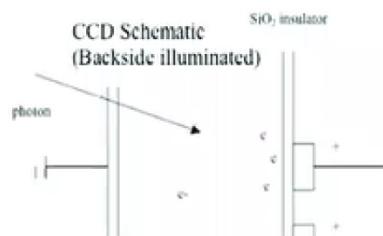
Nowadays scientist can use two types of sensors: CCDs and CMOS.

The CCD works as illustrated in Fig. 1.2. It is composed by an unique piece of silicon equipped with electrodes that create electric fields. During observations, photons from the source hit the silicon and produce electrons below the silicon substrate (using the energy given by photons, electrons pass from the ground state to an excited state).

In this phase electric fields are fixed and electrons are collected below electrodes, confined in a limited area. At the end of the exposure, positive electric fields conduct the electrons in specific directions and move them out the visible fields. If a photon hits the sensor on the border between two pixels, it isn't lost: it goes in one of the two pixels depending on its position.



A CCD picture element



CCDs IN ASTRONOMY 259

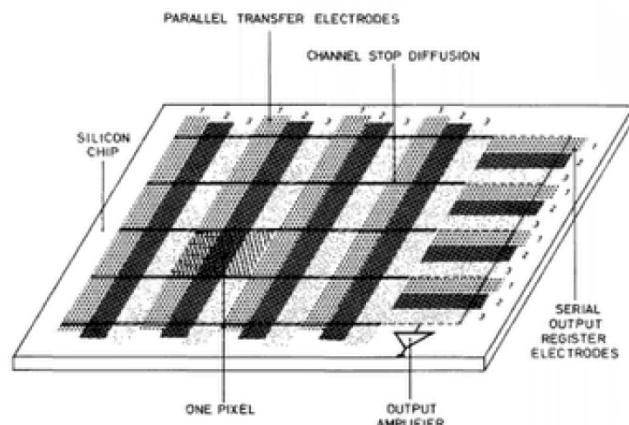


Figure 2b The basic layout of a three-phase two-dimensional CCD. The sequence 1, 2, 3 on each set of electrodes indicates the normal direction of charge transfer in the parallel and serial registers.

Figure 1.2: Functioning of a CCD

Sometimes, at very bright illumination level, there are so many electrons collected under the electrodes that creates a big negative field invading the pixels nearby (in any case they aren't lost). This is called **blooming effect**.

Then there is the read out phase. Electrons are counted and the signal is amplified by an amplifier. This phase takes some time and also some electrons can be lost. In any case the amplifier is only one so, with a good construction of the silicon substrate and electrodes, the quantum efficiency is almost uniform.

Modern CCDs' efficiency is about 90%. In particular the peak of efficiency is shifted a bit on the red side so real peak is between 600 and 700 Å while the peak of the visible is at 550 Å. The difference is not too big but there's some slope to take into account.

The efficiency drops down in the UV band where there are some reflection effects due to the silicon substrate. In general it is necessary to take into account that old photography plates, on which many international systems are calibrated, have the peak of sensitivity in the blue band so they're almost flat in the optical range.

CMOS detectors works differently since they amplify the signal from pixels one by one and then the amplified signal is sent to the electrical connections. This requires more electric power so the CMOS is way less uniform than the CCD even though is faster in read out phase. For these reasons CMOS is used for fast photometry, adaptive optics and other special uses.

CCDs are very stable also because the efficiency depends on the capacity to transform photons into electrons, that gives the necessary energy to electrons to make a bump from ground state to an excited state. However this amount of energy is fixed by quantum transformations (quantum levels are extremely stable) o the CCD is an extremely stable device.

1.6 The S/N ratio and the limit magnitude

When light comes from an astronomical source, it crosses the atmosphere where it is subjected to absorption, deviations and scattering effects. Here 15% is lost. Then it passes throughout the optical train (the telescope). Here there is the most important loss of photons: about 55% of them is lost, due to the reflections of the mirrors, filters and others. Finally only the 30% of the incoming photons are detected, even though the sensor efficiency is about 90% in the visible band.

The signal registered by the detector is given by:

$$S_{tot} = s \cdot t$$

where S_{tot} is the total signal collected, s is the signal per unity of time and t is the exposure time. From this equation we can observe that the signal goes linearly with time, due to the linear trend of the detector. However during the exposure and the reading out there are different sources of noise, defined as:

$$N = \sqrt{S} = \sqrt{s \cdot t}$$

so

$$N^2 \propto t$$

The most important quantity to establish is the ratio S/N .

Therefore depending on the equations above, for a fixed source, going to an infinite exposure time the ratio S/N goes to infinity either. So from a mathematical point of view you can compensate the size of the telescope by increasing t . In general the study of this ratio is a fundamental element planning for the observations. The dissertations about S/N depends on noise nature and, so, on detector characteristics. Then the limit magnitude depends on the ratio S/N . In this case we will treat only stellar images, relatively small with respect to the sensor size and for which stars occupy only a few pixels.

Assuming a strong linearity og the signal as function of incident photons, there are different sources of noise:

- Read Out Noise (RON) during the read put phase of the signal
- Dark Noise or thermal noise
- statistic Poisson noise (shot noise) of the signal and of the sky, which is totally unpredictable
- spatial noise or pattern noise (SPN)

RON - The Read Out Noise is contained in a set of ten pixels so it is negligible for deep images r for bright objects with high S/N . Instead it is important in case of images with low sky level and low signal, for example in spectroscopy or for small exposure time. In general it is almost constant so it is a fixed number for every device.

Dark Noise - Also dark noise is negligible in many modern CCDs. It is caused by the functioning electronics that produce some casual jumps of electrons from ground state to conductive state. they have a Maxwellian distribution of velocity that depends on temperature.

Many CCDs are cooled down to very low temperatures in order to keep all thermal electrons in very low level.

Shot noise - This is the most important contribution of noise. For its nature it varies with the square root of the signal.

In general astronomers observe very faint objects and they have always to take into account the sky background. If the source is comparable to the sky there can be several problems.

In general the sky can be removed if it is known, however any mathematical operation increases the noise N because you don't know exactly the sky level so you use average values and then you can extrapolate it. So if you try to remove the sky background the noise grows up.

In case of a very strong source, the sky level is negligible because the intensity registered is dominated by the source intensity. Instead for fainter stars, the signal is dominated by sky background.

So it is valid

$$N = \sqrt{S + SKY} \quad N^2 = \sigma_{shot}^2 + \sigma_{sky}^2$$

and for very faint stars, for which the signal S is negligible compared to the sky level so

$$N = \sqrt{SKY} = \sqrt{sky \cdot t}$$

where sky is the intensity of the sky level and t is the exposure time. Usually SKY is a fixed quantity.

So in this case, the S/N ratio is:

$$S/N = \frac{s \cdot t}{\sqrt{sky \cdot t}} = \frac{s}{\sqrt{sky}} \cdot \sqrt{t}$$

If $t \rightarrow \infty$ we have that $S/N \rightarrow \infty$ independently from the size of the telescope. Therefore in general the sky and the telescope aren't such a barrier to the sensor capacity to reach the deep sky. Indeed you can observe sources much fainter than the sky background if the exposure time is long enough in order to get the signal S high enough compared to the noise N .

So the real barrier to reach a very high S/N is the SPN.

SPN noise - SPN, also called spatial noise or pattern noise, is the real limit in observations to get a very high S/N and is specific for each sensor.

Even though the CCD is almost uniform (uniformity is almost constant with time) there are some small sensitivity variations from pixel to pixel. Therefore the quantum efficiency changes from pixel to pixel as well the scale background and the scale background and the noise profile of a star is affected by the different sensitivity of pixels.

Mathematically, considering also this type of noise and knowing that all these contributions are independent, the total noise is the square root of the sum of all contributions:

$$N^2 = A \cdot RON^2 + A \cdot sky \cdot t + A \cdot Dark \cdot t + A \cdot s \cdot t + (SPN \cdot (sky + s) \cdot t)^2$$

where A is the area occupied by the source on the sensor ($A = \pi \cdot FWHM^2$). Then S/N becomes:

$$S/N = \frac{s \cdot t}{\sqrt{sky \cdot t + const}}$$

where $const \propto t$. So S/N cannot increase. You can compensate this noise taking some **flat fields**, calibrations files taken before or after the observation, realized illuminating the dome or a background surface uniformly in order to obtain sensitivity distribution, pixel by pixel. These calibrations are useful to reduce SPN but not to remove it because typically you're exposing at different wavelengths so a white light on the dome is not the real light from a star or of the sky background.

Therefore SPN is the real barrier: a 1% of this kind of noise means that the faintest star is about 100 times fainter than the sky background. In terms of magnitude, for ground based observations, sky background is about $V \simeq 21.5$ SO with a factor of 1% the limit magnitude is 27. For a factor of 2% - 3% the limit magnitude is 26 - 25. From space $V = 23$ so with a factor of 1% the limit magnitude is 28, not too different from ground based observations.

There are some techniques to obtain a better correction of SPN:

- **Drift scanning** - Usually, after an exposure, during the reading put phase, there is a shutter that stops the accumulation of photons from the source. In order to smooth down the variations from pixel to pixel it has been created a device that don't close the shutter during the read out and makes the read put very slowly. At the same time the CCD moves with the same speed of the reading put but in the opposite direction respect to the telescope. In this way the image increases in brightness and different pixels with different sensitivity are interested so statistically, variations from pixel to pixel are smoothed down. With this method we can reduce SPN from a factor of 1% to 0.1%, which means a limit magnitude of 28.

However there are some problems with this method. First of all there is no uniformity in S/N since you smooth down SPN only in one direction (the direction of the reading out) so you obtain a very flat image in a direction while on the other direction SPN is 10 times higher. For this reason the image presents some vertical strips.

- **Differing** - This technique consist in taking an image of the source, then a second one a little bit shifted and so on. To obtain the final image is necessary to re-shift the images taken before. In this way different pixels are exposed and statistically the variations between each pixel are smoothed down. In this way you can gain a factor 2 on the SPN, depending on the number of shifted images.

Of course there are some disadvantages. First, the common area is smaller than the full field of the sensor. Second, you miss a lot of observation time reading out the images and storing them on the computer. Nowadays this method is the most used because it is easy and you can reduce SPN more or less depending on the target, the scientific goals etc.

1.7 Exposure time

It is important to find the right exposure time. Suppose a telescope observing a very faint star which intensity is comparable to the sky level, for which is valid:

$$S = s \cdot t \Rightarrow S \propto t$$

$$N = \sqrt{sky \cdot t} \Rightarrow N \propto \sqrt{t}$$

In this case $S/N \propto \sqrt{t}$ and N has a parabolic trend as function of time t while RON is a constant. SPN is an additional noise, proportional to the intensity of the star or the background and linear to the time t , indeed it could be represented as a straight line. This is not a random noise but a spacial pattern that represent variations in sensibility from pixel to pixel.

For a relative short exposure time, SPN is considerable lower than the Poissonian noise while with very long exposures the SPN dominates. So the noise changes from a random trend to a sort of fixed pattern. This is the ultimate limit imposed to observations.

You can correct SPN partially but you can't remove it. Moreover sometimes corrections could worsen the situation.

Suppose we have a 1 hour long exposure, which means a factor 1000 for the sky background. Then suppose to correct this factor with a flat field of the same factor. The operation is:

$$\frac{1h \rightarrow \sqrt{1000}}{FF \rightarrow \sqrt{1000}}$$

This operation implies the propagation of errors so it doesn't reduce the noise but it increases it. You have to go to very high counts if you don't want to degrade significantly the S/N . Of course, to use a flat field with a very strong level of intensity, you have to be sure that the detector is extremely linear.

1.8 Accuracy

Suppose to observe a very faint object and wanting the best S/N : what is the highest accuracy reachable? In general the best accuracy is 70 ppm. The main problem for the accuracy is the variable sensitivity of pixels. One technique you can use to solve this problem is **defocus** the image.

With a defocused image, the shape of the brightness profile of the star gets close to a cylinder, smoothing all the pixel variations.

Then there is another problem: the Poisson noise. To improve it you need to use more counts so we need a wider stellar image.

In general, to have enough electrons, which means enough photons from the source, you have to take into account how long should be the exposure time, If the telescope is not very large, it is necessary a much longer exposure time.

Noise connected to the area - Consider the area occupied by the intensity distribution of the source: it is like a gaussian. At first we can thin that the total area is $A = \pi r^2$ but that's not right because the wings of the distribution are negligible in photometry. So we could use the FWHM but in this way we underestimate the area, loosing an important fraction of light. Therefore the best way is to use the FWHM as radius:

$$A = \pi \cdot FWHM$$

1.9 Bolometric correction

In observations it is necessary to pass from measure in some band to bolometric magnitude m_{bol} , the magnitude corresponding to all the flux of the star from far radio to X. Combining some spectra and applying some corrections you can easily extrapolate the black body profile and, after an integration, the bolometric magnitude.

It is very important to know m_{bol} and even the relative correction, the **bolometric correction BC**.

The transformation between visual magnitude into luminosity and vice versa is done through the BC that allows to pass from optical magnitude to the bolometric one.

$$m_{bol} = V + BC$$

So the BC is:

$$BC = m_{bol} - V$$

which means it is a color index. BC is calibrated on F5V stars with 6600 K of temperature that correspond to an emission peak at the center of the visible band. Outside the normalization point, the BC assumes always negative values, bigger going to higher and lower temperature. Because BC assumes negative values, m_{bol} is smaller, which means brighter.

In general the BC depends on parameters of stellar atmosphere so it is sensible to the luminosity class and the metallicity of the star.

The BC is responsible of many effects visible in color-magnitude diagrams of evolved populations, auch as:

- Rapidly decrease in terms of luminosity of the horizontal branch to the blue limit. In this branch there are stars which burn He into C in the core and H into He in the surrounding shells. Here stars are so cold that relevant fraction of energy is emitted outside the visible filter. So near the giant type the luminosity in the visible band is going down because BC increases.
- The growing slope of main sequence stars with lower temperature. In this area BC increases and turns down the MS line which would be instead much more linear with bolometric luminosity.
- The turn off in asymptotic branches of massive globular cluster rich in metals.

1.10 Temperature and color index relation

Now the problem is to pass form color index information to temperature information. In Fig. 1.3 x-axis is $\log \lambda$ while y-axis is $\log B$. In this graphic are visible black body profiles defined by specific temperature T and specific color index, in this case B-V. In particular, at lower T , the color index increase because of the magnitude scale. So color index is an indicator of temperature, until the emission peak is more or less inside the range of the wavelength filter used for photometry.

Color index (for ex. B-V) as indicator of stellar
photosphere temperature

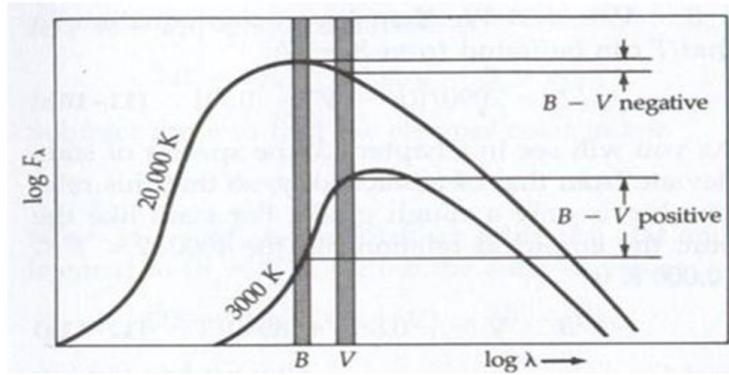


Figure 1.3: Color index as indicator of stellar temperature

Suppose to observe a star that is so hot that the color index is on the rising branch which means Rayleigh-Jeans approximation. Here, different black body profiles have almost parallel slopes so, if you are outside the emission peak, color index is almost constant, even if at different temperatures T .

This means that the color index is a good temperature indicator only close to the emission peak. If you are outside of it the slope is almost independent from T .

So what is the relation between the temperature and the color index? It is necessary to enter with λ_B and λ_V in Planck equation to get the difference between two planckian equations at different wavelength. Practically:

$$B - V = -2.5 \frac{\int B_B T_B d\lambda_B}{\int B_V T_V d\lambda_V} + C_{BV}$$

where the factor -2.5 is due to magnitude scale, B_B is the black body distribution and T_B is the transmission in B band and C_{BV} is a normalisation constant. So $\int B_B T_B d\lambda_B$ is the convolution of the black body with the transmission in the corresponding filter.

After some normalization and algebraic operations, you get the relation:

$$B - V = -0.865 + \frac{8540}{T}$$

So, in principle, measuring $B - V$ you obtain the temperature T and vice versa.

