.

Fundamentals of

EXOPLANETS & ASTROBIOLOGY

From the origin of the universe, stars and elements,
and the creation of solar systems and exoplanets,
to the arise and development of life

Uffe Gråe Jørgensen, March 2023.

Part 6: The case for extraterrestrial life

for lectures 2023, March 28 and 30

1

Life beyond Earth ?

1.1 Is there extraterrestrial life in our solar system ?

The Universe was created slightly less than 14 billion years ago. It was almost homogeneous and the baryonic matter consisted of only hydrogen and helium. Fluctuations from homogeneity made clouds collapse and form stars and galaxies throughout the Universe. Simple calculations showed us that very basic physical principles governed the cloud collapses, so we expect stars as those we see in the sky throughout the entire Universe. Basic physics demanded the collapsing gas clouds to form flat disks around the stars. There were no special demands for planetary formation in the disk. It just required that that the cloud was enriched to a certain level with elements heavier than helium. We know that the whole Galaxy is enriched with such elements due to a homogeneous distribution of exploding supernovae and red giant winds throughout our Galaxy and throughout time. We have all reasons to believe that some kind of planetary systems are the rule of star formation, and extrapolations from results of exoplanetary research indicates there are more planets than stars in our Galaxy, with the majority of the planets being Earth-sized. We also have reasons to believe that large cold giant planets are rare and will form only in the outer parts of solar systems, and only around stars of relatively high metallicity, while small solid planets can form in all parts of the nebula-disk and around all types of stars. We have reasons to believe that it was a normal process that brought the water that formed the oceans on our planet, even though standard models of the solar system formation would predict Earth-like planets to form dry, and there still is no consensus about how, when and from where the Earth's water came. Water may therefore be normal or very rare on Earth-like exoplanets in Earth-like orbits, and it may or may not have relation to the existence of giant gas planets in large orbits. We can also see a hint of a process that could form the first biological cells in a natural non-mysterious process in the ocean; maybe based on large pre-biological molecules that came from space, or maybe formed entirely on Earth itself, depending on the composition of the unknown early atmosphere. We understand to a large extend why life selected the molecules it selected, or at least why it selected some of the most basic molecules at work in the living cell. We understand how this first primitive cell can have evolved through evolution over giga-years to intelligent animals like ourselves, driven by the principles of evolution, as outlined in Darwin's theory. It seems that intelligence was not a driving force in evolution, and its appearance seems to have been caused by a random road to present day life, involving among other things the huge evolutionary effect of a small number of random cosmic collisions. We suspect that intelligence may therefore be rare in the Universe, but at least life in some form could be common – it developed based on common universal principles, and even intelligence might in principle be common if we have not yet understood the basic evolutionary principle that might be behind it.

At this stage we badly need some observational facts to guide the further development of our theories and understanding. The discovery of lifeforms outside the Earth, elsewhere in space, would be the greatest step forward in this understanding, which I personally can imagine at this moment. The discovery of signs of life on an exoplanet (in the form of a low-entropic atmosphere, out of chemical equilibrium) would be a tremendous new input to our understanding of life. However, we may be disappointed in finding something which we cannot really understand (e.g. an atmosphere out of equilibrium in a way we cannot easily model, or which we can model as due to biological as well as due to geological processes), or to realize that we cannot distinguish primitive bacteria from intelligent higher life forms in this way. An even larger break through in understanding of life, may therefore be finding any living cells (or fossils of such) on another body in our solar system, although such a finding might be confused with pollution of microbial life from Earth. Where could we reasonably imagine that we might find alien life forms in our own solar system?

1.2 Venus

Venus is the planet in our solar system that originally must have resembled the Earth the most. It has approximately the same mass and size as the Earth, meaning that independent of whether the volatiles that formed the atmospheres came from outgassing or from collisions with volatile rich objects from the outer parts of the solar system, the amount on Earth and Venus must have been the same, which is the basic assumption of why the "climate model" presented in Fig. 1.1 of the relation between sunlight intensity, $S_{\rm eff}$ (=$f_\odot^x$(t)/$f_\odot^\oplus$(today)), and surface temperature is the same for Earth and Venus. Right here $f_\odot^x$(t) explicitly refers to the solar constant of one of the two planets, x∼Venus or Earth, while we will at other places just use the more commonly adapted $f_\odot$ for $f_\odot^\oplus$(today) as meaning the solar constant at Earth today, i.e. the amount of radiation energy that reach the Earth per area per second outside the atmosphere. Venus distance from the Sun is 0.7 AU, meaning that the solar constant at Venus today is,

$$f_\odot^{\rm Venus} = \frac{1}{r^2_{\rm Venus,AU}} f_\odot \approx 2 f_\odot \;\;\Rightarrow\;\; T_{\rm eff,Venus} \approx \sqrt[4]{2}\, T_{\rm eff,\oplus} = 1.2\, T_{\rm eff,\oplus} \tag{1.1}$$

At the time of the planetary formation, the solar luminosity was approximately 30% lower than it is today, and the early radiation influx on Venus has therefore been only ∼0.7 times of what it is today, or ∼0.7 x 2 = 1.4 times the present day solar influx on Earth, as illustrated in Fig. 1.1 together with an estimate of the temperature development (i.e. a "climate model") of the two planets as function of increased solar luminosity.

We found in the exoplanet chapter on habitability that $T_{\rm eff,\oplus}^{\rm today}$ = 246 K = -27°C, which would imply that when the Sun's luminosity was 0.7 times the present luminosity the Earth's effective temperature would have been $T_{\rm eff,\oplus}^{\rm t=0}$ = $\sqrt[4]{0.7}$ x 246 = 225 K = -48°C (which we used in the part about life on Earth to conclude that life changed the greenhouse heating on Earth from ∼60°C in the beginning to 40°C now in order to keep the surface temperature on Earth constant at 13°C throughout Earth's history).

These same arguments give us that $T_{\rm eff,Venus}$(today) ≈ 1.2 × 246 = 295 K = 22°C and ∼ 1.2 × 225 = 270 K = −3°C when the Earth and Venus were young. In other words, if Venus had had a completely transparent atmosphere (and an albedo of 0.39 as assumed in deriving the 246°C of Earth today) it would have had a temperature of −3°C at t=0 and 22°C today. If it had been able to replicate the (presumable life-driven) development of the atmospheric greenhouse development that Earth had, it would have started out with a surface temperature of ∼60°C and kept that temperature throughout its entire history of existence.

We would consider 60°C a good temperature for the origin of life, so why did life not originate on Venus and kept the temperature right for itself at 60°C all through Venus' history, as we (maybe erroneously) argued that life did on Earth at 13°C? Obviously something went "wrong" for life on Venus, and the atmospheric greenhouse heating rised to ∼450 K rather than going toward 0 K (which would have given a pleasant "holliday temperature" of 22°C).

Remark as a curiosity that it is a consequence of the above calculations that if one added the present-day Earth's greenhouse heating of 40°C to the 22°C, one would likewise reach a present day surface temperature at Venus of "only" ∼60°C, and if one added it to the initial effective temparature of Venus of -3°C and assumed that Venus had had this greenhouse heating of 40 K all life long, the surface temperature of Venus would have developed from 37°C in the beginning to 62°C today, while Earth with similar assumptions would have developed from −9°C in the beginning to +13°C today. Usually the habitable zone is defined as the distance from the host star where the surface temperature of a planet is between 0°C and 100°C (water in liquid form) if the albedo and greenhouse heating is the same as present day Earth. We therefore see the funny conclusion that in the standard definition of the habitable zone, only Venus in our solar system has continously been inside the habitable zone throughout all of its lifetime, whereas Earth has been outside the habitable zone (too cold) most of Earth's life, and even today is on the lower temperature end of the zone.

This obviously aught to make us take some precautions when we see postulates in literature (or in the media) about some specific exoplanet being in the habitable zone of its star. What is really wrong with our definitions, our understanding, and our way of looking at the whole problem? We just don't know. The picture is too simplistic and our understanding of what is required for life to arise and develop is at best insufficient, and at worst completely un-understood.