When T is very high, the second term goes to zero and it is eventually negligible, compared to the constant. The consequence of this is that the color index is no more sensitive to the temperature. If the peak of an object is outside the filter, one possibility is to change wavelength. The situation is quite complicate for lower temperatures for which there are some absorption due to the complex molecules in the atmosphere and you don't have anymore a black body.

So the basic concept is: the derivation of the temperature from color index must be studied depending on the type of star. In general we need a very high photometry accuracy in order to reach a very high accuracy on T in order to get a precise age or chemical composition.

Modern alternatives to color index - There is another way to determine a star's temperature which consist in using the ratio between absorption lines in the visible spectrum. Indeed their depth is related to the temperature for Boltzmann and Saha equations. Typically iron lines are used since they're common in stellar spectra and the temperature calculated in this way is called *excitation temperature*. It is not the effective temperature: one comes from observing the atmosphere while color index comes from the photosphere which is an opaque layer so they should be different somehow.

Temperature from the lines is derived assuming some constant values of energy at which the excitation occurs but they are not very well known. For this reason temperature from color index is still the most used.

Infrared and radio astronomy - If we go to larger wavelengths (infrared and radio) scientists never use the color index because they are in a regime where the slope is almost constant so color index isn't sensitive to the temperature. Radio astronomers measure the brightness at a specific wavelength so in the graphic in Fig. 1.3 they identify a specific point and we know that there is only a specific black body distribution that crosses that point since planckian functions never cross each other and every black body distribution corresponds to a specific temperature. So it is not important the shape of the black body but the value of the brightness.

The main problem is to get the brightness because they need the absolute brightness. The brightness by definition is not depending on the distance so the problem is not to know the distance. It is also affected by interstellar reddening but it is quite negligible since it is low at long wavelength. So the problem here is the following: the measure of surface brightness requires measurements of the flux and the angular size of the target.

$$B \propto \frac{F}{\text{solid angle}}$$

Usually the angular size is unknown. Only with interferometry you can get the angular size for giant stars. For other stars you don't have this information.

You may think that if you have the distance and the model of the star you can get the angular size but this is a very articulate way because you need some other parameters that you have to derive.

So in radio they don't measure stars but mainly extended objects like HI clouds, stellar formation clouds, supernovae remnants and so on.

1.11 From instrumental to international magnitude

In general the color index is used to derive the temperature of a star. However we should consider even the interstellar absorption effect: it changes the color index measured because of the absorption at different wavelengths.

Since the first international photometric system was established there has been an evolution of detectors such as the replacement of photographic plates with solid state detectors with a different spectral response.

CCDs have a more extended red response while their sensitivity goes down in the blue band. If we refer to the international system and we make a convolution of transmission of modern CCDs we can notice that there is a shift of the convoluted band to lower wavelength on the red part of the spectrum. So the sensitivity of the CCD is not flat. Instead, on old photographic plates the function of sensitivity is much more flat, especially on the blue side of the spectrum and it decreases at longer λ .

Therefore if we take the right filters but different detectors we get a different spectral sensitivity. To solve this we have some possibilities:

- choose a different international system
- change filters

Nowadays, to solve this problem, scientists observe the same object in different wavelengths. The consequence is that, when doing the calibration of data, we have to proceed through a process which is the conversion from instrumental magnitude to international system, depending on the specifics of the detector. So it is necessary to convert instrumental magnitude to the international one. For doing so we have to follow two steps:

1. First it is necessary to observe and measure a number of standard stars with different temperatures and then create a plot like the one in Fig. 1.4.

In this graphic V is the calibrated magnitude for that star (what we want to know), v is the instrumental magnitude (what we measure) and $B - V$ is the color index in the international system (available in catalogs). In general the instrumental magnitude is given by:

$$v = -2.5 \log C$$

where C are the counts obtained during observations with a specific detector (eventually normalized).

So v is negative for definition because counts of electrons are positive. Consequently $V - v$ is a positive number, generally a big one.

So for standard stars, if the system is right, the spectral sensitivity is identical to the international one, it is expected a constant plot which means a set of stars having $V - v$ constant (blue plot).

In particular $B - V$ about 0 corresponds to hot standard stars while $B - V$ about 1.4/1.5 corresponds to cooler stars. This is the saturation limit for this kind of stars.

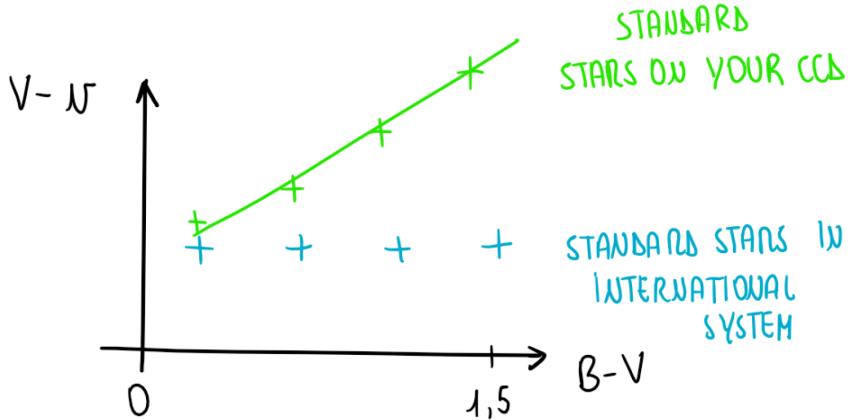


Figure 1.4: $(V - v)$ vs $(B - V)$

Now suppose that, instead, the sensitivity is pushing the average wavelength a bit shifted to the red (as in modern CCDs). More red photons are collected than in the international system. So a very red star could produce more flux in red than it is expected in international system. As consequence plotted stars create a straight line with a slope, smaller in the blue and bigger in the red range (green plot). Mathematically $v < 0$ because counts are growing so $V - v > 0$, growing as well.

In general this plot does not depend on brightness: it depends on color index, so on temperature. The slope indicates that in our instrumental system we have more flux than a given standard star with a specific temperature. If the system used is not strongly different from the international one, the green plot can be approximated with a straight line.

Recapitulating, the first step consists in measuring a number of standard stars, eventually spread in a wide temperature range. Then you have to create a plot like the one in Fig. 1.4 that shows how the sensitivity of the device is different from the international photometric system.

Then the following relation is valid:

$$V = v + K_V(B - V) + C_V$$

where V is the calibration magnitude (what we are searching), $B - V$ is the international color index on x-axis and C_V is a constant to normalize the equation. K_V is the angular coefficient of the straight line and indicates the slope, which is always positive. In general the order of magnitude for a relative good system is about 0.1 mag in an interval of 1 mag.

So in the case of a good system $V - v$ should be constant but if there are some differences from the international one $V - v$ grows and, with it, also the slope of the straight line.

2. Now the goal is to determine V for unknown stars which have been measured obtaining the instrumental magnitude v . Remember that $V = v + K_V(B - V) + C_V$.

K_V is known, derived from the plot for standard stars: it is the angular coefficient of the straight line. Usually is about 0.1 in 1 magnitude in color in x-axis so ignoring this term leads to an error of 10%.

C_V is the normalization constant: it is known as extrapolation to zero point.

However $B - V$, the color index calibrated on the international system is NOT known. It is known for standard stars but not for unknown stars. You should know this calibrated color index or eventually the temperature T to transform into the color index, or eventually the spectrum or something other that allows you to find where you should put the star on the x-axis in order to get the right correction.

From algebra it is necessary to find another independent equation in order to have a linear equation system you can solve. This equation is given by using a different filter for the same star with same color index so it is exactly the equivalent of the first one. We get the linear system:

$$\begin{cases} B = b + K_B(B - V) + C_B \\ V = v + K_V(B - V) + C_V \end{cases}$$

Subtracting the first equation to the second one we have

$$B - V = \frac{b - v}{1 + K_V - K_B} + \frac{C_B - C_V}{1 + K_V - K_B}$$

where $B - V$ is the calibrated color index, $b - v$ is the instrumental color index, K_V and K_B are the slopes of the straight lines in the two bands determined from standard stars, C_V and C_B are the zero point also determined from standard stars.

Knowing the color index calibrated on the international system you can find V and B , the calibrated magnitude.

The main consequence is that there is not only an offset of the instrumental one, there is also the factor $1/(1 + K_V - K_B)$. This term is very important since it makes the scale $B - V$ different from $b - v$ depending if it is smaller or bigger than 1.

- $1/(1 + K_V - K_B) > 1$ means that the scale of $B - V$ is expanded compared to the instrumental one. In this case $K_B > K_V$ because going to shorter wavelength the deviation from the original system is bigger, telling us there is a relevant shift in the band. In particular the scale of $B - V$ is larger and the scale of $b - v$ is compressed compared to the standard one. When they are extremely close, you miss completely the sensitivity because the difference is so small that you can't see differences in temperatures. In this case is necessary to go further away to increase the sensitivity.
- $1/(1 + K_V - K_B) < 1$ means that the scale $B - V$ is compacted compared to the instrumental one.

Instead $C_B - C_V / 1 + K_V - K_B$ is just the zero point, a constant that shifts the plot.

So analyzing the color-magnitude instrumental diagram, the calibration is not just a shift: it is also a change in scale because of the **color correction term** $1/(1 + K_V - K_B)$. In particular K_V and K_B tell us how much is the deviation of a specific system from the standard one. In a perfect instrumental system these terms should be zero so that the color correction is 1. In this case $B - V$ and $b - v$ scales would be identical and there should be only a simple shift in according to the difference of the zero points.

In general this correction is good but not perfect an it is just a model correction. Now it is necessary to take into account the interstellar effect and correct it. What we have done here is just the **first order approximation**.

There are several issues behind this simple process

- The derivation of the plot in Fig. 1.4 is very delicate and depends strictly on the choice of standard stars used for calibration. In principle standard stars should be the same kind of the stars you have to observe in order to have a similar spectral distribution, so a similar metallicity.
- Stars should have a comparable reddening (absorption at shorter wavelengths is bigger to the one at longer wavelengths). In this case the shape and slope are different from the intrinsic one. It is not just a scaling down, it is also distorted and the peak is quite different. Of course absorption lines are exactly in the same position. So, using a mixture of stars with different interstellar reddening and different kind of spectra you may have a spread of points, a dispersion around the straight line in Fig. 1.4. Of course, stars having higher $B - V$ have higher probability to be reddened stars.

ATTENTION: U band is strongly affected by Earth atmosphere because it is close to the visible limit. So, transmission in this band is affected by the atmosphere every night because of the aerosols changing the slope of calibration. Therefore the calibration in the U band should be repeated every night.

1.12 Red-Leak effect

A second reason for deviation from the standard system is the red-leak effect, defined as the contribution of radiation detected at a wavelength greater than that of the main band, typically between 8000 and 9000 Å, where transmission can be around at $10^{-3} - 10^{-4}$ with respect to the peak of the main band. This is a small value, but not negligible when comparing the photometry of stars with very different temperatures between them. The resulting effect of the red-leak is to make a positive contribution to the color term K .

From the technical point of view the red-leak derives from the difficulty of finding filters able to effectively cut the red-infrared radiation, while maintaining good transmission in the blue part of the spectrum.

In the past, copper sulphate (CuSO_4) has often been used because it completely blocks the radiation in the red, in the liquid state in a suitable bulb. But the liquid in this state is unstable and corrosive, while in the solid state the

crystals are hygroscopic so it absorbs the water present in the atmosphere and difficult to find in large formats. In modern filter systems today, layers with thermal reflectors against infrared are used, but these are not easily available.

The problem is more frequent than one might think, and it is also difficult to identify and quantify both in the laboratory and from experiments in the sky. However it must be carefully evaluated especially for the bands U and B . A very simple qualitative test to reveal the presence of a significant red-leak consists in examining the appearance of the flat field in the bands considered with respect to those of the extreme red. A certain resemblance, compared to intermediate bands, can make one suspect the presence of red-leak.

1.13 Second order interstellar reddening effect on photometry

We have mentioned the need to know distance and interstellar extinction (and/or reddening) for determining the ages of star clusters. The distance could be obtained by using as a reference the brightness of the horizontal branch that varies little with the age in the HR diagram where the turnoff point (sensitive to the age) is measured. The level of the horizontal branch (HB) is however sensitive to the helium content, a difficult parameter to determine, and to the metallic content.

The HB brightness follows a relation that has not yet been defined with the precision required for measurements of ages within 10% in the metallicity range from the solar one to the lowest metallicity of the halo (about 100 times lower). The picture is further complicated by the role of the excess of alpha elements found in the galactic halo and second-order effects. The reddening can be obtained in different ways, for example by comparing the spectral type with the continuum or with the color index, or by comparing reference sequences in the cmds and theoretical models.

In principle, if the reddening, or color excess $E(B - V)$, is known, through the ratio R_V between visual extinction and reddening, the extinction can be directly determined.

$$R_V = \frac{A_V}{E(B - V)}$$

However, accurate measurements require a more detailed analysis including second order effect analysis. The reddening in fact acts selectively, with greater absorption in the short-wavelength part of the spectrum with respect to longer wavelengths. In the case of color-magnitude diagrams, the finite width of the bands produces a distortion of the diagrams themselves (rotation) and contraction of the color index scale compared to unreddened diagrams. It is not easy to find an analytical solution of immediate application to the data because the value of the absorption A_V , the ratio R_V , and the excess $E(B - V)$ are linked together and we have different possible approaches to the problem depending on whether we start from excess of color (which usually constitutes the only observational data) or absorption.

Schmidt-Kaler has shown numerically that, having established the ratio R_V , the excess of color in $B - V$ depends on the intrinsic color of the source. There is an excess of decreasing color for cooler stars, as a consequence of the different dependence of the extinction to the wavelength within the B and V bands. The ratio R_V as a function of absorption and intrinsic color $(B - V)_0$ is:

$$R_V = 3.0 + 0.14(B - V)_0 + 0.025A_V$$

this implies a variation of about 5% in R_V between a star with $(B - V) = 0$ and another with $B - V = 1.0$ with the same A_V . The dependence on A_V becomes appreciable instead, and comparable to the effect due to the variation of the spectral type, only for absorption values over 5 magnitudes. Alternatively the author has obtained an empirical expression of R_V as a function of the excess of color (instead of absorption) which is of more immediate application:

$$R_V = 3.0 + 0.2(B - V)_0 + R_1 \cdot E(B - V)$$

with $R_1 = 0.026 \cdot 0.007(B - V)_0$.

Note that the author assumes $R_V = 3.0$ as average value for blue stars, while the current values are around 3.1. The second order term R_1 is not important for low reddening values and it is not very sensitive to the intrinsic color $(B - V)_0$, while R_V significantly changes, with a contribution of 0.2, between the color index $B - V = 0$ and 1.0. The increasing R_V with the intrinsic color index (cooler stars) is due to the more rapidly decrease of the color excess compared to the extinction ($R_V = A_V/E(B - V)$).

Schmidt-Kaler has tabulated, for integration on the spectrum, the relationship between the excess of color of a given spectral type and that of a star B0:

$$\eta = \frac{E(B - V)}{E(B - V)_{B0}}$$

where η is a complicate function of reddening and color excess of the star. Knowing η is very important: it allows to obtain the reddening for a specific spectral type, knowing the B0 one, used as reference. Indeed the B0 stars, very hot, are very common in literature and all reddening laws are set on this type of stars. Obviously, for B0 stars, $\eta = 1$ and, for example, for M5 type, $\eta = 0.89$.

1.14 Interstellar reddening maps

It is fundamental to know where dust and gas are located and in which quantity in order to estimate the absorption. In Fig. 1.5 is visible a galaxy map. There are two lines of sight, one directed to the galaxy center and the other one

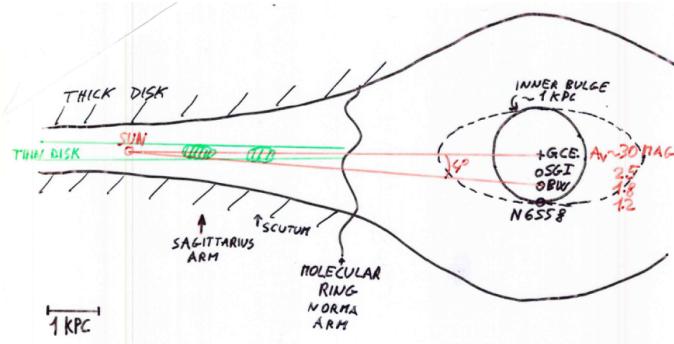


Figure 1.5: Reddening map of our galaxy

4° from the GC.

The absorption is very high to the GC where A_V is about 30 in magnitude, then there are some windows shoving that the reddening is dramatically decreasing at very low angles. Indeed inside the thin disk there are different gas and dust clouds.

If observations are directed perpendicularly to the galactic disk the extinction is minimal but sometimes not negligible because the reddening is distributed non-uniformly and there are small clouds also at relatively high galactic latitude.

Example: the GC - The reddening effect is very high looking to the GC. In infrared band the reddening effect is almost completely deleted.

Sun position - The Sun is in a very lucky position because it is situated in a sort of empty zone in the galactic disk called **chimney**.

Chapter 2

Measures of distance

The most used methods of measurement of distances are:

1. Radar
2. Annual trigonometric parallax
3. Group parallax
4. Spectrophotometric parallax
5. Nebular parallax
6. Differential galactic rotation (derived from Oort equations)
7. Signal dispersion (pulsar)
8. RR Lyrae type stars and the luminosity of the Horizontal Branch
9. Wilson-Bappu effect
10. Dynamic parallax
11. Size of the HII regions (equation of ionization-recombination)
12. Brightness/diameter for supernovae remnants
13. Cosmological distances (Hubble law)

These methods are divided into **direct** and **indirect**. The first one are based on simple geometric considerations and can be used over relatively short distances.

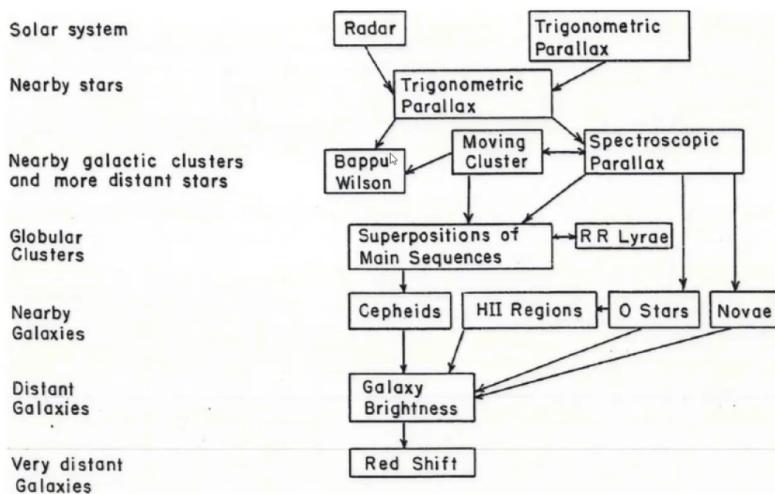


Figure 2.1: Summary of methods to measure the distances

2.1 Radar

Among the direct methods, radar is the conceptually simpler tool, but with the shortest range of action. It is based on the time between the transmission of the signal and the reception of the echo. Half of the time elapsed is simply the product of the propagation speed and the distance from the celestial body that produced the echo, then $d = ct$, where c is the propagation speed of the electromagnetic signal (in the vacuum).

The uncertainty of the measurement is due to the shape of the celestial body responsible for the echo and to the assumption on the propagation speed of the signal that is altered by the presence of matter of different nature (atmosphere, ionospheric and interplanetary plasma ...). The radar can only be used for very short distances because the intensity of the echo decreases with the fourth power of the distance.

Earth-Sun distance - Measurements of distance with respect to Venus, several small planets and, more recently with respect to Mars, have allowed us to derive the Earth-Sun distance with precision of the order of meters or less. The method is based on Kepler's third law and uses the equation expressed in the following form, using the Earth period and distance as a reference:

$$a + x = aP^{2/3}$$

where a represents the Earth-Sun distance, x the measured planet-Earth distance, P is the planet orbital period measured in Earth years. The term on the left of the equation represents the distance of the planet from the Sun. Then, once the term x is known, the distance a is immediately obtained.

The first measurements of the Earth-Sun distance date back to Aristarchus of Samos (310-230 BC) which derived the distance of the Sun from the Moon-Sun angle to the first quarter and concluded that it had to be about 20 times the Earth-in terms of Moon distance. This measure, as Keplero has shown, is far below the real distance. The cause of the error was due to the lack of knowledge of atmospheric refraction at the time. Later Cassini (1672) measured the astronomical unit from the distance of Mars obtained through the parallax. The first precise measure in modern times is that due to Jones (1932) who used the minimal approach of the Eros planet (22 million km) to derive the distance with the parallax (with an error of about 100,000 km).

A further significant improvement was possible with the use of distance measurements from Venus, with the radar (Pettengrill, 1966) which led to an accuracy of about 1000 km. Radar measurements on Mars, until recently, have not led to significant improvement due to the orography. The best current determination, derived from the telemetry distance of the probes sent to Mars, is reported by Pitjeva (2005):

$$\text{AU} = 1.495978706960 \cdot 10^{11} \pm 0.1 \text{ m}$$

Note that due to eccentricity the Earth-Sun distance varies from 1.02 to 0.98 AU.

Several authors have, however, found a progressive increase in the Earth-Sun distance of about 15 m/century, significantly greater than the measurement error, which does not find an obvious explanation.

2.2 Annual trigonometric parallax

It is a very simple method and the fundamental direct one. It is based on the well-known principle that consists in the observations of a celestial body at a distance of six months, relative to distant objects, using as base the axis of Earth revolution. Earth diameter is too small and we would have to detect very small angles in order to measure the distance of far stars.

As a first approximation we can assume that the accuracy of the measurement can be expressed as:

$$\sigma = 0.5 \frac{FWHM}{S/N}$$

where σ (in arcseconds) is the error on the position of an image in arcsecond, $FWHM$ (in arcseconds) is the Full Width at Half Maximum of the gaussian luminosity profile of the source, S/N is the signal to noise ratio of the same image, 0.5 is an empirical coefficient that could be different. This equation is empirical but also quite universal.

Taking a typical seeing of 1" and a signal-to-noise ratio around 100, we can easily see that distances up to the equivalent of about 0.01" of parallax can be measured with an accuracy of 30%, which means, by definition of parallax, up to a distance of 100 parsecs.

More accurate calculations show that accuracy depends on the seeing motion component and therefore also on the exposure time. The adaptive optics systems allow to obtain today much higher accuracy, since the FWHM of the corrected stellar images, in the infrared, approach those of diffraction width which, for the K band, in a telescope of

8 m in diameter corresponds to 0.08''. Thus disregarding image motion effects due to seeing, the 30% accuracy would be obtained up to distances of 1000 parsecs.

From Hipparcos space, parallaxes were obtained up to 1600 pc, while with the Space Telescope ACS, up to about the double.

2.3 Group parallax

The group parallax can be summarized simply as the distance measurement of a cluster of stars that moves with a radial velocity known from the spectra, compared to the apparent size variation.

We can write the equation $\theta = D/r$ where θ is the angle below which we see the cluster, r the distance and D its diameter, and the derivative of the angle with respect to time becomes: $d\theta/dt = -v\theta/r$ from which one immediately derives the distance r from the radial velocity v known and the measured variation of the angular dimension with the time. In practice the method is of limited application because it assumes the isotropy of the system and then because the measurement errors make it usable only for very close star clusters.

A more complex variant is based on the measurements of the proper motion vectors and the identification of the convergence point, but the uncertainties on the method remain the same.

2.4 Spectrophotometric parallax

This method requires the knowledge of the spectral type of the star under examination and is applicable only to stars of spectral types whose absolute magnitude (main sequence stars and giants) is known. The comparison between apparent magnitude and absolute magnitude allows to directly derive the distance modulus, and the distance itself once interstellar absorption is known. This method therefore belongs to indirect methods.

2.5 Wilson-Bappu effect

It is an empirical, little known, indirect method of which no detailed physical explanation is known. It is based on the equivalent width of the weak emission that appears at the center of the absorption of the lines H and K of the CaII. This quantity, called W by Wilson and Bappu (1957), correlates with the absolute magnitude of G-K spectral stars. Calibrated on the Sun and on known parallax stars, this method allows to obtain the distance of stars of which we have the high dispersion spectrum in the H and K lines of calcium with an uncertainty of about 10%.

The lack of knowledge of the mechanism that links the emission of the lines with the absolute magnitude constitutes a limit to the method. An attempt to explain it is that the absolute magnitude and therefore the brightness of the star would be connected with the size and motion of the convective cells and from these the existence of an emitting layer. At the moment it is not clear what the influence of the actual temperature or the spectral type might be.

The original relationship of Wilson and Bappu is:

$$M_V = -14.94 \log W + 27.59$$

The width W is measured from the ends of the line and the data are corrected by instrumental enlargement.

Practically, we measure the equivalent width of CaII line (from which we get W), then we enter it in the plot and we measure the corresponding absolute magnitude M_V . Then, using magnitude law with apparent magnitude, we get the distance module and from this the distance. Of course, in doing so it is important to know the interstellar reddening.

Using this method it is possible to determine the distance of individual solar type stars, which fall in the range $1.2 < \log W < 2.0$, a very common range including the vast majority of G stars.

This method is usually good but sometimes some points could have big dispersion and could move away from interpolation.

2.6 Nebular parallax

It is applied to expanding clouds (theoretically also to shrinking clouds) and is based on the simultaneous measurement of radial and transverse velocity by proper motions. If you measure the distances in parsecs, the proper motion in arcseconds per year and the radial velocities in km/s, you get:

$$D = \frac{v}{4.74 \mu}$$

where μ is the proper motion due to the expansion of the cloud.

2.7 Pulsating stars: RR Lyrae, Cepheids, W Virgins

The method is based on the absolute magnitude of the variables stars once their class, period and/or metallicity is known. Absolute magnitudes have been calibrated by nearby variables with distance measured by annual parallax (for example the RR Lyrae) or when they belong to star clusters of distances known from other methods.

The RR Lyrae are characterized by periods less than a day, and by color indexes between $B - V = 0.15$ and 0.45 . Instead, Cepheids have periods from a few days to tens of days. A classic example of Cepheid is the Polar Star. Current equations for RR Lyrae calibration are:

$$M_V = 0.16[\text{Fe}/\text{H}] + 0.98 \quad (\text{Jones, Baade-Wesswlink method})$$

$$M_V = 0.15[\text{Fe}/\text{H}] + 0.80 \quad (\text{Harris, 2010})$$

$$M_V = 0.21[\text{Fe}/\text{H}] + 0.75 \quad (\text{HST data, Benedict et al., 2011})$$

RR Lyrae are the most used photometric indicators, visible up to huge distances and easily identifiable thanks to their short and stable period of few hours.

They are evolved stars located on the Horizontal Branch so they burn He into C in the core and H into He in the envelope. Those stars have a color index un-reddened and very clear of $0.15 < B - V < 0.45$. As all variable stars, RR Lyrae in color-magnitude diagram are located in the instability strip, a vertical section that crosses the diagram and contains:

- in the upper section, variables with long periods and bigger radius so low density
- in the central part there are the RR Lyrae and more compact objects with shorter periods, about a few days or hours
- in the lower part are located all those compact objects such as white dwarfs or pulsar with extremely rapid period and rotation

But the uncertainty due to different biases (method of measuring the brightness of the RR Lyrae, range of metallicity, effects of age and evolution etc.) is wide. The metallicity coefficient ranges from 0.16, corresponding to most theoretical models, to a maximum of 0.30 (Sandage, from the period-shift method). The calibrations for population Cepheids I and W Virginis (or population Cepheids II) are more uncertain even if these distance indicators have the advantage of being more brilliant, and therefore observable at greater distances.

For the classical population I Cepheids, Feast et al. (1997) give the following relation period-luminosity, calibrated through parallaxes of Hipparcos:

$$M_V = -2.81 \log P - 1.43 (\pm 0.1)$$

Udalski (1999) proposes for the I band the following relationship adopted for extragalactic distances:

$$M_I = -2.962(\log P - 1) - 4.904$$

which is often more convenient because the amplitude of the Cepheid variability is greater at a longer wavelength. The strong dependence on the period is evident (the RR Lyrae brightness has no appreciable link with the period).

A Cepheid can reach a brightness of more than thousands of times higher than the Sun and therefore be visible up to a large distance (at $V = 25$ we get $m - M = 30$, corresponding to 107 parsecs, or 10 Mpc, almost the distance of the Virgin cluster, estimated today at 18 Mpc).

2.8 Dynamic parallaxes

This method is based on the knowledge of the semiaxis of the orbit of the double visual stars. The method is apparently limited because it can only be used with resolved and well known double stars, but in reality it is of great interest because it is geometric, and it can constitute an independent method for measuring some stellar populations.

The equation that is derived from the observations of double visuals, based on the third law of Kepler is:

$$a^3 = P^2 \cdot \text{const} \cdot (M_1 + M_2)$$

so it is sensitive to the masses. With this method the distance of the Pleiades from the double star Atlas was measured recently and it was found to be $d = 135\text{pc} \pm 2$, against 118 ± 4 which instead results from Hipparcos. The reason for this discrepancy is not yet clear.

2.9 Novae and supernovae

Novae reach their maximum luminosity which is:

$$M_V = -9.96 - 2.31 \log(\text{decl})$$

where decl is the decline time of 2 mag. Type 1a supernovae have $M_V = -19.3 \pm 0.1$.

They are candles observable from long distances, used also to establish the expansion of the universe. The accuracy of the determinations depends on the calibration and on the distinction between subtypes. There is also the dependence on interstellar absorption which is not always easily deducible from the same observations of novae and supernovae.

Distance of supernovae can be determined using the photometric relation showed before but also in a geometric way, through the determination of the angular expansion of the cloud in radio range.

Chapter 3

Young stellar population

3.1 Color-Magnitude diagram

The color-magnitude diagram is an useful item to study stellar populations. We have two axis:

- In the x-axis we can put the color index (observational) or the temperature (theoretical). These two quantities are related through the relation $B - V = -0.856 + 8540/T$. It can be used also the spectral type. The temperature is the effective temperature so the temperature of the photosphere and not the one of the core.
- In the y-axis we can put the absolute magnitude in the visual band (observational) or the luminosity (theoretical)

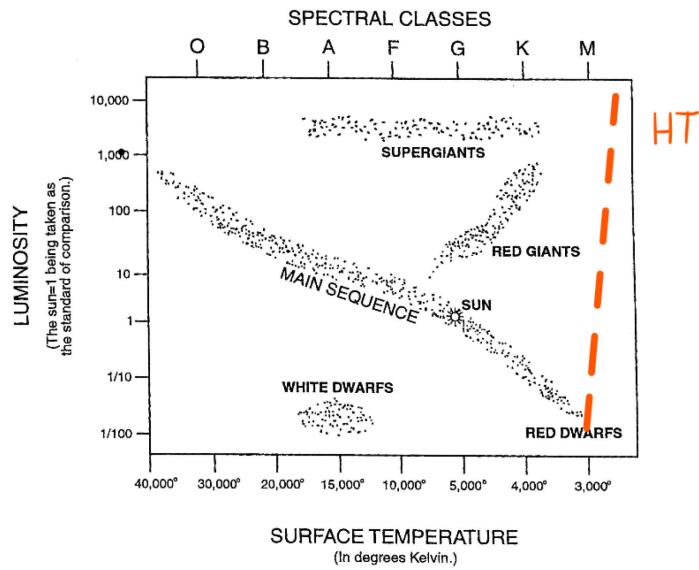


Figure 3.1: Color-magnitude diagram

Most of the stars, near the Sun, are located in the main sequence which is a relatively narrow region where stars burn H into He in the inner parts. Stars in this region can occupy the same point for very long time.

In the low part of the MS there are partially convective stars (lower mass, fainter and cooler) while at the top of the MS there are almost totally radiative stars (higher mass, brighter and hotter). Along MS stars are located according to the mass and they are characterized by a uniform chemical composition. The mass range of stars is quite small while the sensitivity of the position along the MS from the star mass is very high. These stars goes from $0.5 M_{\odot}$ to $5/6 M_{\odot}$. Going to lower masses, they are more numerous but also fainter.

Above this zone there is the zone of the giant and super giant stars, brighter than the sun but also redder. Between the giant zone and the MS there is the Hertzsprung gap, due to a rapid evolution of the stars that exit from the MS. We don't see a lot of stars due to a selection effect because we loose the fainter ones.

There is a small group on the lower left with respect to the MS that contains the white dwarfs, small and evolved stars, much dense and compact.

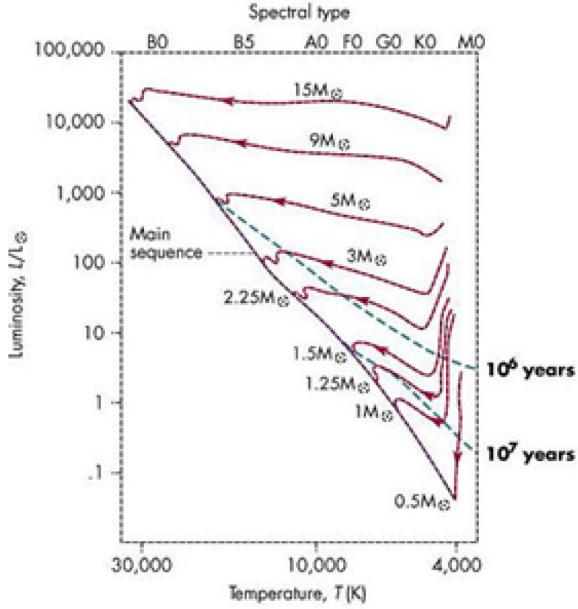


Figure 3.2: CMD from the theoretical point of view

In CMD, stars occupy only the allowed region, a sort of triangle between the MS and a vertical line called Hayashi line. The left side of the MS is not allowed by theoretical models because this section corresponds to a star that evolve completely mixed, a totally convective star.