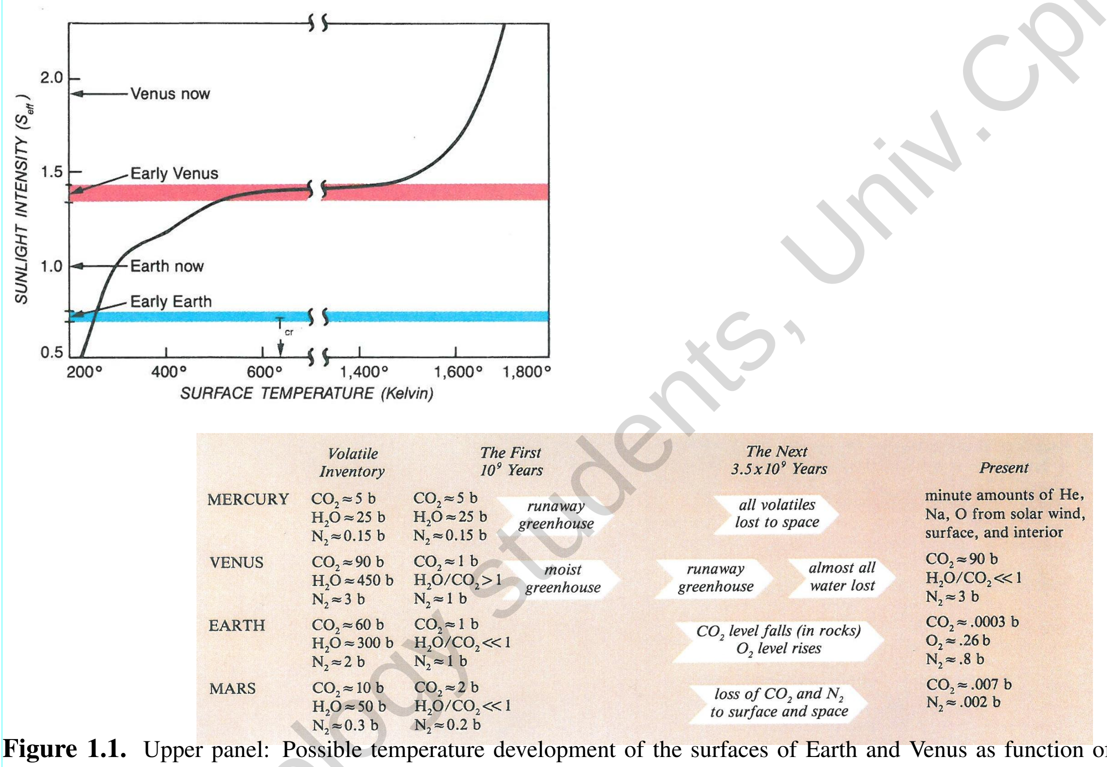

Figure 1.1. Upper panel: Possible temperature development of the surfaces of Earth and Venus as function of the increasing intensity of sun light from the formation of the planets and untill today. Lower panel: Possible development of the composition of the atmospheres of the terrestrial planets from the formation of their atmospheres until present time.

According to the climate model plotted in Fig. 1.1, Venus surface temperature at the very beginning was close to the critical temperature, $T_{\rm cr}$, i.e. slightly above 600 K, where all of water in the atmosphere is in the form of vapour. Such estimates depend critically on the assumed albedo and composition of the primordial atmosphere, and the problems are well illustrated by the lacking accuracy in predicting the correct present day temperature of Venus and the lack of ability to predict the almost constant temperature of the Earth due to the biological response to atmospheric temperature changes. The figure is from a well-renomated astronomy text book, and I reproduce it here only for the illustration, there is nothing better elsewhere, the problem is just simply difficult and there exist no well restricted model yet. The model in Fig. 1.1 assumed that Venus was identical in albedo, atmosphere and total water content 4.5 Gyr ago, but in spite of this shortcoming the predicted trend of Venus entering a run-away greenhouse state only shortly after its formation may be qualitatively correct, where the temperature increased with several hundred degrees in response to a modest increase in the solar influx over a relatively short time scale.

However, the 2 times higher solar input on Venus should have caused a factor of 2 higher organic molec-

	Venus	Mars	Titan	Earth with life	Earth without life
CO_2	96.5%	95%	0.01 ppm	0.03%	98%
N_2	3.5%	2.7%	82-99%	79%	1.9%
O_2	traces	0.13%	0.0%	21%	0.0%
Ar	0.007%	1.6%	$< 1 - 6\%$	0.9%	0.1%
CH_4	0.0%	0.0%	1-6%	1.7 ppm	0.0%
H_2O	0.01%	0.006%	0.0%	1%	
$T_{surface}$[°C]	456	-53	-179	13	240-340
$P_{surface}$[atm]	90	0.007	1.5	1	60

$$\Delta T_{GreenHouse} = T_{surface} - T_{eq} \qquad T_{\rm eq} = \sqrt[4]{\frac{(1-A)L_*}{16\pi r^2 \sigma}}$$

Figure 1.2. Measurements of the relative abundances of the most abundant gasses in the atmosphere of Venus, Mars, Titan and Earth (i.e., the only four solid bodies in the Solar system with a substantial atmosphere). Also listed is the measured surface temperature and pressure. Finally, the last column list an estimate of what the Earth's atmosphere would have looked like today if there had never been life on Earth. Under the table is repeated the definition of greenhouse warming and formula for calculating the equilibrium temperauture (i.e., the surface temperature the boddies would have if there was no atmosphere, the planetary albedo was A, the distance from the host star r (in astronomical units) and the host star luminosity was L_*.

ular production rate in the Venusian ocean than on Earth, according to equations we developed in a previous chapter. We know too little about the origin of life to say whether this in principle would have helped Venusians to appear quicker than Earthlings. However, the higher solar influx on Venus compared to the Earth, resulted in a higher evaporation rate from the oceans, and since water is the dominant greenhouse gas (also on Earth today) due to its very complex and pronounced infrared spectrum, then the surface temperature will have increased more steeply from its effective temperature as function of time on Venus than it will have done on the Earth. Model calculations indicate that H_2O was photo dissociated in the upper atmosphere of Venus and escaped to interplanetary space. In these models Venus soon became a very dry planet, surface temperatures rose to the present day value of 750°C, and the atmosphere developed high acidity and a surface pressure of 90 atm. Life as we know it from Earth seemingly had no chance of adjusting the atmospheric conditions, and hence the planetary surface temperature, fast enough to avoid the catastrophe – maybe. Whether life could have originated while Venus was still Earth-like is an open question, and whether it could in that case have managed to adapt to the rapidly changing conditions or escape to better places on the planet is another questions. Most scientists is of the clear opinion that it has not been possible.

Occasionally, attempts have been made to investigate whether e.g. the cloud layers in 50 km height (where temperature and pressure are as at Earth's surface) inhabits some kind of Venusian microbes. A short-lasted excitement arose in 2020 when scientists claimed to have found traces of biologically produced phosphine in the clouds of Venus. Soon, however, the general conclusion in the scientific literature reached a concensus that the claims were due to erroneous interpretation of the observations, and the general opinion today is that we have not so far seen any signs of life on Venus, neither past or present.

1.3 Mars

Mars' diameter is approximately half that of the Earth, so its surface is 4 times smaller and its volume and mass 10 times smaller. Theoretically we should therefore expect that if atmospheres and oceans are coursed by geological out-gassing, then the Martian ocean and atmosphere must have been considerably less massive than that of the Earth (more surface per volume). Mars would have been poorer in keeping volatiles from impacts, but the vicinity to the asteroid belt may have compensated for this in terms of higher volatile influx, if carbonaceous chondrites played an important role for the planetary oceans and/or atmospheres. From an

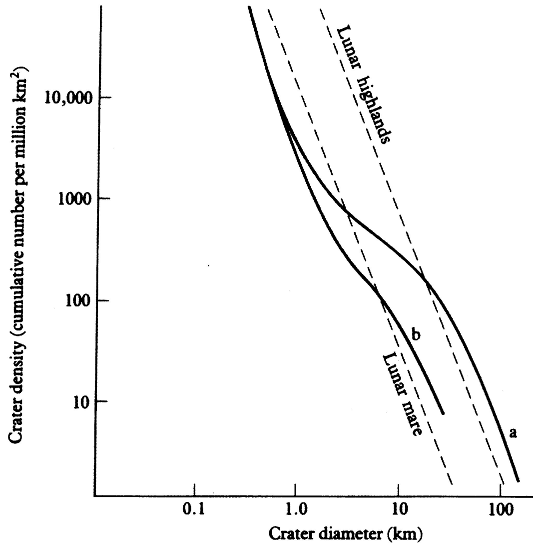

Figure 1.3. The crater size distribution of the martian uplands (a) and Chryse Basin (b), compared with the crater distribution in the lunar highlands and lunar mare regions.

observational point of view, the crater density in the oldest regions on Mars is comparable or a bit higher on Mars than on the Moon (see Fig.1.3). This may indicate that the volatile impact rate on Mars was larger than on Earth, and that would lead us to expect a dense early martian atmosphere and a substantial ocean, if they formed by impacts. Another empirical argument that Mars had a dense atmosphere in the beginning, comes from realizing that the lunar and martian crater densities are similar for large craters, while small craters are under-represented on Mars (also Fig.1.3). This could indicate that small projectiles burned up in the atmosphere and/or small craters were eroded away on Mars. This interpretation seems to be supported by the fact that the old craters in the southern highlands are weathered (i.e. not "sharp"). The younger craters in the northern plateau are sharp as the lunar craters, indicating that the atmosphere had gone at the time when these craters formed.