There are different reasons for MS stars to go up to giant region:

- One is over-production of energy. When the star starts to burn H in shell, there is an He core which is unstable and contracts because there is no production of energy inside. This contraction produces thermal energy by gravitational contraction which creates a surplus of energy and this makes the star expanding
- the most important driver to the giant region is the average increase of molecular weight due to the transformation from H to He. Therefore the radius increases then the evolution of giant stars depends on mass.

The line that cannot be crossed is called **Hayashi line** and corresponds to the track of full convection. Indeed, moving to higher color index, which means lower surface temperature, the surface opacity increases in a strong function of temperature. So if temperature decreases, opacity increases and vice versa. When the layers are getting opaque then the energy transfer can not be anymore radiative and becomes convective. So, as T decreases, convection stars to be the main mechanism to transport energy in external layers and the star gets cooler. At the same time convection goes deeper in the star until it is fully convective along the Hayashi track. Stars cannot be more than fully convective so there are no stable configurations on the right side of Hayashi track.

3.1.1 CMD of young population

The color magnitude diagrams of young populations are normally based on the near-infrared (JHK) bands to minimize the problem of interstellar absorption. They are characterized by a main sequence where the most massive stars are found, while going down to smaller masses, from a certain point, the sequence appears truncated because the stars are still in the pre-sequence phase, on the right of the diagram. In some ways the diagram appears the opposite of that of the old populations where the main sequence is truncated in the upper part.

The young population corresponds to pre-MS stars, as shown in Fig. 3.2. The blue dotted line is the isochrone curve that represent a simple stellar population with the same age, so constant in time. Isochrones are not evolutionary paths and are useful as reference in the observational plot. The red solid lines represent the evolution track of stars with different masses.

There are some limits from the phenomenological point of view to the study of pre-MS evolution on HR diagrams. These are due to:

- the position on the HR diagram is influenced by the presence of excess infrared color due to the presence of disks etc.
- the regions of recent star formation are rich in dust which causes a strong reddening with consequent difficulties in deriving the intrinsic color temperatures.

Moreover, from the modeling point of view, in the main sequence we have the **Vogt-Russell's theorem** that says that the position of a star on the MS does not depend on its pre-MS history but only on its mass (with fixed age and chemical composition), like saying that for a given mass there is only one equilibrium position for stars that burn hydrogen in the core.

Color-magnitude diagrams of young populations are usually based on the near-IR bands in order to minimize the problem of interstellar absorption.

3.1.2 Evolution towards the MS

Stars form in a big cloud at very low density and temperature by gravitational contraction. In this phase temperature is low but luminosity is relatively big so the star starts from the upper part of the diagram. Then it reaches the upper limit of Hayashi line: now the protostar is formed and goes down through the Hayashi line. Going down the star contracts, reducing radius and reducing the emitting surface. This is a fast evolutionary phase and there is a lot of dust so stars in this phase are not observable in the visible band.

Evolution proceeds in fully convective phase contracting and heating up. In this phase the energy comes from gravitational contraction. Temperature increase, opacity decrease and therefore temperature gradient below the convective shell so the inner part becomes radiative. Evolution slows down and the star increases the surface temperature with a little increase in luminosity.

Stars leave the Hayashi line and enter the main sequence triggering the thermonuclear reactions. In this phase the gas surrounding the star is cool enough to condensate and becomes a flat rotating disk: this is the formation of a protoplanetary disk.

3.1.3 Time-scale of evolution

For a solar-type star the phase of the pre-sequence is around 10^7 years, and only the evolution along the Hayashi line takes 10^5 years.

3.2 The ages of young population

The age can be estimated from the position and shape of the pre-main sequence turnover. The diagrams often appear to be dispersed due to high field stars contamination, reddening and differential absorption and the presence of stars with infrared excess due to protoplanetary disks.

- Contamination: different techniques allow to clean up the diagrams from the contamination, for example with the proper motions or with the typical X emissions of young stars.
- The reddening and the differential absorption instead can be corrected by obtaining for each star the reddening in the two-color diagram $J - H$ $H - K$, with respect to the position of the unreddened points, and then correcting each point on the cmd diagram.
- The stars with infrared excess are identified in the two-color diagrams with the technique described below and treated separately.

3.2.1 Two colors diagram

For the study of young populations, in order to minimize the problem of interstellar absorption and to over-plot the isochrones to get the age, it is fundamental the preliminary study in the two-color infrared diagram (generally $J - H$ vs $H - K$).

In this diagram the place of the points occupied by the stellar photosphere models, Both of MS stars and giant stars, is well defined and limited between about $J - H$ and $H - K$ about 0.0 for hot stars (OB type), up to about $J - H = 0.8$ and $H - K = 0.4$ where the first extreme is reached by the giants while the second one is reached by the dwarfs. In particular, we can see the track from O stars to M stars, then the line splits: the giant one goes up while the dwarfs one goes down. The black solid line corresponds to the reddening limit for all the stars so all stars are shifted by

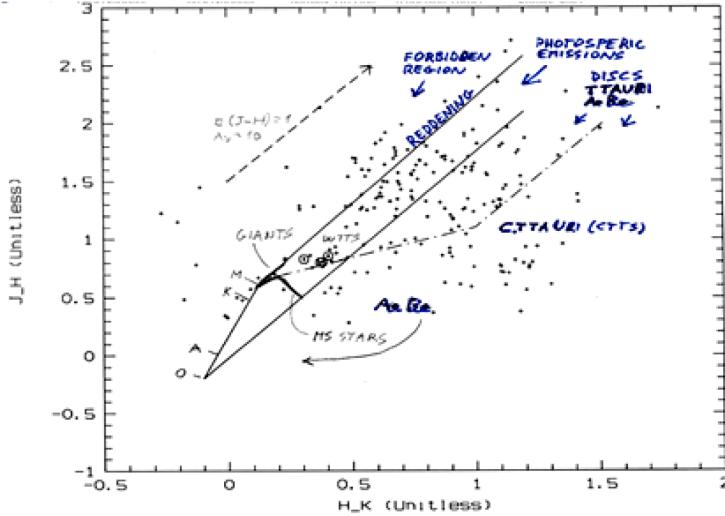


Figure 3.3: Standard JHK diagram

the reddening effect. As visible in the figure, the reddening vector corresponds to $E(J - H) = 1$ that is equal to an absorption A of about 10.

The reddening lines in this diagram can be easily calculated from the $A_\lambda/E(B - V)$ ratios that have been tabulated. From literature:

$$\frac{E(J - H)}{E(H - K)} = 1.94$$

which is the slope of the reddening lines in the discussed two-color diagram. Note that the absorption in the K band is almost a tenth (in magnitude!) of that in the visual band.

Above the upper black line there is the forbidden area: there are no stars above the reddening line because it is the limit to the color index of M stars but however we can see some points. They are basically measurements errors because we are talking about very faint stars.

Between the two black lines we have the permitted area where the majority of points is located. Under the bottom black line there is the disk region: here stars present a protoplanetary disk and it is populated by T.Tauri stars, Ae and Be stars. All these points are real stars located below the permitted area for the presence of the disk that causes an $H - K$ excess that moves the points to the right. There is also a reddening effect that shifts up all the points. The result is a large distribution of points below the permitted area.

From the two color diagram it is possible to individually correct the stars for reddening and in some cases also for the infrared excess due to the presence of the disk.

The photometry so obtained can be used in the infrared color magnitude to obtain ages and distances from the fit with the presequence isochrones. The accuracy of the age from the fit with the isochrones in the pre-sequence phase is limited by the accuracy of the isochrones themselves. Uncertainties derive from hydrodynamic factors that are neglected in the models, and the stability of the stellar stratification outside the convective regions. The ages derived from the models, on eclipse binary stars between 1 and 20 solar masses give deviations between 15 and 400%. Moreover, considerable discrepancies (200%) remain between the ages derived from the models of presequences and the turnoff of the main sequence. Despite these difficulties the theoretical models constitute a fundamental reference for the calibration of the other methods.

3.2.2 Other methods

For young stars there are four other different age estimation methods:

Lithium content

The decrease of photospheric lithium over time is considered one of the most reliable methods of dating the presequence stars. Observations are made on the Li line at 6707 Å.

The method is based on the rapid burning of lithium at temperatures lower than those of the PP hydrogen cycle (2.5 vs. 10 million degrees) which can then take place in pre-sequence stars. More precisely the ignition temperature, for solartype stars, occurs at the base of the Hayashi line, where the star is still entirely convective and therefore the mixing between the photospheric material of the star can occur, at the age of just over a million years. The reaction takes place directly by capturing a proton or an alpha particle and can lead to total exhaustion of lithium in a relatively short time (about a hundred million years). The exhaustion of lithium obviously depends on the temperature of the star. For very low temperatures (so small masses) the reaction is obviously slower. But, at high masses and temperatures, the decrease in lithium can be limited by the rapid development of the radiative core that pushes the base of external convection towards greater distances, bringing it to a region where the temperature is too low for the reaction of lithium. This explains why the models expect greater abundance of lithium at the extremes of very low and very high temperatures. The calibration was done in the age range between 1 and 30 million years, using atmosphere and isochrone models, but it is probably extendable even to a hundred million years.

The method is therefore dependent on models, in particular on temperature. The extrapolation of the lithium burning models shows that the decrease in lithium is more accentuated for cold M-type stars, for which the ages thus derived are considerably higher than those obtained from the isochrones.

This method is often used for the study of age of individual stars.

Disk fraction against age

the identification of stars with disks is obtained by studying the two-color infrared diagram $J - H$ and $H - K$, as explained before. The method, used for sets of contemporary stars, is calibrated in the range of 1 to 10 million years. It is found that stars with ages less than about one million years are almost always surrounded by a disc, detectable in JHK (excess in K) with more than 90% probability. This probability drops to 50% at 2-3 million years and 5% over 5 million years. This small fraction of stars with discs remains up to over 10 million years.

The excess in the K band corresponds to the inner part of the disc where the temperature, as already mentioned, is around 2000-3000 K. The excess of $K - L$ color instead reveals the relatively cooler part of the disc. The trend is similar to the previous one, with a scale time of about 5 million years. It can be concluded that within an uncertainty of a couple of million years the disk disappears almost simultaneously from the hottest part, where the photoevaporation is more efficient, to the colder one.

The reasons why there is this gradual reduction of stars with a disk, going from the youngest to the oldest are the following:

- planets, which form in 10 million years, have collected all the material from the disk
- disks are removed partially by stellar winds

Asteroseismology

Stellar oscillations are an effective means of deriving the ages of isolated stars. The fundamental point is that the oscillations give direct information on the stellar structure. In particular, what is called the small frequency separation represents the observable speed of sound where thermonuclear reactions are generated. The limit of the method is given both by the availability, for now limited, of data, and by the models involved for the interpretation of the observations.

Gyrochronology

The rotation speed of young stars decreases with time. It is known that for T Tauri stars the rotation period is around ten days, while it is known that the Sun rotates with a period of about 25 days. There is therefore a braking mechanism not yet well identified. The method is applicable to cold stars, of the F-M type, with convective envelope and age of over 100 million years. In these limits the period of rotation can be expressed as: $P = f(B - V) \cdot g(t)$, where f and g are two functions dependent on the color index and age. The calibration is adapted in order to reproduce the solar parameters.

Unfortunately this observable relation between rotational speed and age has a wide dispersion due to the different nature of stars. There are the rapid rotators in hydrodynamic collapse, and the slow rotators.

So the main methods for measuring the age of young stars are:

- methods that can be used only for single stars, measured one by one:
 - Li abundance
 - rotational speed of the star
- methods for groups of stars
 - isochrones
 - fraction of stars with a protoplanetary disk

Chapter 4

Physics of planets

4.1 Basic equations

Gasses in Earth atmosphere can be treated as perfect gasses so they are described by the equation of perfect gas:

$$PV = nRT$$

where n is the mole number and R is the constant of gas. An equivalent expression is:

$$PV = NKT$$

where N is the molecular number and K is the Boltzmann constant.

Another fundamental equation is the following that allows us to derive the kinematic energy of gas particles:

$$\frac{1}{2}mv^2 = \frac{3}{2}KT$$

where m is the mass particle (depending on gas type), v is the velocity of the particles and T is the temperature as seen before.

This equation means that kinematic energy is equal to the energy of motion of particles due to the temperature. So we can derive:

$$v = \sqrt{\frac{3KT}{2m}} \propto \sqrt{\frac{T}{m}}$$

Therefore we can derive the velocity due to the thermal motion.

Relating kinetic energy with gravitational energy we can compute the escape velocity:

$$\frac{1}{2}mv^2 = \frac{GmM}{r^2}$$

so we obtain the escape velocity which is equal to:

$$v_{esc} = \sqrt{\frac{2GM}{r^2}}$$

That means that if we want to keep the atmosphere bounded on a planet the escape velocity must be larger than the thermal velocity. The process of atmospheric loss is very important, because the presence of the atmosphere is fundamental for life.

4.2 The processes of atmospheric loss

Thermal processes

The loss of atmosphere due to thermal motions (Jeans mechanism) becomes significant when the thermal velocity of the molecules is comparable with the escape velocity. It is estimated that an atmosphere such as the terrestrial one is totally lost in a time scale of the age of the solar system when the thermal speed is >20% of the escape speed, even if it is lower than the escape velocity. The reason is that there is a fraction of particles, in the high velocity tail of the Maxwell distribution of the speeds, with a velocity higher than the escape speed. Once these particles are lost

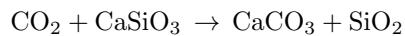
the Maxwell distribution is normalized again and further particles with a velocity higher than the escape speed are generated. This mechanism proceeds, eventually, until most of the atmosphere is lost. This is the reason why hydrogen is not present in the atmosphere.

This is not the dominant process that removes planetary atmosphere, indeed there are other removal effects, explained in the following subsections.

Non-thermal processes: chemical-physical capture ("sequestration")

While the loss for solar wind does not seem to be significant at least on Earth (but it is also thought on Venus), instead the capture of gas by the crust is important. It is believed that on Earth the carbon fixed on the rocks constitutes about 250 times that of the atmosphere. Given that the current carbon dioxide content in the atmosphere is of the order of 380 parts per million, the release of all carbon in the form of carbon dioxide could bring the atmospheric pressure to values around a hundred atmospheres, similar to that of Venus. This gives an idea of the importance of the mechanism of chemical capture of atmospheric gases by rocks.

The capture mechanism is based on the Urey reaction, which can be simplified in the following form:



where CaSiO_3 is calcium silicate, CaCO_3 is calcium carbonate and SiO_2 is quartz. This reaction is very slow and the combination between rocks and gases is not possible if there is no water since it acts like a catalyst. This happens thanks to rainfalls.

The carbonates thus formed, taken deep inside the Earth, by the motions of the crust. Under conditions of high pressures and temperatures, the carbonates can again release carbon dioxide in the form of volcanic eruptions. This mechanism is critical for maintaining the thermal balance on the earth's surface. The time scale of the reaction is very uncertain. Recent estimates give between 100,000 years and some millions of years.

Solar winds removal

Solar winds are composed by electrons, protons and other energetic particles that travel at a speed of about 400 km/s with very high energy. When these particles impact with the atmosphere the impact can transfer kinetic energy to atmospheric particles sufficient to allow them to escape. So solar winds can rip the outermost layer of the atmosphere off.

On Earth this is not efficient due to the magnetic field that surrounds the Earth and deviates the solar wind thanks to the Lorentz force.

Therefore magnetic fields, are very extended but don't protect Earth atmosphere completely. However the loss of atmosphere for solar wind does not seem to be significant at least on Earth.

UV effect

Also the UV component of the radiation emitted by the Sun can cause an atmospheric loss but it is necessary to study more this effect.

Bombardment erosion

This effect is caused by the impact of the Earth with asteroids etc. The impact causes a removal of some fraction of gases due to mechanical and thermal reasons. This mechanism was effective at the early beginning of our planet, not now.

Calculation of surface temperature

The most important parameter in studying a planet is the surface temperature. The effective temperature depends on the energy received from the Sun and it is the balance temperature obtained by the received energy from the Sun and the emitted energy from the planet (that emits like a black body). This because the crust is an optimal insulation, so the thermal energy is enclosed inside the Earth.

4.3 Atmospheric pressure distribution in hydrostatic equilibrium

In conditions of hydrostatic equilibrium, considering a perfect gas, the equation valid are the perfect gas equation and the one for the hydrostatic equilibrium:

$$\frac{dP}{dz} = -g\rho$$

From these two we can obtain:

$$P_z = P_0 \exp(-z/H)$$

where $H = KT/mg$ is the scale height that depends on the temperature, on the composition of the atmosphere and on the gravity of the planet. It increases with gravity as it increases with temperature.

4.4 Tidal forces

Consider now tidal forces. In a simple configuration they are defined as:

$$F_{tidal} = G \left(\frac{Mm_1}{(D-r)^2} - \frac{Mm_2}{(D+r)^2} \right) = GMm \left(\frac{1}{(D-r)^2} - \frac{1}{(D+r)^2} \right)$$

that, doing an approximation, becomes:

$$F_{tidal} = GMm \frac{4r}{D^3}$$

So a body with mass $2m$ (composed by two identical masses m) is in equilibrium conditions if the internal force of gravity exceeds one of tide that is:

$$F_g = \frac{Gm^2}{(2r)^2} > F_{tidal}$$

then

$$\rho > 4M/D^2$$

Often the value 10 is used as the coefficient instead of 4 to take into account other effects including, for example, rotation. If this condition is respected, a static tidal force has only the effect of an elastic deformation.

4.5 Electromagnetic emission, effective temperature, greenhouse effect

The electromagnetic emission of the planets includes the diffused radiation of solar origin, which dominates in the visible, and the emission of thermal radiation. It is easy to see that the thermal emission has a peak in the far infrared, between about 10 microns of Mercury and the 100 microns of Pluto, while in optical the radiation is largely dominated by the diffused solar component with a peak at about 0.5 microns.

The bodies of the solar system are therefore visible only because solar radiation is largely concentrated between 0.4 and 0.8 microns. The solar component, however, is reduced to a negligible intensity in the infrared-radio.

Part of this radiation is absorbed, part is diffused back to the space. The relative proportion depends on the absorption coefficient, or rather on the quantity defined as the **Bond albedo**, which is the ratio between the radiation diffused to the space by a sphere in all directions and the radiation received from a celestial body. The maximum value of the albedo is 1 for a perfectly reflective surface, while it tends to zero for a black body. The average albedo of the Earth is between 0.32 and 0.36.

Said A the albedo of Bond, the quantity of relative radiation absorbed is therefore $1 - A$.

The effective temperature of a body of the solar system can be defined as the equilibrium temperature of a sphere that emits as a black body, located at a distance D from the Sun, of diameter r and albedo A . We can write a relation between the Sun energy absorbed by the planet and the irradiated one, in the hypothesis that the thermal emission is that of a black body:

$$S(1 - A)\pi r^2 = 4\pi r^2 \sigma T^4$$

where S is the solar flux a the distance of the planet and σ si the Stephan-Boltzmann constant.

The equation is valid if the object is in rapid rotation, otherwise the term in the second member must be divided by two, to take into account that the radiation is dominated by a single hemisphere. The T variable, so defined, is the effective temperature. We see that T depends on the albedo and the heliocentric distance, while it is independent of the size of the object. This is true, however, up to the size of particles of the same order as the wavelength, where the law of Mie must be taken into account. In the opposite case of very massive bodies we will have to instead take into

account the sources of internal energy that can, as in the case of Jupiter and Saturn, be comparable to the amount of energy received by the Sun.

The actual temperature of the Earth thus calculated, assumed an albedo of 0.33, gives 263 K, about 30 degrees lower than the average surface Earth temperature.

The presence of an atmosphere around a planetary body modifies its temperature in the sense that it acts as a filter for the infrared radiation sent back towards the space. In the latter case, part of the radiation diffused by the soil is reabsorbed by the atmosphere which in turn will radiate both towards the space and towards the ground. It is the case of **greenhouse effect**, which on Earth is due to the absorbing action of water vapor and carbon dioxide of the infrared radiation emitted from the ground.

Greenhouse effect

The basic concept of the greenhouse effect is that we are treating the atmosphere like a region of the space above the ground that can be compared to a thin screen which absorbs part of the IR radiation and re-emit it into two directions: to space and to the ground. In particular the IR radiation hits the molecules that can absorb this wavelength and re-emit the radiation. This is a normal process of ionization and recombination.

There are special components in the atmosphere that absorb IR radiation. The dominant gases, N and O, are neutral, completely transparent to IR radiation. More complex molecules like carbon dioxide or water vapor can create much more absorption lines and bands.

Once the IR radiation is absorbed, the absorbed energy is used by ions to do an electronic jump to higher levels. Then they go in a excited status that could be rotational, electronic, etc. This status is unstable so ions re-emit the radiation. In this case the greenhouse effect could not occur. In order to heat the atmosphere, it is necessary to transfer a certain amount of energy to the dominant gases, N and O, by collision. This process can heat up the temperature of a planet only if the time of decay is bigger than the time of collision, because the energy absorbed must be transferred to other molecules.

To quantify the greenhouse effect we can use a simplified model where the atmosphere is seen as a partially opaque body, with an emissivity coefficient that is currently estimated around 0.7. The emissivity can vary from a minimum of 0 (for a perfectly transparent or reflective body) to a maximum of 1 (black body).

This additional radiant element has a temperature similar to that of the soil and radiates infrared radiation both towards the outer space and towards the ground in equal proportion. It is therefore a question of adding an additional component to the first member of equation for the effective temperature which takes into account this additional energy contribution in thermal equilibrium:

$$\frac{1}{2}\varepsilon 4\pi r^2\sigma T^4 + S(1 - A)\pi r^2 = 4\pi r^2\sigma T^4$$

We then easily obtain:

$$T^4(4\sigma - 2\sigma\varepsilon) = S(1 - A)$$

These equations are based on the energy equilibrium.

An alternative way to express the correct temperature for the greenhouse effect is to evaluate the ratio between actual temperature (T_e) and temperature with greenhouse effect (T_s):

$$\frac{T_s^4}{T_e^4} = \frac{1}{(1 - \varepsilon/2)} \quad \text{or} \quad T_s = 1.106T_e$$

This result indicates a contribution of the greenhouse effect of about 30 degrees, so $T_s=293$ K. From the equations we see that the greenhouse effect is determined by the emissivity coefficient (if we put it at zero, $T_s = T_e$). The theoretical estimate of ε as a function of the content of gases in the atmosphere is a crucial problem of climatology.

In general it can be assumed that the emissivity coefficient is given by the sum of the contributions of the gases that have absorption lines in the mid-infrared, in particular water vapor (at 70%) and carbon dioxide, in addition to other gases present in traces like ozone and methane. The water vapor content grows rapidly with the temperature and therefore also increases the produced greenhouse effect, while the carbon dioxide reacts slowly to the temperature, not necessarily in a positive way, with a more complex mechanism, basically following Urey's reaction. From the equation of the effective temperature it is evident that a significant increase of the emission coefficient of the atmosphere, up to a maximum of 1.0, would increase the temperature a few degrees (263x1.19=312 K). However the above calculation do not takes into account the temperature difference between the surface and the atmospheric layer. A more accurate calculation takes into account the equilibrium temperature of the atmosphere and a multilayer model, from the flux

balance.

Asssuming that the surface temperature is different than the one of the atmosphere, input and output radiation including an atmospheric "thin screen", with emissivity ε , can be written as:

$$S(1 - A)\pi r^2 = 4\pi r^2 \sigma T_s^4 (1 - \varepsilon) + 4\pi r^2 \varepsilon \sigma T_a^4$$

where $S(1 - A)$ is the fraction of absorbed energy per surface unit, T_s is surface temperature and T_a is the atmospheric temperature; σT_s^4 is the black body emission.

The thermal equilibrium of the atmospheric "thin screen" is:

$$\varepsilon \sigma T_s^4 = 2\varepsilon \sigma T_a^4$$

The latter gives:

$$T_a = T_s 2^{-1/4}$$

so $T_a = 241$ K: this is the temperature of a high altitude layer. So, knowing ε and T_a we can get T_s .

Replacing T_a we get:

$$T_s^4 = S(1 - A)/4\sigma(1 - \varepsilon/2)$$

For $\varepsilon=0.77$ and $A=0.33$ we get $T_s=288$ K, just a bit higher than the real temperature of about 287.9 K. This calculation is not perfect because "thin screen" is just an approximation, while atmosphere is actually an extended layer composed by many different layers at different temperatures. A multilayer calculation can give a more accurate result, for example putting an infinite number of layers and integrating the equation. This approach is mathematically correct but it's not so easy because of the structure of the layers.

Effective temperature

The effective temperature of a body is the temperature of a black body that would emit the same total amount of electromagnetic radiation. Effective temperature is often used as an estimate of a body's surface temperature when the body's emissivity curve is not known.

The area of the planet that absorbs the power from the star is A_{abs} and is a small fraction of the total area of the planet $A_{tot} = 4\pi r^2$. This area intercepts some of the power spread on a sphere of radius D .

The planet reflects some of the incoming radiation so we need to incorporate the albedo a . The expression of the absorbed power is then:

$$P_{abs} = \frac{LA_{abs}(1 - a)}{4\pi D^2}$$

Although the entire planet is not at the same temperature, it will radiate as if it had a temperature T over an area A_{rad} which is again a small fraction of the entire area of the planet. The emissivity factor ε represents atmospheric effects. The Stephan-Boltzmann law gives an expression for the power radiated by the planet:

$$P_{rad} = A_{rad}\varepsilon\sigma T^4$$

Equating these two expressions gives an expression for the surface temperature:

$$T = \sqrt[4]{\frac{A_{abs}L(1 - a)}{A_{rad}4\pi\varepsilon\sigma D^2}}$$

Common assumption for the ratio of the areas are 1/4 for rapidly rotating bodies and 1/2 for the slow ones. This equations does not take into account any effects from the internal heating of the planet.

Anti-greenhouse effect

The above calculations don't take into account neither the effects of aerosols neither the uneven distribution of the temperature across the Earth surface. Recent calculations give a lower "effective" temperature, then the need of a higher greenhouse contribution. The aerosol contribution has been discussed by Turco et al., (1991). They calculated the temperature, including aerosols (they call them "smoke"), after the pollution due to a nuclear explosion, which has an effect very similar to an asteroid collision. The complete equation is:

$$\frac{T_g}{T_0} = \left\{ \frac{f}{2 - \varepsilon_s} \left[1 + e^{\tau_s/\mu_0} \left(\frac{2 + \varepsilon_a(1 - \varepsilon_s)}{2 - \varepsilon_a} \right) \right] \right\}$$

Here f is a coefficient related to the two albedos, s means smoke, and a means clean atmosphere, T_0 is the effective temperature.

A very heavy aerosol can counterbalance the greenhouse effect

4.6 Generalities of the Solar System

The Solar System is constituted by the Sun and the bodies rotating around it. There are 8 planets, Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus and Neptune (Pluto is a minor planet) and a set of minor bodies such as asteroids, comets, interplanetary dust and a solar wind composed of nuclear particles, mainly helium and hydrogen nuclei. The mass spectrum of these minor bodies covers an interval of as many as 40 orders of magnitude in mass.

The Sun, together with the solar system, is on the periphery of the celestial system to which it belongs, the Galaxy or Milky Way, at a distance of about 8 kpc (more precisely the recent data give 8.2 kpc) from the galactic center, to only a dozen parsecs from the symmetry plane, inside the fragment of a spiral arm known as the Orion arm.

The orbits of the planets all lie on the same plane with small deviations, in particular of the planet closest to the Sun, Mercury. The ecliptic is defined as the maximum circle that describes the Sun on the sky in the course of the year and it corresponds physically to the plane of the Earth's orbit around the Sun. The Sun rotates on itself with an average period around twenty days, corresponding to an equatorial tangential speed of about 2 km / s. The orbital characteristics of the planets are largely dominated by the gravity of the Sun.

The planets rotate around the Sun in the same direction and the orbits are almost strictly circular, with the exception of Mercury, the planet closest to the Sun, and Pluto, which is the farthest. Mars also has an orbit with an appreciable ellipticity ($e=0.09$ vs. 0.02 of the Earth) a property that, together with its proximity to the Earth has allowed Kepler, from the study of the motion of Mars, to derive the three laws that regulate the motions of the planets around the Sun. Kepler's laws are empirical, approximate laws, which are often used in the study of the solar system due to their simplicity and immediate application which express some important relationships between the orbital elements.

The planets in turn rotate around their own axis (axis of rotation) which, for most of the planets, is almost perpendicular to the plane of the orbit. Important exceptions, only partially explained, are that of Venus, which has a retrograde motion, and of Uranus, which has the axis of rotation almost parallel to that of the orbit. The planetary orbits are characterized by parameters called the orbital elements, the most important of which are the **eccentricity**, the **period**, the **inclination** (with respect to the plane of the earth's orbit around the Sun, called the plane of the ecliptic when it is seen projected in the sky) and the position of the **line of the nodes** (the line of the nodes is the intersection of the planet's orbit with the reference plane of the Earth's orbit).

4.6.1 Regularities and properties of the Solar System

Distribution of the distances

The solar system has important regularities, only partly explained in terms of general formation processes of planetary systems.

An important regularity in the solar system is the clear separation of the planets into two main families with distances from the Sun, and distinct physical properties. The 4 planets closest to the Sun (Mercury, Venus, Earth and Mars) are all within about 250 million kilometers from the Sun, while the next, Jupiter, is at a distance of almost one billion kilometers. The outer planets (Jupiter, Saturn, Uranus, Neptune) are also called Jovian planets and are distinguished not only for the distance, but also for the lower temperature, for the large size and for the density. Finally, the farthest planet or minor planet, Pluto, is not easily classifiable, being very distant, of small size and mass. The surface temperature values of the planets that can be measured from the Earth in the range from about 300 C of Mercury, in the part facing the Sun, to around -250 C of Pluto.

While the interpretation of temperatures is fairly obvious, there is still no consistent explanation regarding the distribution of the masses. It should be noted in this regard that the planetary systems around other stars so far discovered show several cases of very massive planets even at very small distances with respect to the star around which they rotate.

Mass and density distribution

The distribution of the masses of the planets follows a characteristic pattern, with small masses, comparable to the terrestrial one, for the 4 planets closest to the Sun, while the most distant planets all have masses of several orders of magnitude higher, with the sole exception of Pluto, which instead has a mass lower than that of the Moon (0.178 Moon masses).

For these common properties (proximity to the Sun, mass and size), the 4 planets closest to the Sun are called terrestrial, the other Jovians.

Another important difference between the two families is the density. Very high for terrestrial ones (between 3 and 5 times the density of water), to that of the Jovian planets which is comparable to the water.

Satellites

Most planets in the solar system have natural satellites. In the case of the terrestrial planets the Earth possesses only one (the Moon) and Mars two (Phobos and Deimos), while Mercury and Venus, the closest to the Sun, do not have any. It is not clear if this is linked to the process of planet formation or if the gravitational field of the Sun has prevented their formation.

The giant planets, on the other hand, have many satellites and of considerable dimensions. Jupiter has 4 main satellites (Galilean satellites), of dimensions comparable to Mercury, and a few dozen smaller satellites. There are also numerous Saturn satellites. The largest satellite, Titan, is the only satellite in the solar system to have an extended atmosphere. There are also numerous satellites around Uranus and Neptune. Pluto also has 5 satellites. One of these, Charon, has a size comparable to Pluto itself. In this context the Earth is still a rare case (similar to Pluto) because the Moon has a size of almost a third compared to the Earth.

Rotations

The planets of the solar system rotate around their own axis, in the same direction, with two important exceptions: Venus rotating in a retrograde direction, and Uranus which has the rotation axis almost parallel to the orbital plane. While for the case of Uranus the magnetic axis forms an angle of almost 90 degrees with respect to the ecliptic, and therefore we can explain the anomaly with a violent impact, the problem of Venus remains to be explained. According to Colombo the retrograde rotation of Venus is due to the secular effect of the attraction of the Earth with the result of an Earth-Venus resonance.

The rotation periods are of the order of days, or several months, for the terrestrial planets, while they go down to a fraction of day for the Jovian planets.

4.6.2 Minor bodies

Asteroids and meteorites

There are many minor bodies with dimensions smaller than one hundred kilometers. Many of these are found between the orbit of Mars and Jupiter (**asteroid belt**), while others are beyond the orbit of Neptune (**transneptunians**). Depending on their orbital characteristics these objects are divided into families. The most important families are the Apollo-Amor (Near Earth Objects).