1.3.1 Mars' geological periods.

Fig.1.4 illustrates the main geological periods in Mars' history:
The Noachian (and pre-Noachian) periods stretches from the formation 4.6 Ga ago until somewhere between 3.8 and 3.5 Ga ago. The highland craters formed during this period, as did most of the valley systems. Since the valley systems are formed by huge amounts of water, water was abundant on Mars during at least periods of the Noachian epoch. Also the atmosphere must have been dense, judged from the weather erosion seen of the highland craters. Seemingly the temperature must have been warm enough for liquid water, and the composition of the atmosphere must therefore have been such that the greenhouse effect was strong, in particular considering the lower solar influx compared to today.

$$f_{\odot,\mathrm{Mars}} = \frac{1}{r^2_{\mathrm{Mars,AU}}} f_\odot \approx 0.5 f_\odot \;\;\Rightarrow\;\; T_{\mathrm{eff,Mars}} \approx \sqrt[4]{0.5}\, T_{\mathrm{eff},\oplus} = 0.8\, T_{\mathrm{eff},\oplus} \qquad (1.2)$$

The ending of the Noachian period coincides with the first evidences of life on Earth. Hence, Mars had physical conditions as on Earth long enough for the time it took life to arise on Earth, or even longer because Mars most likely cooled down to "biofriendly" conditions before Earth, because of its smaller size. It is therefore extremely important for our understanding of life, to know whether live evolved on Mars during the Noachian period. We may learn this by looking for fossils on the southern hemisphere of Mars. No space probes have yet been there because of the rugged terrain.

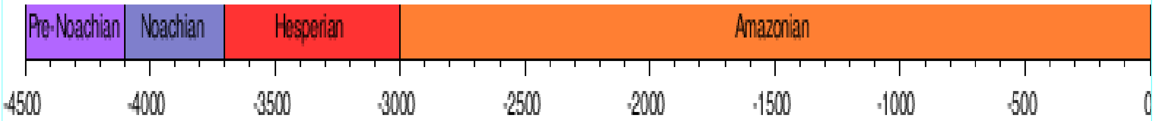

Figure 1.4. The three main geological periods on Mars: the Noachian, the Hesperian, and the Amazonian. The ages in Myr (million years) before now of the transitions between the periods are marked on the illustration, but are very uncertain.

The Hesperian period stretches from the end of the noachian period (between 3.8 and 3.5 Ga ago) until somewhere between 3.5 and 1.8 Gyr ago (billion years is most often written Gyr, but sometimes written Ga derived from the latin word annum for year; there is no difference in the meaning). Craters from this period are much less eroded than the Noachian craters, and the water and the atmosphere therefore most likely disappeared from Mars on the transition between Noachian and Hesperian times. A crucial question for understanding of the origin of water on Earth and Mars is therefore to analyze whether water existed only as a "flash" on the Noachian-Hesperian boundary (supporting a cometary origin 3.9 Gyr ago), or it stretched fart back into the Noachian period (i.e. an intrinsic-outgassing origin) and only gradually disappeared during the early Hesperian. Obviously also a more precise dating on the transition between the periods is crucial, since 3.8 and 3.5 Gyr makes a big difference concerning its relation to the LHB 3.9 to 3.8 Gyr ago. Finally, **the Amazonian period** stretches from the end of the Hesperian until today.

The heat transport from the interior to the surface must have been larger on Mars than on Earth in the beginning, because of the larger surface to volume ration of Mars. This together with the existence of water and a thick atmosphere, makes it likely that Mars has had some kind of continental movement in the beginning. However, soon the crust and lithosphere has grown relatively thick due to the quicker cooling, and the continents will have stopped moving. Water will have rained CO_2 out of the atmosphere, as on the Earth today. On present-day Earth (Fig. 1.14), the rain also erodes the silicate minerals on land and brings them to the ocean where they react with the CO_2 resolved in the ocean, and forms carbonate rocks. Subduction of the ocean floor brings the carbonaceous rocks into the hot interior, where the CO_2 is then released and re-emitted to the atmosphere through volcanoes. Once the continental drift stopped on Mars, the carbon cycle will also have stopped, and the weathering removal of CO_2 will then have resulted in a cooling. Depending on the speed of the cooling, the water in the atmosphere and on the surface will either have evaporated into space as on Venus, or it will have withdrawn as permafrost underneath the Martian surface. Most in-situ observations indicate that the latter was the case.

1.3.2 Channels and basins – signs of past water

Four different types of "channels" on Mars give us important knowledge about the history of water on Mars. The most prominent channels on Mars are actually not channels in the common use of the term. They are the equatorial canyons; most pronounced are the Valles Marines system. Although water seems to have flown in Valles Marines, it was a tectonic event (and not water) that formed the canyon system. They are huge cracks in the martian surface, later widened by erosion. Most likely the same tectonic uplifting that gave rise to the Tharsis bulge, where later the big volcanoes build up, also gave rise to the canyon system at the beginning of the Hesperian epoch.

The oldest type of channels is the *runoff channels*. They formed during Noachian time together with the southern highland craters. Water ran in these channels like in rivers on the Earth, and they are typically hundred meters wide and a few tens of km long. A larger and younger type of channels are the *outflow channels*. They are typically hundreds of km long and more than 10 km wide. It is not known where the water came from that ran in these channels. The flow rates must have been hundreds of times the flow in the Amazonas river (Earth's largest river), but water may have run only episodically. Often the start of these rivers are chaotic terrains, where collapse of the surface seems to have been associated with generation of a massive flood of water, such as the terrain in Fig.1.5. The course of the melting could have been subsurface

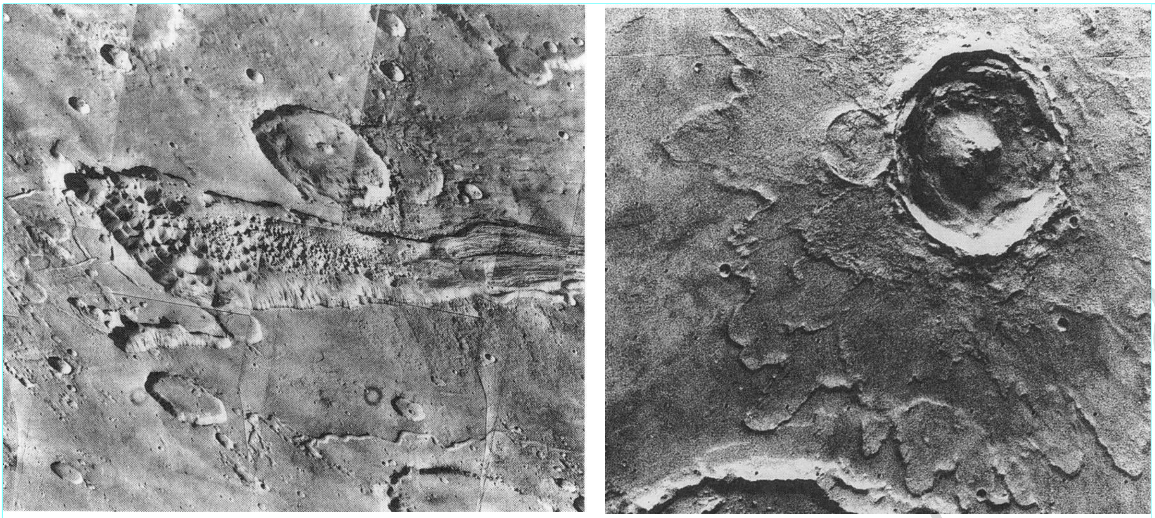

Figure 1.5. Chaotic terrains as the one in the 400 km wide area in the left panel, form the beginning of the outflow channels. Seemingly abruptly melted water in such regions formed the input to the enormous water flows through the outflow channels. The "mud-waves" around the crater in the right panel is another clear evidence of the existence of large amounts of past sub-surface water-ice.

magma movements or external impacts that delivered the water or chock melted existing sub-surface ice, or maybe completely other courses.