Meteorites

The smaller bodies are very numerous and often enter the earth's atmosphere. Depending on their mass and composition they are destroyed at high altitude (50 km) or can reach the earth's surface.

The phenomenon known as shooting stars refers to the high-altitude destruction of small-sized meteorites (around the millimeter), usually of cometary origin, as shown by their periodicity due to the passage of the Earth in a cometary orbit.

Meteorites that reach the ground are important residual samples from the formation of the solar system. Unlike the planets, in fact, they did not undergo those processes of fusion and chemical-physical modifications that occurred instead on the planets.

The meteorites collected are distinguished from the terrestrial samples by the evidence of signs of superficial fusion produced during the passage through the atmosphere. Meteorites cross the atmosphere at speeds around 50-70 km/s. It should be remembered that the Earth rotates around the Sun at a speed of 40 km/s, while the acceleration that an external body undergoes to the solar system that reaches the Earth, can give it a speed of 30 km/s. Taking into account also the acceleration suffered by the earth's gravity, we see that the maximum combined speed of a meteorite is just over 70 km/s, while the minimum speed is around 30-40 km/s, in excellent agreement with the observations. An exception are some meteorites that reach 100-120 km/s, according to measurements made with the triangulation by radar, of doubtful interpretation.

Meteorites are divided into two main categories: ferrous and non-ferrous.

- The ferrous ones are made of a high purity iron-nickel alloy, compact, and of high specific weight, and constitute most of the collected meteorites. They are also the most easily identifiable ones.

- Non-ferrous meteorites can be basaltic or carbonaceous and can appear compact or as chondrites (the name “condrule” indicates inclusions of small spheres).

There are also some other types of meteorites:

- Basaltic meteorites are rare and difficult to distinguish from terrestrial stones. A sub-category of basaltic meteorites, the shergottites (from the name of an Indian village where the prototype was found), appear compact and homogeneous. The shergottites could be samples of Martian rocks ejected by Mars due to impacts or volcanic ejections.
- The carbonaceous chondrites, on the other hand, are more easily distinguishable and often appear as friable agglomerates. By their nature they are the meteorites that can hardly survive the impact with the earth's atmosphere. It is therefore believed, despite their scarcity compared to ferrous ones, that they constitute the most numerous samples in space.

Carbonaceous chondrites do not have an obvious terrestrial counterpart. The dating performed by the radioactive isotope ratios with respect to their derivatives (eg uranium 238 with respect to lead 206) allows the samples to be dated. It is shown that carbonaceous chondrites are the oldest relicts of the solar system and their upper ages converge towards 4.6 billion years, much older than the oldest terrestrial rocks (except for some zircons recently discovered in Australia). This dating, together with some lunar samples, constitutes the basis for the estimation of the age of the solar system. It is also used by astronomers for models of stellar evolution that are calibrated on the age of the Sun (estimated equal to that of the solar system).

Let us remember for completeness the existence of another category of meteorites of more uncertain origin, the glassy tektites, often concentrated in defined areas, which could be lunar samples ejected by meteoric impacts and then fallen on the Earth.

Comets

There is no clear distinction between comets and asteroids, although the former are constituted by volatile materials which, when the comet approaches the Sun (within the orbit of Jupiter), evaporate and give rise to the coma and the characteristic cometary tail.

The theory on the formation of comets derives from the original idea of Wipple, subsequently refined according to the dirty snowball model.

The nucleus of a comet would consist of a set of ice and dust. With the evaporation of the ice (mainly ice water) dusts of various sizes are also dragged into the dust tail and leave a long trail in the comet's trajectory (hence the meteoric rains when the Earth passes through these trails). It should however be noted that in ground spectra no water molecules are observed, which should instead constitute the dominant fraction of the molecules, in the inner part of the coma. Most of the molecules visible in the spectra instead, are those of the CN, CH, Na, C₂ etc. which result from the decomposition of the primary molecules by solar radiation, in the immediate vicinity of the nucleus, therefore in a region that is not resolved from the ground.

According to studies from Oort, comets are present in large numbers in the form of inert nuclei, at a great distance from the Sun, well beyond the orbit of Pluto (at about 50.000 - 100.000 astronomical units), with orbits characterized by very long periods. As shown by the images of the Halley nucleus, the appearance of the cometary nuclei is of very dark objects, with a diameter of a few kilometers. Perturbations due to the passage of stars can bring comets into a very eccentric orbit and therefore penetrate the inner part of the solar system, until they reach the Sun.

Also the comets constitute a sample that has undergone few chemical-physical changes since their formation. Comets can be the main source of interplanetary dust that reaches Earth. It is interesting to note that, according to some theories, on the comets there would be suitable conditions (solar radiation, graphite particles, ice), for the synthesis of complex prebiotic organic substances. The evaporation of comet ice occurs at the distance of several astronomical units from the Sun. Indeed, at 7-8 AU a significant evaporation of the ice of carbon dioxide and carbon monoxide occurs, while water vapor evaporates significantly only between 3 and 4 AU, at a temperature of 120-140 K.

It is clear that at each step the cometary nucleus loses mainly the ice of carbon dioxide and carbon monoxide, while the water ice persists longer, with an estimated loss of around ten meters for each passage to the perihelion (within an astronomical unit). It is difficult to estimate the mass of cometary nuclei: from gas and dust losses they are estimated between 10¹⁴ and 10¹⁷ g, while the most precise estimate, by the Giotto probe on Halley's comet, gave 10¹⁷ g, corresponding to about 10 orders of magnitude less than the mass of the Earth. It is therefore easy to conclude that the total mass of comets in the solar system does not exceed the mass of the Earth.

4.7 The formation of the Solar System

Protoplanetary disk formation

It is now believed that the stars are formed by the gravitational contraction of large clouds of gas of a sufficiently low temperature to allow the prevalence of the gravitational force over that of the internal gas pressure (Jeans theory). The process, once started (ex. from the passage of the density waves of the spiral arms of the Galaxy or from shock waves due to supernovae explosions), proceeds with increasing speed, supersonic, with the gas in free fall towards the center.

The equilibrium is reached when the density of the gas and its temperature are sufficient to oppose the gravitational force. A steady-state equilibrium situation is not completely reached, however, because the gas continues to lose energy by radiation and therefore the contraction continues, but in much larger scale times, through states that can be considered as quasi-stationary.

A spherical central body of high temperature is thus formed that will give rise to the Sun, while in the external part the contraction leads to a high rotation speed (due to the conservation of the momentum) and therefore to a flat disk. From the current dispersion of the planets orbits with respect to the ecliptic plane, it is estimated that the thickness of the protoplanetary disk is of the order of one tenth of an astronomical unit. There is no common agreement about the temperature of this disk, while it is estimated from the extent of the tidal forces exerted by the Sun that its minimum mass is around 0.1-0.01 solar masses. The same result comes from the chemical analysis of the planets and from the estimate of the lost mass of volatile elements during the planet formation.

The temperature of the disk in its innermost part was higher than in the outer part. On the basis of the thermal equilibrium, between energy received from the Sun and lost by radiation, some authors have estimated a temperature between about 300 K up to 600 K for the gas that should have been 4.6 billion years ago at the distance of the Earth from the Sun. The uncertainty is due not only to the difficulty in estimating the opacity of the interplanetary medium, but also to the brightness of the Sun during the first evolutionary phases. The temperature on the outer regions of the disk, near the orbit of Pluto, would have instead reached a few degrees above the absolute zero.

Formation of planetesimals

The process that leads from the protoplanetary disk to the formation of the planets consists of three main phases:

1. **condensation** of the gas into solid particles (grains)
2. **aggregation** of grains to solid particles up to the formation of planetesimals of size around one kilometer
3. **agglomeration** of planetesimals in planets

In the end we have a process of intense solar wind that removes from the solar system gas and residual particles.

One of the most critical point in the formation of the solar system is the aggregation of the grains up to the size of the planetesimals. Although the density of solid particles was high enough to result in a mean free path around the centimeter, it is necessary to postulate a physical mechanism such as to favor aggregation rather than destruction following impacts.

This phase is still not completely clear today in its quantitative aspects. Computer simulations show that a rotating disk of planetesimals tends to break in a series of rings and these in a series of clouds. Even the growth of planetesimals until the formation of protoplanets requires special collision conditions. The speed of impact must be low, much lower than the escape velocity of the growing planet, and the materials must have cohesive properties.

Urey and Lewis equilibrium condensation theory

Urey and Lewis have listed the compounds that can condense (sublimate) into solid particles at various distances from the Sun in the protoplanetary disk under physic-chemical equilibrium conditions.

It is evident that only the most refractory compounds such as iron oxides and titanium compounds can condense at the distance of Mercury, while in the terrestrial orbit they can also condense iron and metallic nickel in addition to silicates, feldspars (silicate oxide with Al, K, Na, Ca, Fe ...) and troilite (iron sulfide, FeS) where the last is the result of chemical reactions between the iron and the sulfur, still in the gas phase. The presence of troilite both on land and on many meteorites demonstrates the validity of the equilibrium condensation model.

It is important to note that water ice cannot condense in the Earth's orbit. This process can only take place at

temperatures of about 170-200 K, much further from the Sun, near the orbit of Jupiter, in a region called the **snow line** or **frost line**, as it is also indicated by the evaporation of ice from the nuclei of comets during their approach to the Sun. Recent works that keep into account the opacity of the disk and of the outer "flaring", give the limit of condensation of the water around 2.5 astronomical units. Consequently the accumulation of water on the Earth can occur either indirectly from hydrated compounds of higher condensation point (such as serpentine) or from later processes of meteoric bombardment of bodies rich in ice like nuclei of comets, when the Earth surface was cooler enough to keep in the atmosphere the resulting water vapor.

A fundamental proof that the condensation model is correct, at least in first approximation, is the decreasing density of the planets of the solar system as they are farther from the Sun. Iron oxides and metallic iron, in fact, have a specific weight higher than silicates, and their specific weights, in turn, are higher than those of water or methane ice.

Water on Earth - Recent works that keep into account the opacity of the disk and of the outer "flaring", give the limit of condensation of the water around 2.5 astronomical units. Consequently the accumulation of water on the Earth can occur either indirectly from hydrated compounds of higher condensation point (such as serpentine) or from later processes of meteoric bombardment of bodies rich in ice like nuclei of comets (**late bombardment**).

Indeed, if we study the ratio Deuterium/Hydrogen on planets we can see that Earth has a value different from the other planets, but similar to comets. So, this implies that water on Earth comes from far objects as comets. This bombardment happens when the Earth surface was cooler enough to keep in the atmosphere the resulting water vapor.

Formation of protoplanets

When a protoplanet has grown enough, by aggregation of planetesimals, it can also capture gas directly from the interplanetary medium and create an atmosphere.

This gas must have been mainly constituted by hydrogen and helium, the two by far the most abundant elements both in the cosmos and in the Sun. According to recent studies, subsequent impacts can, due to soil erosion, produce gas and contribute to creating an atmosphere.

Elimination of the residual cloud

After the formation of planets the solar nebula contains a significant amount of dust and gas. This material must have been expelled because its presence in the long time would have led to the destruction of the planets making them spiral towards the Sun.

One of the proposed mechanism is the solar wind during the so called T Tauri phase. In this phase there is a strong solar wind generated by the convection of the outer layers, which is able to eliminate all the gas from the protoplanetary system.

Migration of planets: the Nice model

The Nice model is a scenario for the dynamical evolution of the Solar System. It proposes the migration of the giant planets from an initial compact configuration into their present positions, long after the dissipation of the initial protoplanetary disk. In this way, it differs from earlier models of the Solar System's formation.

This planetary migration is used in dynamical simulations of the Solar System to explain historical events including the Late Heavy Bombardment of the inner Solar System (which allowed to deliver water to Earth by comets), the formation of the Oort cloud, and the existence of populations of small Solar System bodies such as the Kuiper belt, the Neptune and Jupiter trojans, and the numerous resonant trans-Neptunian objects dominated by Neptune.

4.7.1 Evolution of a planet: the planet Earth

After the formation of the planets there had been a bombardment that produces a degassing and increased the temperature of the surface of the Earth: the surface was totally melted, it was a magma ocean. So the conditions were: very high temperature and new gases in the atmosphere. At the same time the original gases were collected by the gravitational force on the Earth. Originally the atmosphere was very rich in H and He, but these light gases were then removed due to the high temperatures or impacts. Then a secondary atmosphere was generated by other impacts and volcano eruptions when the surface started to cooling down.

An important problem is the **origin of water**: one option is that it has come indirectly through rocks containing hydrates, which are present in carbonaceous chondrites.

According to the Nice model, instead, there have been an episode, called **late bombardment** that could have played

a key role in enriching the Earth with water.

With the arrival of water the Urey reaction started and fixed the CO₂, while ammonia had been decomposed by UV radiation into nitrogen and hydrogen.

A dramatic decreasing of temperatures occurred when oceans were formed. The Earth entered in the so called **snowball phase** in which it was covered in ice. The albedo dramatically increased. This phase is not reversible but, somehow, Earth exited from this phase. Probably there were some strong volcano eruptions that increased the CO₂ in the atmosphere and, due to the greenhouse effect, the Earth left this phase. However it is difficult to understand how Earth kept its climate stable despite the minor luminosity of the Sun in that phase. This is called the **young Sun paradox**.

Sources of heat and the differentiation

Planets larger than a hundred kilometers collect enough material to produce a considerable heating of the rocks due to the decay of radioactive isotopes. The most important ones in this phase are ²³⁵U, ⁴⁰K and thorium. At the time there was also the Aluminum which has the fastest decay time and produces more energy in a short time therefore it is the responsible for the fusion process in the first evolutionary phases of the terrestrial planets.

Therefore, heating can melt the inner rocks of the planet and maintain the state of fusion for a period sufficient to allow the differentiation of the material, i.e. the separation between heavy materials (iron and nickel) towards the interior and light materials, such as silicates, on the surface.

The condition for differentiation is expressed by the relationship between the surface temperatures $T(R)$ and the central temperature $T(0)$:

$$T(0) = T(R) + (a/6c)R^2$$

Where a is the production of energy per unit of volume and c the thermal conductivity. In order for differentiation to occur, the difference in temperatures must exceed 1800 K for rocky bodies composed of silicates and about 150 K in the case of ice. The calculation with the typical values of the planets gives a radius, in both cases, of a few kilometers. It is therefore concluded that the process of differentiation was very common both between the planets and between the asteroids. Only smaller bodies have retained the original composition.

4.8 Dating the surface of terrestrial planets

Dating planetary surface is fundamental in order to understand planetary evolution. The dating of the surface through the counting of meteoric craters represent a very important technique given the variety and uniqueness of the surface structures.

The meteoric impact on a planetary surface forms a crater with dimensions that can be expressed empirically according to the kinetic energy of the body that causes the impact:

$$D = (498.7 dr^3 v^2)^{1/3}$$

where D is the diameter expressed in km, v is the speed, d the density and the number is a constant to normalize the relation. Usually big craters are rare and surface is more uniform going to smaller diameters.

In absence of atmosphere or other erosive agents, meteoric craters grow in number with time and therefore their surface density increases.

In older lands there could be even a **saturation** of the ground when the number of craters no longer grows.

This method is based simply on the assumption of a progressive increase in density of the crater with age. Furthermore, the number of craters is inversely proportional to the square of their diameter:

$$N = \text{const} \cdot D^{-2}$$

Obviously, in a planet with an atmosphere and other erosive agents we must take into account the destruction of the smaller over time. the opposite effect is observed instead on the Moon where there is an increase of the small craters due to the production of secondary impacts from material expelled by larger impacts that fall around the primary crater.

Chapter 5

Exoplanets

5.1 Research techniques

There are different ways to discover an exoplanet:

- Direct imaging
- Astrometric perturbation
- Radial velocities
- Photometric eclipses: the transit method
- Radio detection
- Pulsation
- Radar analysis

Every technique allows us to know specific parameters about the planetary system or about the planet but every method has also some biases, a kind of selection of which kind of planets we are able to detect, due to the intrinsic nature of a specific method.

5.1.1 Direct imaging

It is the most obvious way but also the most difficult one due to the difference in brightness between the star and the planet.

Taking the planet Jupiter as an example, at the distance of 5 parsecs would appear at $V_J = 24.6$, while the Sun would be a star of apparent magnitude $V_\odot = 3.4$, with a separation from Jupiter (about 5 AU) corresponding to $1''$. The difference in luminosity between a planet like Jupiter and a star like the Sun is 21.3 magnitudes (about a ratio of $3 \cdot 10^8$). Actually there is no technique for the direct detection of a planet with these characteristics. Direct detection has recently been obtained on some planets where the separation is wider (e.g. 2M1207b has a separation of 54 AU) and the star/planet luminosity ratio is much lower (about 100).

The situation gets more complicate if we are looking for terrestrial planets since their visual magnitude is even smaller.

Other techniques of direct detection in the radio domain, at large wavelengths where synchrotron magnetospheric emission of planetary origin is strong, are interesting for possible future developments.

Obtainable parameters

From direct imaging we can obtain:

- brightness of the planet
- eventually the distance of the planet from the star if you know the distance of the star from the Sun.
- the period, even if this parameter is very difficult to find because there is a strong bias

Therefore this method is suitable only for spectroscopy but nothing more.

Bias

We can detect only planets who are not so close to their star, at 30/40 AU, like Pluto. Their period is very long so it is almost impossible to follow their orbit.

5.1.2 Astrometric perturbation

This technique is based on the astrometric measurements of the motion of the star around the common center of gravity.

From classical mechanics, the equation that describes the equilibrium around the barycenter is:

$$M_s a_s = M_p a_p$$

where M_p and a_p are the mass and distance from the center of gravity of the planet, while M_s and a_s are the ones relative to the star.

The mass of the star is dominating, can be deduced from stellar models. Observing the motion of the star around the barycenter position, we can deduce the radius a_s which is a small number. Then, observing the star moving, following its path, you can derive the period P .

Now you just have to use the third Kepler's law and you have a_p :

$$P^2 = \frac{4\pi^2}{G(M_s + M_p)} a_p^3$$

The only unknown parameter is the mass of the planet M_p .

Obtainable parameters

From this method we have the following parameters:

- distance from the center of gravity of the star a_s
- period of rotation P around the center of gravity
- distance of the planet from the center of gravity a_p
- mass of the planet M_p

What is important is to understand that we can get the mass of the planet because the motion of the star is driven by gravitational force.

Bias

The discoveries are based on the amplitude we can measure. The amplitude of the astrometric orbit depends on two factors: the distance of the planet from the barycenter and the mass of the planet. Therefore the selection favors the discovery of planets far from the star and of great mass, and/or orbiting planets around stars of small mass.

This is the oldest methods used in the detection of exoplanets however it has never been successful because you need a very high accuracy.

5.1.3 Radial velocity

Almost 1000 exoplanets known to date have been discovered, or studied, through the study of relative star motion (velocity) around the common center of gravity, along the line of sight, from the Doppler effect.

It is known from classical astronomy that the radial velocity, along the line of sight of the observer, of a mass star M_s , due to a companion of mass $M_p \sin i$, with orbital period P , eccentricity e and inclination i is given by:

$$K = \left(\frac{2\pi G}{P} \right)^{1/3} \frac{M_p \sin i}{(M_s + M_p)^{2/3}} \frac{1}{(1 - e^2)^{1/2}}$$

where K is the half amplitude of the graph (R_v, t) shown in Fig. 5.1 so it is known from observations of radial velocity as function of time. The period P is known from the graphic as distance in time between two peaks. The eccentricity is derived from the analysis of the radial velocity curve while the mass of the star is known from isochrones or from the spectrum. Then the term $(M_s + M_p)^{2/3}$ is known due to the fact that M_s is actually bigger than the planet mass so M_p is negligible.

Therefore you can obtain $M_p \sin i$. The inclination i is the angle between the direction of the observations and the normal to the orbit.

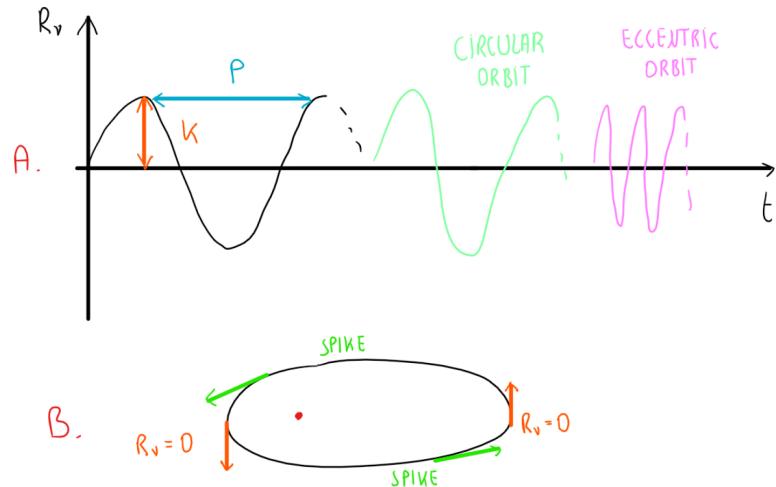


Figure 5.1: Radial velocity

Obtainable parameters

The most important parameter we can get through this method is the mass of the planet M_p . From the radial velocity graph we can also obtain the period P and, using the third Kepler law, we can derive the axis a

Bias

- K decreases with the increase of the period P , linked to the orbital radius, so the discovery of close planets is favoured.
- K increases linearly with the mass of the planet but depends on the projection of the orbit along the line of sight.
- K increases with the eccentricity of the orbit e .
- K decreases as the mass of the star increases.

In conclusion the method favors the detection of planets with great mass, high eccentricity and short periods (like hot Jupiters), possibly around small stars.

The accuracy of the radial velocity depends on the stability of the spectrograph, as well on S/N . However the bigger issue is the one related to the degeneracy of $\sin i$.

5.1.4 Photometric eclipses: the transit method

An alternative method of detection is based on the observation of the transit of planets in front of the star. The variation of brightness of the star is proportional to the ratios of the projected areas, i.e. to the squares of the radii:

$$\Delta I = \frac{R_p^2}{R_s^2}$$

If the radius of the star is known from the spectrum (usually it is assumed to be a main sequential star), the radius of the planet is then immediately obtained from the depth of the measured transit translated into a difference in intensity.

From the repetition of the transit we can also obtain the period and, if the mass of the star is known, we can also compute the distance of the planet from the star.

There can also be an occultation, with a minor depth, when the planet passes behind the star. In this case the total intensity is given only by the one of the star.

We can't measure the transit of Earth-like planets from the ground for several issues. First of all the limitation

is due to the statistics of the photons and to the atmospheric scintillation which depends on the wavelength, the pupil size and the distance of the turbulent layer of atmosphere from the pupil. Therefore if we want to detect this kind of planet we have to go into space.

In order to detect a transit, it is necessary a specific geometric condition. That means that we can observe a transit only if the planetary system is on the line of sight. The probability of observing transits is very low because it requires that the tangent of the angle of inclination between the planet's orbit and the direction of observation is smaller than the ratio between the radius of the star and the radius of the planet's orbit around the sun:

$$\tan \alpha \leq \frac{R_s}{a_p}$$

and this occurs only for small angles.

Obtainable parameters

This method allows to find the period P and the radius of the planet. We can also get the orbital axis. Of course knowing the radius from transit and the mass from radial velocity we can also compute the density, a fundamental parameter to investigate the nature of the planet.

Bias

- planets transiting near the star are favored because of their short period compared to the transit time
- small mass stars and therefore small radii are favored
- the continuity of observations and the duration of the same is very important in order not to miss the rare transit event
- instrumentation plays an important role because signal to noise ratio increases with the telescope diameter both because it improves the photon statistics and because, from ground observations the effect of atmospheric scintillation decreases.

Summarizing, this method introduces a selection for smaller stars and bigger planets with a small tilt in orbit. Of course there could be some false positive: in order to understand if it is a double system of stars or a planetary system we need confirmation from the radial velocity method.

Chapter 6

Supernovae remnants

Supernovae are traditionally classified (old, simplified classification) into two main categories based mainly on spectral characteristics. While in the type I supernovae spectra the hydrogen lines do not appear, in the type II SNs the lines of the Balmer series are identified, in particular the H α .

Type I

Origin:explosion - SNI comes from a binary system, usually composed by an evolved giant star and a white dwarf. The process implies transfer of mass from the giant star to the dwarf till the last one reaches a critical mass and explodes in a thermonuclear explosion generating heavy elements, like iron.

Features - This type of supernovae has a light curve with a period of about 50 days around the maximum followed by an exponential decrease with a scale of around 60-80 days. The color index is around $B - V = 0.5 - 0.7$. The optical spectrum is characterized by wide emission bands. Type I SNs are observed in all type of galaxies and for this reason they are associated with stars of intermediate mass of population II.

Recent studies suggest a further subdivision of this class into SN Ia and SN Ib. The last ones are systematically weaker and therefore closer to the SN II.

Type II

Origin:implosion - SN II came from intermediate stars with masses of the order of $8-9 M_{\odot}$ which generate all the heavier elements than H in thermonuclear reactions. This SN occurs thanks to a collapse of the iron core when it reaches a mass about Chandrasekar limit, forming a neutron star with $1.4 \simeq M_{\odot}$.

This collapse is translated into thermal energy and kinetic energy ejecting the outer layers. Moreover the implosion triggers a shock-wave with very high velocity.

For their origin these SN are called also **core-collapse supernovae**.

Features - These SN have different light curves. The classification is based mainly on the characteristics of the spectrum that shows, a strong continuous in the blue and weak Balmer lines.

Type II supernovae are characteristic of spiral galaxies and are associated with population I stars.

Statistics indicate that SN Ia are about twice the one of type II. However we have to take into account that there are some selection affects that tend to favor SN Ia since they have an higher magnitude at the maximum ($M_{v,Ia} = -19$ vs $M_{v,Ib} = -18$).

6.1 SN remnants

A supernova remnant is the structure resulting form a supernova event. The SNR is bounded by a shock wave and consist of ejected material expanding. The interstellar material is swept up and shocked along the way. In Fig. 6.1 there are some examples of SNR:

1. Crab Nebula in Taurus
2. Cassiopeia A
3. Cygnus Loop

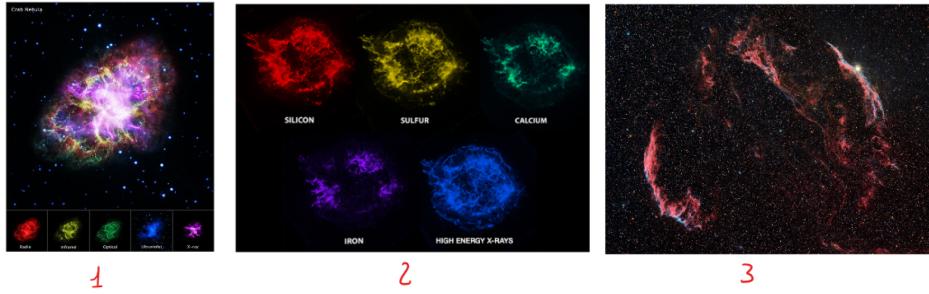


Figure 6.1: SNR examples

6.1.1 Statistical results on supernovae remnants

An important diagram for study the evolution of supernovae is the one plotting the surface brightness and a function of the diameter of the source, both in logarithmic scale.

If the expansion mechanisms were identical and the energy involved comparable for all supernova remnants, the meaning of the diagram would be purely evolutionary.

An expanding shell decreases in brightness exponentially over time. We therefore expect that the diagram can be interpolated with a power law:

$$\Sigma = \text{const} \cdot D^{-3.7}$$

where D is the diameter and the slope is 3.7. This graphic is shown in Fig. 6.2. From this figure we can see that points

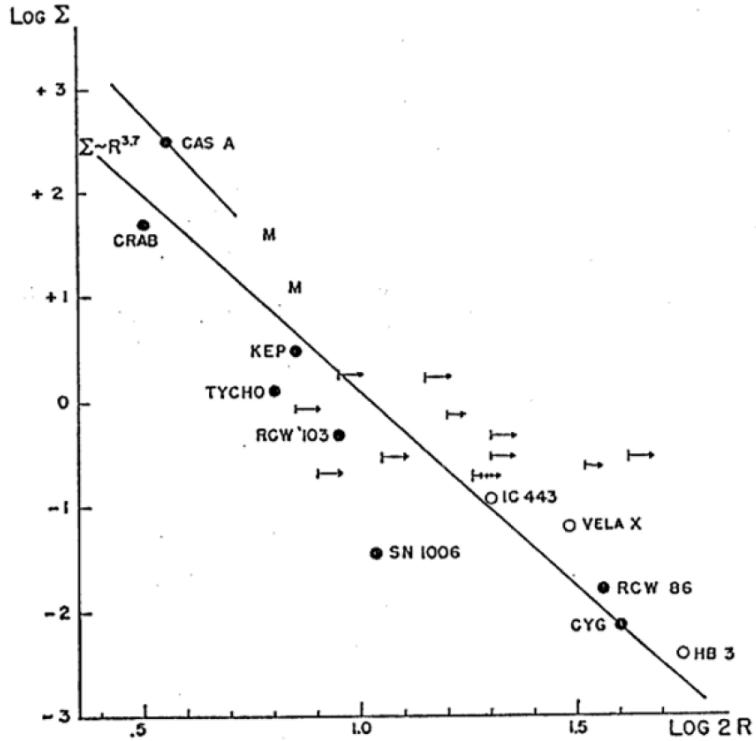


FIGURE 1: The Σ - R diagram. ●—well determined distances; ○—distance from interaction with H II region; \rightarrow — H_2OH , or C_2H_2O absorption distance; \leftrightarrow —same but poor; M—Magellanic Cloud objects. The line through Cas A represents its present evolutionary track.

Figure 6.2: Surface brightness vs real diameter

are in according with the evolution: the diameter increases and the brightness decreases, becoming cooler with time. In spite of their double origins there is small dispersion: this means that they somehow lost memory of their origins quite soon after the generation of SN remnants. This implies that the process is driven by other factor than the initial velocity ejection and initial energy.

Pay attention: in order to know the real diameter of the remnant we have to know the distance.

Alternative methods for SNR distance

Another empirical result is that once we have this relation for well known SNR we can observe any kind of SNR, then using the surface brightness we can find the diameter. then, comparing the real radius to the apparent one we can get the geometric distance, independent on reddening and position.

Therefore while the nebular method can be applied only to nearby stars visible in optical range, this last method is more powerful and can be applied also to SNR detected in radio.

Brightness-diameter relation

Clark and Caswell made an interesting work about the relation between brightness and linear diameters of SNR. From this work they found out that the most of supernova remnants expand in a phase in which the expansion is not free but it is conditioned by the interstellar medium. There is also a considerable dispersion. This indicates that galactic supernovae produce expansion velocities in a broad interval. The phenomenon of evolution could be much more complex than what we assume with the hypothesis of a single phase of free expansion.

Another important result is that there is no correlation of the spectral index with the diameter of supernova remnants, suggesting a rapid expansion compared to the internal energy decay of relativistic electrons. In other words, the kinematic evolution is faster than the loss of energy of SNR.

6.2 Evolution of supernovae remnants

Current evolutionary models foresee, as originally described by Woltjer and from Ilovaisky and Lequex in 1972, three distinct phases of evolution:

1. short-term free expansion in 100-300 years
2. adiabatic expansion with shell development
3. isothermal expansion accompanied by cooling by radiation
4. total mixing with the interstellar medium

Most supernova remnants (SNR) would be found in the second phase or between the second and third one. Only the youngest, such as Cassiopeia A, may have just come out of the first phase, while at the other extreme we have the Cygnus Loop which should be in isothermal phase.

6.2.1 Free expansion

The free expansion phase assumes that the density of the interstellar medium is negligible compared to that of the rest of supernova (SNR). SNR gas acts as a piston that compresses the interstellar gas and the speed of the ejecta does not undergo an appreciable deceleration.

The expansion occurs between low density gases with mean free path of hundreds of parsecs. This implies that there is no direct collision between particles. The interaction between SNR particles and the interstellar medium occurs through magnetic coupling. In the typical conditions of supernova explosion the perturbation of the medium is quite strong, with velocity much higher than the speed of sound, causing the formation of shock waves.

Unlike sonic waves, the approximation of $dP/P < 1$ (normally used for sonic waves) is not applicable and the wave deforms with the peak which tends to anticipate the entire perturbation, given that the speed of propagation of the perturbation is greater in the center, at higher density. Therefore it creates a discontinuity of limited spatial extension where the material of the surrounding medium undergoes a sudden compression and thermalization which means that interstellar gas is suddenly shocked and reaches very high pressure and density. Gas is not removed but it is shocked.

The classic solutions of the shock wave (currently called shock), in this situation, they can easily be obtained by imposing the condition of continuity and the equivalent of the laws of motion, in the condition generally verified in the initial phase, of adiabatic transformations. The post-shock density, pressure and temperature, as a function of density of the medium and the speed of expansion (Sedov solutions), are derived:

$$\begin{aligned}\rho &= 4\rho_0 \\ P &= \frac{3}{4}\rho v^2 \\ T &= \frac{3}{16}m\frac{v^2}{K}\end{aligned}$$

where P and ρ are the post-shock pressure and density, ρ_0 is the density of the medium, T is the post shock temperature, v the velocity of the shock wave and m the average molecular mass.