The fourth type of channels, called *gullies*, were only discovered recently, and they are indications that liquid sub-surface water still exist on Mars today. Detailed surveys of the Martian surface from the Global Surveyor cameras revealed stripes of moving material on the cliffs of larger channels and crater rims, seemingly carved by flowing water. Fig.1.6 shows gullies on a cliff near the south polar region and on the wall of the Newton Crater. Evidently small amounts of liquid water occasionally still flows on Mars today, but the detailed origin of the water that forms the gullies is unknown.

Several independent indications that substantial amounts of water was present on the surface of Mars in its earliest geological period, together with the fact that most of the northern hemisphere is lower than the southern hemisphere (compared to a fitted overall geosphere) and that the northern hemisphere is younger (in the sense that it has a lower crater density) than the southern hemisphere, might be evidence for a substantial early Martian ocean covering most of the northern hemisphere. While the gullies represent small amounts of localized water, the runoff channels represent larger amounts of water, possibly having run over longer periods of time. The largest amount of water, however, seems to have run through the outflow channels. The latter all ends along what seems a coastline toward the northern lowland plains, often after a fall of several km along the channel. This obviously has lead to speculations about the existence of an ur-ocean on Mars during the Noachian and early Hesperian period, covering most of the northern hemisphere. The Pathfinder mission landed in the delta of such a runoff channel, and did indeed find strong evidences for sedimentation in water. However, the ocean-interpretation is not without problems. Some measurements indicate that the coastline is not horizontal, as we would expect from an ocean, but may slope as much as 150 meters from one end to another. This may hint at a more viscous liquid, such as lava, or it could simply mean that the inclination of the coastline has shifted since the ancient time where the oceans existed and the coastline was horizontal. Another challenge is that we should expect carbonate rocks covering the bottom of the dried-out ocean floor, following the scenario outlined in Fig.1.14, which is not seen and which may indicate a flaw in our arguments or that the ocean was such a short event in Mars' history that noteworthy amounts of carbonate rocks didn't have time to accumulate. The clarification of the timescales of the existence of water on the Martian surface is paramount for our understanding of the possible biological evolution on the planet and in our solar system.

Although our knowledge is still quite limited about exactly how much ocean and atmosphere existed

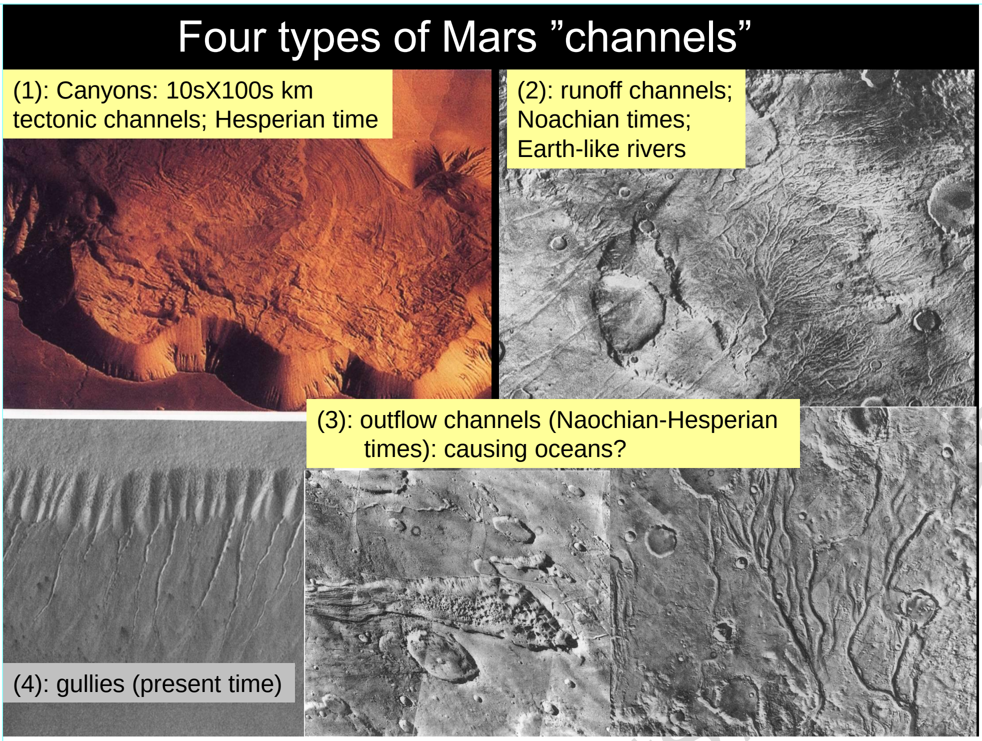

Figure 1.6. Runoff rivers and tectonic channels from various martian geological periods, including the two km wide image of gullies near the north polar cap, showing possible signs of present day running water.

on Mars in the Noachian and early Hesperian times, at least some numerical models show the Martian atmosphere and greenhouse effect to have been large enough in the beginning to make it take 1.5 Gyr before enough atmosphere had escaped Mars' low gravitational field to let the oceans disappear. In other words, the conditions on Mars' may have been similar to the conditions on the Earth long into the age where life had originated on Earth. Knowledge about the possible origin of life on Mars, therefore will teach us not only about Mars, but about life on our own planet as well. If life never arose on Mars, then it may mean that life does not automatically originate as soon as the right physical conditions are present. This would have profound consequences, in negative direction, for our expectations for finding life elsewhere in the Galaxy. If on the other hand we were able to find fossil life on Mars, but no living organisms, then it means that life was unable to keep the planet habitable over geological time scales (which would e.g. considerably weaken the theory that life is able to not only adapt to the physical surroundings, but also manipulate them under changing cosmic conditions on large scale in order to keep them fit for life).

1.3.3 Polar caps and sub-surface ice – present day water

While liquid surface water has been abundant only in the past (but still seems to appear in small amounts occasionally), ice is still abundant on Mars. The reason that liquid water cannot exist at the surface of Mars today is the low atmospheric pressure, which is always below water's triplepoint, as is seen from the phase diagram of water in Fig. 1.9. Hence, water will transform directly from solid to gas phase on Mars, as we are familiar with for CO_2 ("dry-ice") on Earth. No child (or grown up) will therefore be able to make a snowball of the snow that falls on the ground during the martian winter; it will be "dry".