6.2.2 Adiabatic expansion

These conditions remain valid also in the second phase, when the SNR density, as a result of expansion, has become comparable to that of the medium (or, in other words, the amount of swept matter from the shock wave it becomes comparable to that of the supernova itself). At this point the speed starts to significantly decrease. This condition is reached when the radius $R(t)$ satisfies the relation:

$$\frac{4}{3}\pi R(t)^3 \rho_0 = M_0$$

where M_0 is the ejected mass and ρ_0 s the density of the interstellar medium.

However, energy conservation conditions continue to apply (we can also demonstrate that kinetic and thermal energy are conserved separately):

$$\frac{1}{2}M(t)v^2 = E_0$$

where E_0 is the initial energy and is comparable to the kinetic energy released by the explosion.

The total mass is variable and increases at expense of the interstellar medium. Under the same conditions of the explosion, the transit to this phase is shorter, greater is the density of the interstellar medium. The only energy losses at this stage are due to free free transitions ans free-bound emission lines. The previous equation can be written as:

$$\frac{1}{2}\left(\frac{4}{3}\pi R(t)^3 \rho_0 + M_0\right)v(t)^2 = E_0$$

where the first term is the mass og the interstellar medium while M_0 is the mass of the original ejected material. Of course, as $R(t)$ increases with time, as a consequence of increasing mass the velocity v decreases and therefore shock waves slow down. The integration of the last equation gives the radius as a function of time:

$$R(t) = \text{const} \cdot \frac{E_0^{1/5}}{\rho_0} t^{2/5}$$

This last equation together with the ones that regulate density and temperature in the wave fronts, describe the conditions of the supernova remnants in the phase of adiabatic expansion. This equation is also important to derive the characteristic parameters of the SNR.

Assuming a density of interstellar gas around to a hydrogen atom per cubic centimeter, we obtain that the enrgy release from the explosion of the supernova remnant in kinetic form is around $10^{51} - 10^{52}$ erg.

This is about two orders of magnitude lower than the total graviational collapse energy released (mainly as neutrinos energy) by the supernova. Once this parameter is set, the relationship radius–time is therefore defined. In a histogram of distribution of the number of lower SNRs at a certain radius, against the radius we should therefore expect a straight line with slope 2.5 on a logarithmic scale. Moreover from the observed radius the age can be easily obtained.

Energy losses are small, due mainly to the free free emission expressed by:

$$-\frac{dE}{dt} = \text{const} \cdot n_e^2 t^{1/2} \cdot \text{volume} \propto t^{3/5} E_0^{4/5}$$

As the evolution proceeds the gas cools down, the shock wave gets slower and the loss of energy through free free slowly increases as a result of the increase of the total mass which compensate the temperature decrease.

When the temperature drops to a few million degrees, the radiation from free-bound and bound-bound transitions, due to elements heavier than hydrogen (C, O, N), gets dominant. In this situation energy loss became relevant:

$$-\frac{dE}{dt} = \text{const} \cdot \rho_0^{1/5} E_0^{9/5} t^{12/5}$$

The high power index of time, compared to the free free losses, demonstrates what has been said. We can define a time t' , at which half of the initial energy is irradiated and beyond that the transformation are no more in adiabatic conditions:

$$t' = 5E_0^{4/17} \rho_0^{-9/17}$$

6.2.3 Isothermal expansion

Then SNR enters the isothermal phase: the energy transferred to circumstellar material (CSM) is nearly immediately radiated leaving temperature T unchanged. Therefore this phase is characterized by constant temperature in a very thin outer layer. At the same time, just behind this thin layer, density is rapidly increasing to a value which is much more than a factor 4, due to the cooling. Indeed gas can not oppose anymore with thermal pressure and the density increases. Moreover in this phase the momentum (mass per velocity) is constant.

6.2.4 Last phase

Then there is a total mixing with the interstellar medium at a sonic speed.

We know that the evolution proceeds faster at higher densities of the interstellar medium.

Hot cavity of SNR - Suppose to have an explosion of SN in a star cluster. the heat released makes the gas expanding and decreasing its density. In these condition, a second massive exploding star expands in a hot cavity with very different expansion time due to the minor interstellar density.

This means that we may not see the second explosion of SN until it is so fast to reach the edge of the previous one.

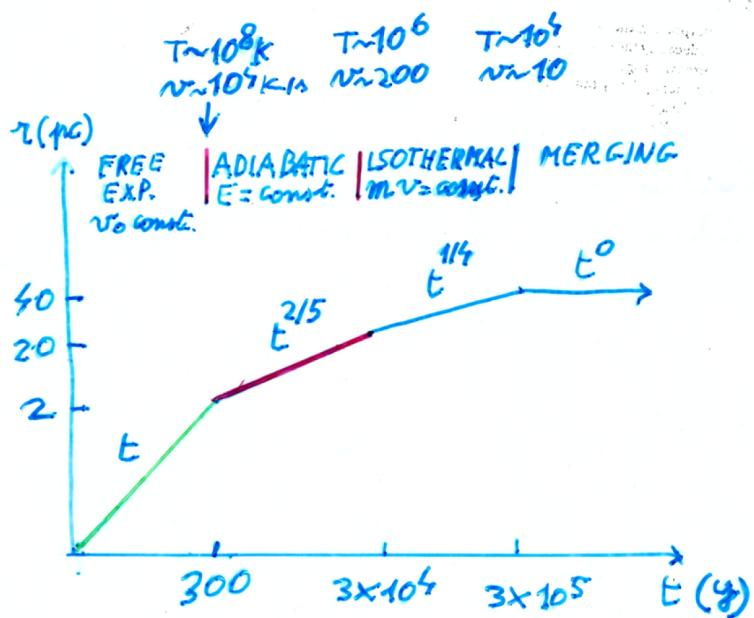


Figure 6.3: Evolution of the size with time

Chapter 7

Maser

7.1 Molecular lines and maser emission in the galaxy

Molecules can have 3 types of quantized transitions: electronic, rotational and vibrational. A molecule has a total energy E_t that could be written as the sum of these three terms:

$$E_t = E_e + r + E_v$$

Electronic transitions with energy E_e are transitions of electrons between different quantum levels. Usually they are not visible in IR or radio because they are very energetic so they are usually visible only in optical and eventually X-ray windows. There are some exceptions like the HI line at 21 cm.

E_v is instead the energy necessary for vibrational transitions inside molecules. The vibrational levels are therefore of low energy with the highest energy level corresponding to the molecular binding energy. Transitions between vibrational states produce lines in the infrared regions up to about 10-20 micron. On the other hand the vibrational transitions are not observed normally in the optical domain for spectra of astrophysical importance. The relative vibrations of two atoms can be seen as harmonic oscillations. These are quantized.

In addition to vibration, the molecules have another degree of freedom which is the possibility of rotating around the molecular center of gravity. Also this energetic state is quantized and the energy differences between the rotational states are in general smaller than those of the vibrational states. The vibrational states are so further divided into rotational levels with smaller separations. Transitions between these states lead to lines especially in the radio field. As a first approximation most of the molecules can be compared to rigid rotator, obtaining as a result a series of equally spaced lines. Transitions between sublevels in which individual rotational lines are divided are also important in the radio region. Important examples are the Λ splitting and the separation of hyperfine structure.

Therefore in general is valid:

$$E_e > E_v > E_r$$

As concerning E_v and E_r we need a condition about electronic dipole emission. If the molecule is symmetric, there is no dipole emission. So in the case of the most common molecule in the interstellar medium, H₂, we can only observe transitions in UV and not in radio or IR.

7.2 Emission of CO

Among the most important interstellar molecular lines certainly the rotational lines of the CO at 2.6 mm occupy a dominant position. CO is among the most stable molecules and derives from the combination of the two most abundant elements after hydrogen and helium. Although its abundance is about 104 times lower than that of the H₂ molecule, the CO has been detected everywhere in the galactic disk, with high concentrations in regions of small diameter and small dispersion of radial velocities, identifiable with compact molecular regions of low temperature (about 10 K). The CO emission maps in the radial-galactic velocity diagrams are very similar to those obtained from the 21 cm of the hydrogen, but at high resolution show a complex structure of compact regions. The CO lines are currently used both to obtain the molecular abundances in our galaxy and for the temperatures and density of the Bok globules (small parts of dark nebulae).

The relative abundances of the lines originating from $^{12}\text{C}^{16}\text{O}$ against $^{13}\text{C}^{16}\text{O}$ allow, in theory, to derive the isotopic interstellar abundances $^{12}\text{C}/^{13}\text{C}$. However, the measurement is made difficult by the frequent saturation of the $^{13}\text{C}^{16}\text{O}$, from which the temperature of the clouds is more easily obtained. For the isotopic ratios it is preferred today to use instead the $\text{H}^{12}\text{C}^{16}\text{O} / \text{H}^{13}\text{C}^{16}\text{O}$ ratio since these lines are almost always optically thin.

7.3 Emission of OH

OH is a quite common molecule and it is an asymmetric molecule that has two different orientation for rotation, with different momentum depending on the rotational axis we consider. A transition between these two axis generates the line at 18 cm called Λ doubling.

It should be noted that, unlike the hyperfine structure transition of the 21 cm hydrogen atom, these transitions are permitted. Their probability of transition is about 4 orders of magnitude higher. This justifies the observed intensity of these lines despite the effect of a factor 7 of less abundance than OH compared to H.

The line at 18 cm has often been observed in emission. The lines are relatively narrow, corresponding to Doppler widening of 100 K. We can calculate the transition probabilities of each of this line and we can check from observational data if they correspond to theoretical expectation. They do not: the relatively intensities are very different from what we expect. Moreover, when they have been detected for the first time, it has been noticed that these lines are very bright, very thin and also with a high degree of polarization, both linear and circular. So nothing to do with other common lines.

All this indicates that the emission of OH does not occur in conditions of thermodynamic equilibrium. This is why the 18 cm lines of hydroxyl are now explained with the maser mechanism.

7.4 Maser emission

The MASER emission mechanism is based on the stimulated emission in environment where there is a metastable state of excitation called "population inversion". This term indicates that the number of atoms or molecules present in a high energy level is greater than that of the atoms or molecules that are at a lower energy level. In conditions of local thermodynamic equilibrium this is not generally possible as the collisions tend to redistribute the populations according to the Boltzmann equation, that is with a population that decreases exponentially with increasing energy levels:

$$\frac{n_1}{n_0} = \frac{g_1}{g_0} e^{\frac{h\nu}{kT}}$$

It is clear from this equation that, for unitary ratios of the statistical weights of levels, there is no temperature value T for which the ratio n_1/n_0 can become unitary. For values of T close to zero the ratio n_1/n_0 also tends to zero, while at increasing values of T , the ratio increases, while remaining always less than 1 and tends asymptotically to this value for T increasing to infinity.

From the formal point of view one can have a population inversion only by introducing in the equation negative temperature values.

From a physical point of view, deviations from this equilibrium condition are possible only if the density is lower than a critical value, which depends on the radiation field and on the atoms / molecules present, to prevent collisions from "thermalizing" the distribution with greater speed of the opposite process that disturbs the balance.

7.4.1 Three-levels maser mechanism

The more efficient mechanism to overpopulate the upper level is the so called three-levels maser.

The molecules are somehow pushed to the excited state (3) by a pumping mechanism. From this upper level (3), by spontaneous transition and emission, mostly in the infrared, they pass to an intermediate level (2) which represents the upper state of the maser transition. If the probability of transition between the two upper levels is much greater than that between the upper (3) and lower level (1), then the depopulation of the lower level (1) can be very strong to the advantage of overpopulation of the upper level (2).

The transition between the upper maser level (2) and the lower level (1) can occur spontaneously or by **stimulated emission** by photons. It is well known that the stimulated emission by photons requires inducing photons with an energy identical to the energy difference between the two levels. From a single exciting (triggering) photon two identical photons are produced.

In the case of population inversion the probability of stimulated emission far exceeds that of absorption. As the radiation of photons, corresponding to the transition (1-2) propagates through the emitting medium, it is amplified by **stimulated**, or **induced emission**. Moreover, since the emitted photons have the same characteristics as the photons that induce the transition, the stimulated radiation is coherent. This process in the laboratory is called laser, but maser in the case considered of emission in the interstellar medium in the radio field.

The maser mechanism is maintained until the energy pumping remains active. There are three main types of pumping:

1. photon pumping, in which the upper states are populated by photon absorption
2. Collisional pumping in which collisions populate the upper state. This type of pumping is often used in artificial lasers where collisions are due to electrons accelerated by electric fields.
3. Chemical pumping. In this case the molecules are preferentially created in excited states, including the upper ones responsible for the maser emission.

In general, however, none of the three mechanisms alone is capable of explaining the observed transitions, even if in some cases it is clear that an intense IR flow excites the higher states of the maser.

A characteristic of maser radiation is that of being strongly polarized. The stimulated emission in fact releases photons that have the same polarization direction as the incident photon. This means that all the photons that derive, directly or indirectly, from the same progenitor, are all polarized in the same direction and therefore the radiation is 100% polarized.

The production of photons by stimulated emission, under certain conditions, can become higher than the re-population of the upper level by pumping. In this case the efficiency of the maser is strictly linked to the efficiency of the pumping mechanism and the amplification of the signal will be strongly limited. In this case we say that the maser is "saturated". The saturation effect can induce rapid apparent variations in the angular dimension of the sources because the amplification, and then the depopulation, is higher along the maximum length segment.

The maser activity requires some fundamental conditions:

- a high optical depth of OH lines in order to have interaction between photons and atoms ($\tau_{OH} > 1$) but it should be low in IR otherwise the photons cannot go inside the cloud deep enough to overpopulate all the cloud ($\tau_{IR} \approx 1$)
- the gas density must be much higher than that of the interstellar spaces (at least 10^5 atom/cm⁻³ while interstellar medium has a density of 1 atom/cm³) to have amplification. Indeed, if density is too low, particles don't interact
- a high brightness source must be present ($L > 10L_\odot$) to provide pumping energy

Amplification

Therefore maser is a continuum mechanism with very low efficiency converting flux coming typically from a star into energy of a peculiar and very amplified line, such as OH. Indeed this line is very intensive but pay attention: the peak intensity doesn't correspond to a physical brightness temperature, which instead is around $10^{10} - 10^{12}$ K, out of thermal equilibrium. It is just a peculiar amplification.

Moreover amplification is very peculiar for two reasons:

- it is an exponential amplification;
- it violates the transfer equation: greater is the optical depth in the line, more intense is the output radiation. Therefore for masers we have a negative temperature T and negative absorption k .

7.5 Stellar and interstellar masers

There are two main types of maser sources in our galaxy: **interstellar masers** and **stellar masers**.

Interstellar masers

Interstellar masers are associated with regions of recent star formation and pumping energy can be provided by young stars type O-B, while the maser is triggered by compact condensations of interstellar material located near these stars

Stellar masers

In stellar masers instead, the emission takes place in the envelope of M type giant or super-giant stars and the energy source is the star itself.

The interstellar and stellar masers differ substantially in the trigger mechanism, that is in the origin of the photons responsible for the stimulated emission.

In interstellar masers these photons are provided by radiation outside the cloud and therefore have a very precise direction (source-cloud-observer), while the stellar ones are usually triggered by spontaneous transitions within the circumstellar shell and can therefore radiate in all directions.

In stellar masers the radiation that reaches the observer comes only from two narrow caps located in the star-observer direction, as the maser amplification requires a uniform speed field within the material where it is generated because velocity gradients produce a Doppler effect and the triggering masers cannot amplify relatively moving molecules.

In general, the spectrum shows a characteristic splitting of the lines due to shell expansion.

7.5.1 Distance measures with stellar maser

We know that the diameter of the shell around the star has a size of several AU. Therefore, even if the radiation from star allows pumping increase in both direction at the same time, from the observer the amplification from the cap below the star, on the opposite side, takes more time to arrive due to the diameter of the cloud. As result, we measure a non-synchronous variation. The difference in time is equivalent to the time necessary to photons to pass all the shell. Therefore, knowing the speed of light, we can get the true diameter of the sphere. Then, measuring by interferometry the apparent size, we can get the real distance using a purely geometric method.

This technique is very useful: in this way we can measure distance stars thought the disk because measurements are in radio so there is no absorption. So this method can be used to map the kinematics of galactic disk as function of the distance from the GC. This is fundamental to study black matter.

7.6 Usefulness of masers

The small angular size of the masers makes them suitable also for high precision astrometric measurements. Exploiting the expansion motions of the various sources of a cloud, or the statistical parallaxes of a cluster, or even the proper motion due to the galactic rotation, it is possible to obtain geometric distance measurements on a galactic and extra-galactic scale.

The OH maser allowed accurate measurements of the Sun-galactic center distance.