It is well known that the polar caps are mixtures of water and CO_2 ices. During the summer months large quantities of CO_2 evaporates from the polar caps and drifts over the entire planet toward the opposite pole, where it condenses. This simple and regular CO_2 cycle potentially contain indispensable information about the role of CO_2 on climate systems in general, if well measured and investigated. The effect on the martian

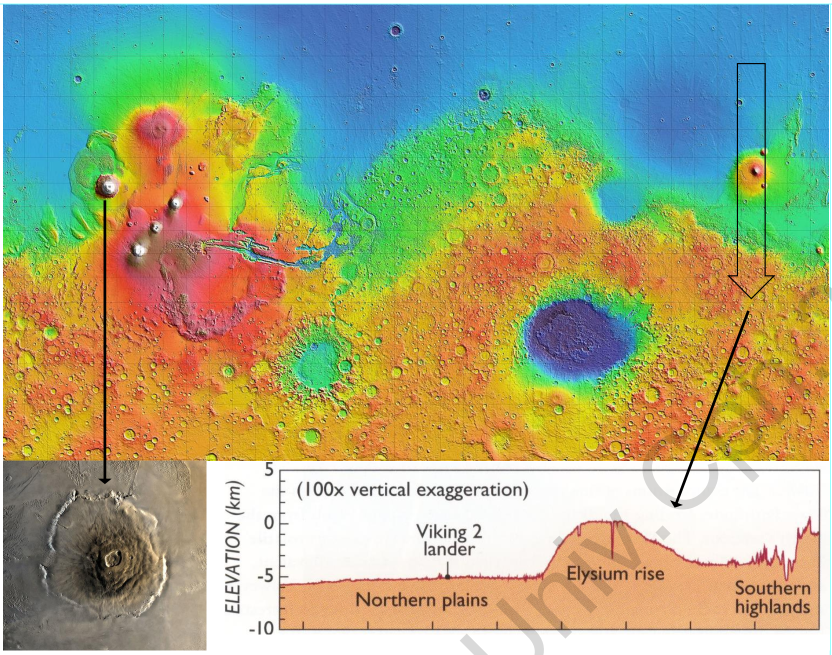

Figure 1.7. Topological map of Mars with the flat low-land northern plains (blue) and the on average 5 km higher southern crater-rich landscape. Some researchers take this as evidence for the existence of a northern Martian ocean during the Naochian epoch. Lower left inset shows the 25 km high Olympos Mons vulcano. Lower right inset shows a vertical cut from north to south.

polar caps are that they are large during the winter months, and shrinks to a residual size of 300 and 1000 km for the southern and the northern caps, respectively. The south polar cap stays at 150 K (which is the freezing point of CO_2) even at mid-summer, but probably it consist of a 2 km thick CO_2 layer overlaying a larger, permanent water-ice cap underneath. The even larger north polar cap rise to above 200 K during the northern summer months (in spite of the lower summer temperature in the north), and the water abundance in the atmosphere increases abruptly over the ice-cap during summer. The main course of the difference in the evaporation patterns is the different dust content (and hence albedo) of the two polar caps.

The largest amount of present-day ice on Mars is, however, not in the polar caps, but in permafrost. In 2002 the Mars Odyssey spacecraft was equipped with instruments that made it possible to make a global mapping of the subsurface water abundance down to a depth of 1 meter (actually of the hydrogen abundance, but it is generally believed that the hydrogen abundance is a measure of the water abundance). The result was that up to 35% (by mass) of the subsurface material down to 1 m is water-ice in the polar regions, and that the subsurface layers contain water-ice in the uppermost 1 m of soil as close to equator as $\pm 45°$. Due to strong isolation because of the thicker crust and lithosphere on Mars compared to the Earth, the temperature gradient downward from the surface rise steeper on Mars than on the Earth. Already at 2 km depth in the equatorial regions (and at 6 km depth in the polar regions) the temperature will have reached 0°C. If water-ice is still present at these depths, it will be in the form of liquid water in the rocks. On Earth, bacteria can thrive under similar conditions to these, and the deep subsurface layers may therefore be the most realistic place to expect martian life. If life arose in the martian oceans and rivers $\sim$4 billion years ago, it may have withdrawn with the water to the deep rocks km below the surface, as it maybe did on Earth during its few hundred million years long period of the ice-covered snowball-Earth phase half a billion years ago.

1.3.4 Searching for life – the Viking experiments

If we would be able to identify living organisms on Mars during coming missions, it will be crucial for our understanding of life, whether the basic structure of their cells is different from cells on Earth (for example

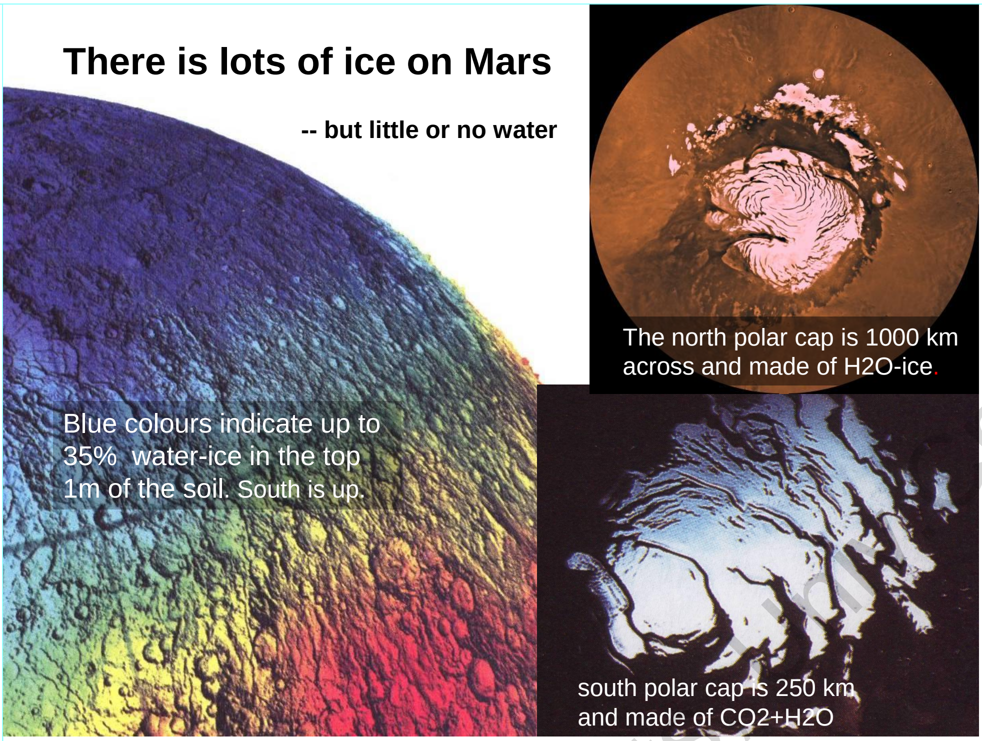

Figure 1.8. There is lots of water on Mars. Most of it is sub-surface, and quite an amount is in the polar caps too.

different kinds or different chirality of the assembled amino acids, a different kind of genetic structure, etc). In that case we could be sure the identified life had originated independent of life on Earth, and we would conclude that life on Mars arose on Mars, and that life on Earth arose on Earth (which would be a far from trivial conclusion in itself). If life on Mars was identical to life on Earth, we would be more in trouble with the interpretation, as we will discuss below, because several interpretations would be open.

Mars is the only place that has been the destination for space probes with the specific aim of looking for life. In 1976 the two Mars-probes Viking 1 and 2 landed on Mars equipped with transportable mini-laboratories that were able to analyze soil from the Martian surface for microbiology. A shovel on an arm on the space probes brought soil samples into three on-board chambers where test were performed on the soil. Samples were taken directly from the surface, as well as from underneath a rock, with the assumption that soil under rocks would have been shielded from biologically harmful solar wind and cosmic ray particles, and solar UV radiation. The three chambers were basically able to test whether some kind of metabolic process ("eating") took place, or whether a process resembling photo synthesis took place. In addition a gas-chromatograph could test whether large organic molecules were present in the soil.

In chamber one, *the gas exchange experiment*, a mixture of organic nutrients dissolved in water (soon termed "a chicken soup") was dripped onto the soil. If microorganisms digested the nutritions, one should expect a change in the gas content in the chamber, which was sought by letting the gas through a gas chromatograph. If the microorganisms resembled those on Earth, O_2, CO_2, and/or CH_4 would result from the digestion of the "chicken soup".

In chamber two, *the labeled release experiment*, the nutrients were labeled with radioactive ^{14}C. Any respiration of the martian organisms that released gases containing carbon (CO_2, CH_4, etc) would then be detected by a ^{14}C detector.

In chamber three, *the pyrolytic release experiment*, the carbon in the martian atmosphere was substituted with ^{14}C. If any photosynthesis-like process took place, the radioactive carbon would be incorporated in the organisms. After some time the air was flushed out of the chamber, the soil was heated to 750° and the

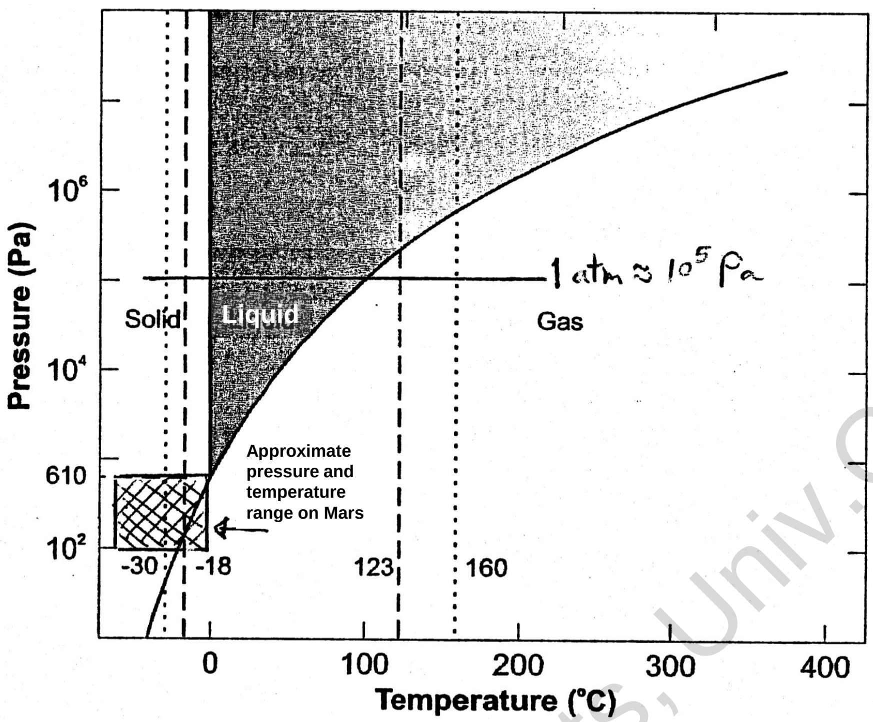

Figure 1.9. The phase diagram of pure water, showing also the approximate range of pressure and temperature found on the martian surface, as well as the temperature ranges on which life forms have been found (-18 to 123°C) or are expected to be able to survive (-30 to 160°C) on Earth.

resulting gases were analyzed for ^{14}C.

All three biology experiments initially gave positive results, consistent with the existence of living organisms. However, the following tests were contradicting the first results, and a general consensus arose that all the results could be explained by chemical reactions, including reactions with strongly oxidizing molecules (first suspected to be the well-known H_2O_2 that we are familiar with from any pharmacy, but only 30 years later finally identified by the Phoenix lander not to be a per-oxide but a per-chloride, ClO_4) and ions (such as O_2^-) produced by the solar UV radiation reaching directly to the martian surface. A fourth experiment on-board Viking was a gas-chromatograph mass-spectrometer (GCMS). It would be able to detect any large organic molecules in the soil, which would have to be there if biological organisms were present. Somewhat surprising, no organic molecules were detected at all to the detection limit of a few parts per billion (which is well below the organic molecular abundance of even Arctic soil; see Fig. 1.11). On this level even in-fall from carbonaceous chondrites on the martian surface should be detectable, and the interpretation was that the martian surface must be highly oxidizing, degrading even the non-biological large organic molecules from in-falling meteorites. Fig.1.11 show the results of test of the GCMS experiment on arctic soil and on samples of the carbonaceous chondrite Murchison. The final statement from the Viking biology team summarized that "the Viking results do not permit any final conclusion about the presence of life on Mars". Maybe the most important lesson we have learned from the Viking biological experiment is that there is a great need to clarify what life is before we can search for it in a meaningful way.

The martian soil analyzed by the two Viking probes was probably representative for the Martian surface in general, because frequent dust storms bring material around most of the globe, and meteoritic impacts mix the surface material substantially down to one or even a few meters depth. What was not sampled was material from a few meters or km depth, which may harbour liquid water at Earth-like temperatures, and therefore may be much more hospitable to Earth-like life than the seemingly very in-hospitable martian surface.

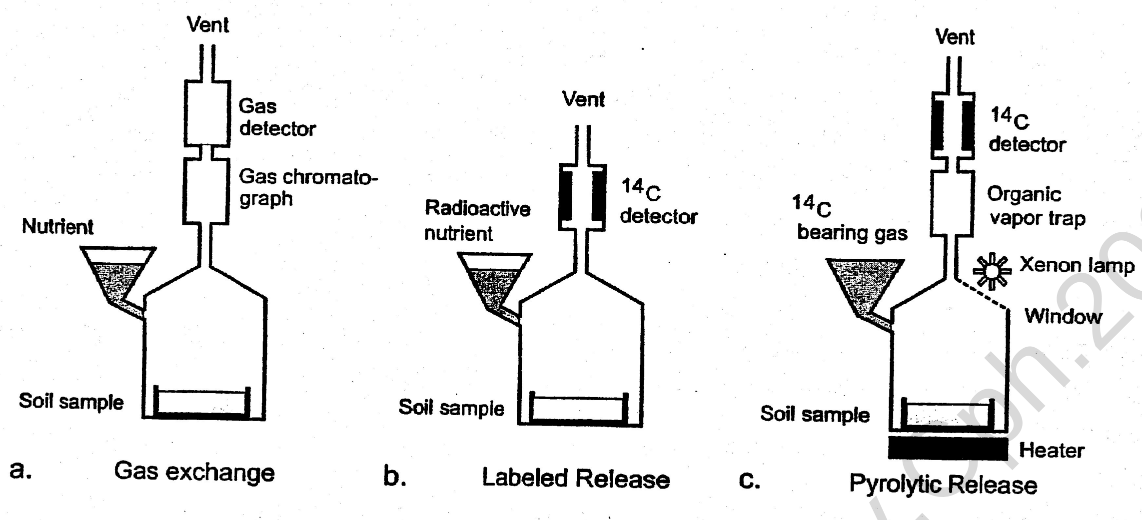

Figure 1.10. The different "metabolism" experiments on-board Viking to search for microbial life in the martian soil.

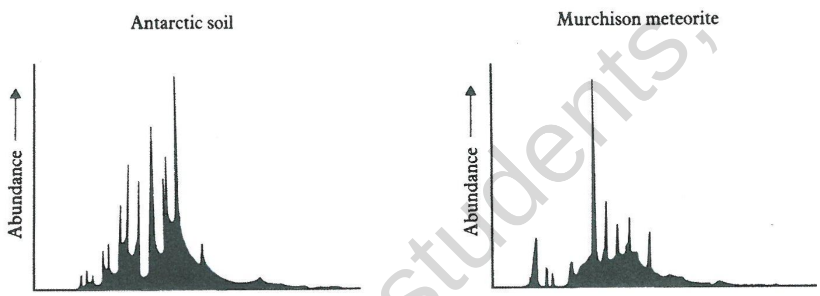

Figure 1.11. Test measurements, by use of the Viking GCMS, of the abundance of large molecules in samples of Antarctic soil and the carbonaceous chondrite meteorite Murchison. On this scale the results from the martian soil measured by Viking would be a straight line at zero.

Furthermore, the Viking experiments of course didn't say anything about possible fossil life – the question about whether Mars was populated in Hesperian or Naochian times. For this, we fortunately got unexpected input from a completely different side than the space probe experiments – a well awaited input that in 1996 should rejuvenate the interest in the dormant Mars exploration, and push a wish forward for promoting a manned space program and possible colonization of Mars in the foreseeable future.

1.3.5 Are we the Martians? – stones and bacteria from Mars

Sixteen million years ago some pieces of Mars was kicked off the surface during a low-angle (with the horizon) impact event. One of the pieces orbited the Sun in its own orbit until 13,000 years ago, where it was captured by the gravitational field of the Earth and felt on Antarctica. Here it was lying until 1984 where it was found on the Allan Hill (ALH) location of Antarctica where the wind and ice-flow patterns are such that meteorites falling at several locations of Antarctica over long time-spans are acummulating. The meteorite hunting expedition that found it judged that it was the most peculiar meteorite they had found during 1984, and therefore termed it ALH84001. At first it was misclassified and stored in a drawer at Harvard University, until it almost 10 years later was realized that this stone actually was a rare piece of Mars. Radioactive dating showed that it had solidified on the Martian surface 4.56 billion years ago. This number has been challenged in several later studies (now often listed as being 4.1 billion years old), but all agree at least on that it is considerably older than any of the other known Martian metaorites. Other measurements showed that ALH84001 had been exposed to cosmic radiation during 16 million years, until it again was shielded against cosmic radiation 13,000 years ago. From this we conclude that it was kicked of from Mars 16 million years ago, orbited the Sun until 13,000 years ago where it felt on Antarctica (and hence since then being protected against cosmic radiation by the Earth's atmosphere).

The high age is in itself exceptional, because it shows that Mars had a solid crust very early in its evolution. If life started in the way we have discussed in a previous chapter, then it can have originated on Mars several hundred million years earlier than on Earth. Kicked-off pieces of Mars may therefore have brought Mars-bacteria to Earth long before they had time to develop independently on Earth itself, and we may therefore ourselves be the true descendants of the first Martians. In that sense the first manned Mars mission will be "nothing other than the final return home after the longest staying away from home we know of".

But the measurement that gave rise to most controversies and debate was tiny inclusions in a crack in the stone. Dating showed that water had run through the crack 3.8 Ga ago. This is a vcery interesting dating in itself, because it states that liquid water was running on Mars allready 3.8 Gyr ago, during the time the sedimentation in water on Earth formed the piece of Earth's crust which is today known as the Isua Gerrnstone Belt, where the oldest potential traces of life on Earth is found. Obviously, Mars must therefore have had similar physical conditions as Earth 3.8 Gyr ago, i.e. an atmosphere and a temperature that allowed liquid water to exist on its surface – in other words, the habitable zone of our solar system included Mars 3.8 Gyr ago!

Detailed studies of the cracks in ALH84001, showed that the water had caused small inclusions to on the side, just like water can leave small chalk deposits on the shower cabin walls when the water has evaporates after a shower. The rims of the small inclusions were found to be strongly out of chemical equilibrium. Furthermore, this area showed small tube like structures, typically 0.1 mm long, which looked like bacterias on Earth. The authors of the paper that announced the discoveries, concluded that the most simple an coherent explanation of the deposits along the cracks, was to postulate that they were the result of bacterial activity on the surface of Mars 3.8 Gyr ago. Since then, many arguments for and against this interpretation has been given, but I think it is fair to say that at this stage there is a general agreement that the final conclusion about whether there has once been life on Mars will not be reached before we have been at Mars and been able to analyse a selection of representative samples both from the surface and from the deep subsurface layers.

Apart from the possible clues of fossil life in ALH84001 and apart from the important clues about the existence of a very early solid crust on Mars, ALH84001 gives us another very important information: It is possible for pieces of Mars to be kicked off from the martian surface, travel through interplanetary space, and land on Earth, without being melted or destroyed.

ALH84001 is only one of several groups of meteorites coming from the surface of Mars. Together we know them as Mars meteorites, or previously often called SNC meteorites, named after their three early identifications, Shergotti, Nakhla, and Chassignite. Most of them have small gas inclusions with the same relative composition as measured by Viking in the martian atmosphere, which is distinct from any other known meteorites and planetary composition. The SNC meteorites themselves were formed on Mars 170 Ma and 1.3 Ga ago, and kicked off much later. In particular, studies of the 7 known Nakhlites have shown that they were never heated above 100°C, and never exposed to high-pressure shock waves. One of the nakhlites had a very peculiar collision story that made it famous all over the world. A local newspaper reported that a piece had hit a dog during the impact, and the owner described how the dog fell dead to the spot. This is the only known reported dead casualty by a meteorite impact, but later investigations by other journalists seems to indicate that the story was made up.

However, even "just" the low degree of heating of the SNC meteorites can seem remarkably enough, and was eagerly debated. Today, it is concluded that it is in agreement with impact models that impacts can take place that are violent enough to accelerate surface rocks to high enough speed to bring them out of the martian gravitational field, without shock heating the rock even close to its melting point. Approximately 1 ton of mars meteorites hit the Earth annually, but as for all other meteorites, it is only a tiny fraction that is ever found. Most of these have orbited the sun for a very long period from when they are expelled from Mars untill they hit the Earth. One can estimate the time that passed between they were kicked off from the

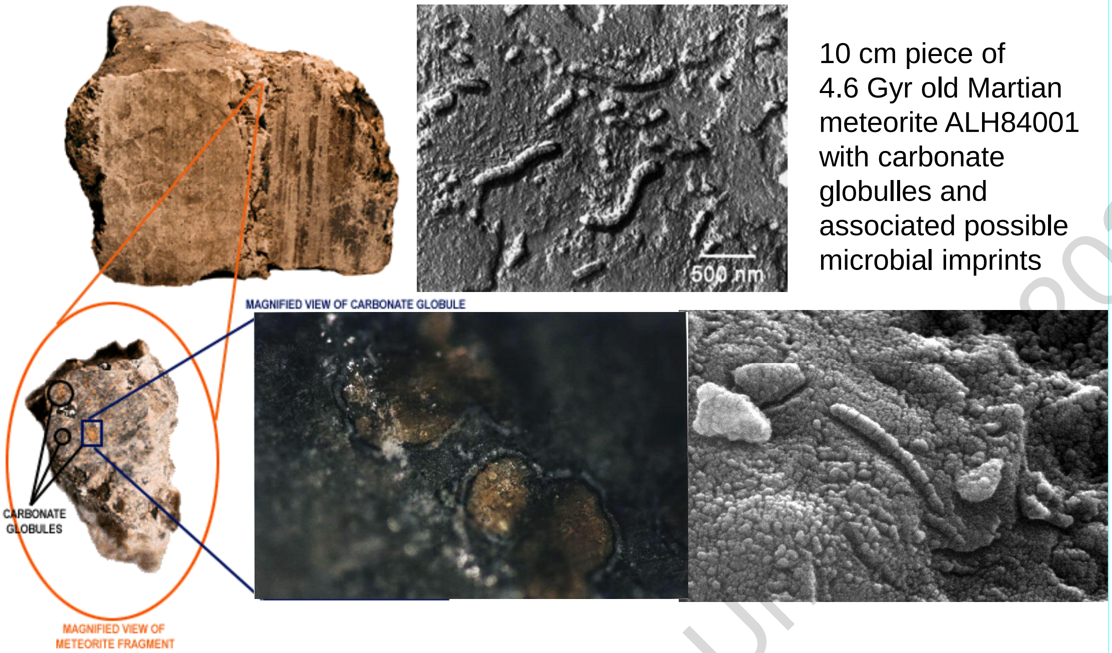

Figure 1.12. The discovery of a Martian meteorite, ALH84001, with tiny structures that can be interpreted as imprints of fossils after early Martian life forms.

surface rocks they were part of on Mars till they fell on the Earth, by measuring how much cosmic radiation they have been exposed to. The Nakhlite group was kicked off from Mars 11 million years before it hit the Earth, and ALH84001 was 16 million years in space on its journey to Earth. However, statistically one out of every 10 million pieces will hit the Earth less than 1 year after it is kicked off the surface of Mars, and on average 10 larger pieces of more than 100 g will fall on the Earth less than 3 years after they were expelled from Mars. This has dramatic consequences for the possibility of transporting living microbes from Mars to Earth (and in principle from Earth to Mars, but due to the higher gravitational field of the Earth, and the position of Mars in and orbit outside the Earth, the likelihood of meteorites traveling from Earth to Mars is on the percent level of the likelihood of travel from Mars to Earth).

Panspermia (which is Greek for “all seeds”) as a hypothesis about organic or biological material being transported through space, was introduced already 2500 years ago (by the Greek philosopher Anaxagoras), and it has been advocated by many leading scientists in modern time, too (e.g., Lord Kelvin, Svante Arhenius, Fred Hoyle, and Francis Crick). However, it has only been through the latest few decades that it has been possible to say anything quantitatively about it, and the surprises have been large. Experiments with sterilizing food products by radiation revealed already in the 1950s a bacteria now called Deinococcus radiodurans which not only survived the food sterilization, but now is known to thrive well even inside nuclear reactors. Another surprise was that several families of bacteria are quite resistant to ultraviolet radiation, previously believed to destroy all kinds of living organisms. Experiments in space has shown that just thin aluminum foil is enough shielding against UV radiation to keep most of certain bacterial spores vial for years in space. At the same time we have learned that not even can bacteria inhabit rocks, but probably they do it with such abundance that the sub-surface biomass on Earth may well be larger than that of all living organisms living on the surface of the Earth. Together all these pieces of knowledge tell us that if Mars once was inhabited with micro-organisms in a similar way as the Earth is, it is likely that pieces of rocks kicked off from Mars would not only fall on Earth regularly, but they would also bring with them living micro organisms (or dormant spores that could be brought to life in the right environment on Earth).

Experiments with the best suitable rocks for transportation of microbial life forms from Mars to Earth, indicate that relatively small dust grains as well as larger regular stones are viable. If they are very small, of course they offer no shielding to the inhabiting bacteria or bacterial spores, but already a few microns of rock is enough to shield from the solar UV radiation in space. More than a few microns of rock will course high-energetic protons to produce harmful showers of secondary radiation inside the shielding material. At the same time it has been shown that dust-sized meteoritic grains are decelerated slow during the entrance into the Earth's atmosphere, and will fall gently to the ground without being heated at all. It is generally believed that the largest mass of meteoritic material that annually falls on the Earth comes from such micrometeorites, totaling an annual infall of several tens of thousand tons. One source of microbial transport might therefore be micrometer sized dust.

Larger meteorites will heat considerably due to friction with the atmosphere during entry, but for pieces as large as the 100 g sized meteorites discussed above, the heath wave will penetrate only through the uppermost few millimeters, and the interior will stay unaffected during the few seconds short flight time through the atmosphere. Experiments have shown that several kinds of bacteria are un-harmed by the strong deceleration during the impact with the surface itself, and therefore it seems that a wide range of meteorite sizes will allow bacteria or bacterial spores of many kinds to be able to make not only the kick-off from the martian surface, but also the following interplanetary journey and the final penetration of the Earth's atmosphere and landing on the surface.

Today such invading bacteria would be so wildly outnumbered by existing local microbes that they would most likely be eaten long before they would have had any chance to adapt to the new and different environment which they would not at first be fit for, unless they are un-digestable. Already Darwin draw a similar conclusion to Pasteur's pasteurization experiments, by noticing that a consequence of the results must be that life can only have started on Earth once. In the very early period before life had arisen on Earth itself, martian microbes would have had much better conditions of time to adapt to Earth and survive without being eaten by locals. Since Mars may have had conditions suitable for life long before Earth had, we may therefore reach the somewhat astonishing conclusion that if life is such that it arise quickly as soon as the right physical conditions are present, then it is more likely that we are the descendants of the Martians than that we originated independently on Earth.

1.3.6 Could life have developed and existed on Mars until present times?

In 2004 the Mars Express spacecraft discovered local areas on Mars with an atmospheric concentration of CH_4 of up to 30 ppb. Consistently, observations from Earth show CH_4 concentrations in the Martian atmosphere of on average 10 ± 3 ppb. The photochemical lifetime of CH_4 under Martian atmospheric conditions is estimated to be 340 years or less. Consequently, something must be adding CH_4 to the atmosphere in order to keep it at this relatively high CH_4 concentration. The source can be geological as well as biological. If it is biological and caused by methane producing bacteria similar to those we know from the Earth, then it indicates that the microbial density in the Martin soil is $\approx 10^{10}$ times lower than on Earth. This is 10^5 times below the sensitivity of the Viking experiments, and therefore consistent with the results from Viking. Furthermore, the Mars Express results, as well as measurements from Earth, indicate that the activity is concentrated to a few specific subsurface areas, which there is no reason to believe the Viking mission should have happened to encounter.

A central question is, however, whether life (as a concept) can really exist in just small isolated areas without spreading globally, but further investigations of these spots are of course of highest scientific interest. Another clue that life may still exist on Mars today comes from some of the younger Mars meteorites. They show a photosynthetic $^{12}C/^{13}C$ ratio. On Earth this observation on graphite inclusions in 3.8 Ga old sedimentary rocks from Greenland, is the main argument for considering that life may have existed on Earth 3.8 Ga ago. Is the same observation in 170 Ma old rocks (meteorites) from Mars an argument that life existed on Mars as recent as only 170 Ma ago? Scientist do not agree on the answer to this question, just as it has been difficult to reach a general consensus about whether or not the non-equilibrium sedimentation in ALH84001 is a sign of ancient Martian biology or not.

If microbial life originated, or maybe even still exist, on Mars, a natural question is whether it ever developed into any higher life forms. Higher organisms like ourselves are much more sensitive to changes in the environments than are microbial life. On Earth small changes in the orbital and spin parameters cause

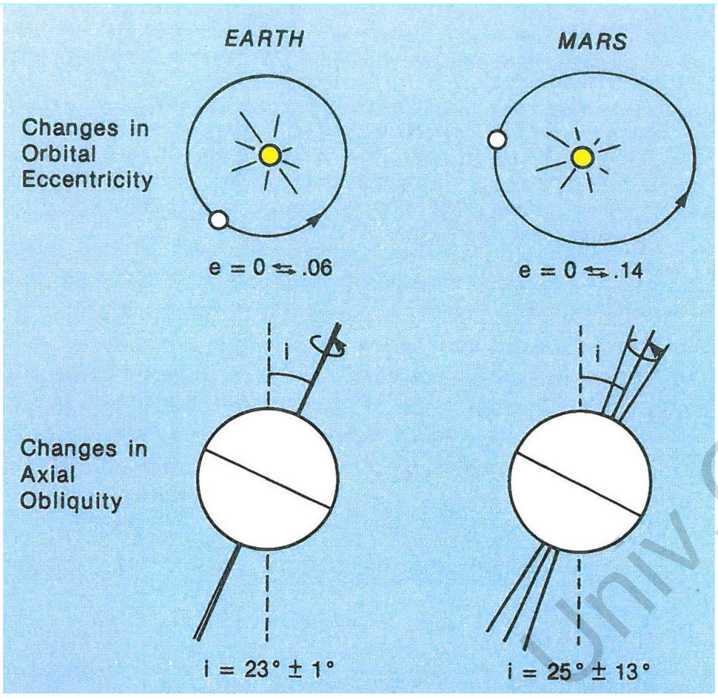

Figure 1.13. The vicinity to Jupiter makes Mars' orbit much more excentric than the Earth's, and the lack of a large moon makes the direction of the pointing of the axis less stable.

changes in the global average temperature of around 10°C. In the coldest end of such variations we have ice ages, and in the warmest end we have periods like now. These changes are caused by the overlap of several periodic and well predictable changes in the relative positions of the planets, giving rise to the so-called Milankovitch cycles. Apart from the Sun and Moon, the largest effect is caused by Jupiter. However, the torque imposed by the lunar orbit on Earth, demands a very high external perturbation of the Earth-Moon system in order to move the Earth's rotational axis, which will therefore always stay within $\pm 1°$ around the present inclination of 23.4°.

The changes between glacial and interglacial periods exert a strong pressure on higher lifeforms, and even as small changes as the 0.5 to 1°C change in average temperature that have been between now and the latest local temperature minimum around 1860, have given rise to considerable changes in the conditions for human societies, and there is a general worry about or civilization's capacity to adapt to possible temperature raises of 2 to 3 degrees during the coming century. Between 700 and 500 Myr ago, geological changes may have amplified the cyclic changes to a positive feed back that resulted in a complete freeze down of the entire Earth ("the snowball-Earth" discussed in a previous chapter). The finally end of this freeze-down period coincided in time with the Cambrian explosion – a complete change in the biology on Earth. Even the shorter and comperably modest climate changes (compared to the snow-ball-Earth freeze down) that took place 245 Ma and 65 Ma ago (possibly due to meteoritic impacts) resulted in extinction of around 90% or all larger organisms on Earth.

Due to the lack of a large moon, combined with the closeness to Jupiter, Mars will be exposed to much more violent and rapid changes in its rotational axis and other orbital parameters, and corresponding violent climate changes. Although the inclination of Mars' rotational axis today is almost identical to Earths', calculations show that it is moving much more than Earth, as shown in Fig. 1.13, and maybe even as much as between values close to zero (i.e., the rotational axis perpendicular to the orbital plane) and out to as much as 60° on timescales of only a few million years. Such marked orbital changes will not only change the surface temperature on the planet, but change the polar ice caps, the general atmospheric circulation pattern, and the existence of subsurface ice and possible water concentrations over large parts of the globe. Like in most other questions about life, we have not more empirical knowledge than the one example we know from Earth, but higher lifeforms on Earth would surely have very difficult in adapting to, or even surviving, such radical and rapid changes in the environmental conditions. It could be a major obstacle for sustaining long-term human colonization of Mars.

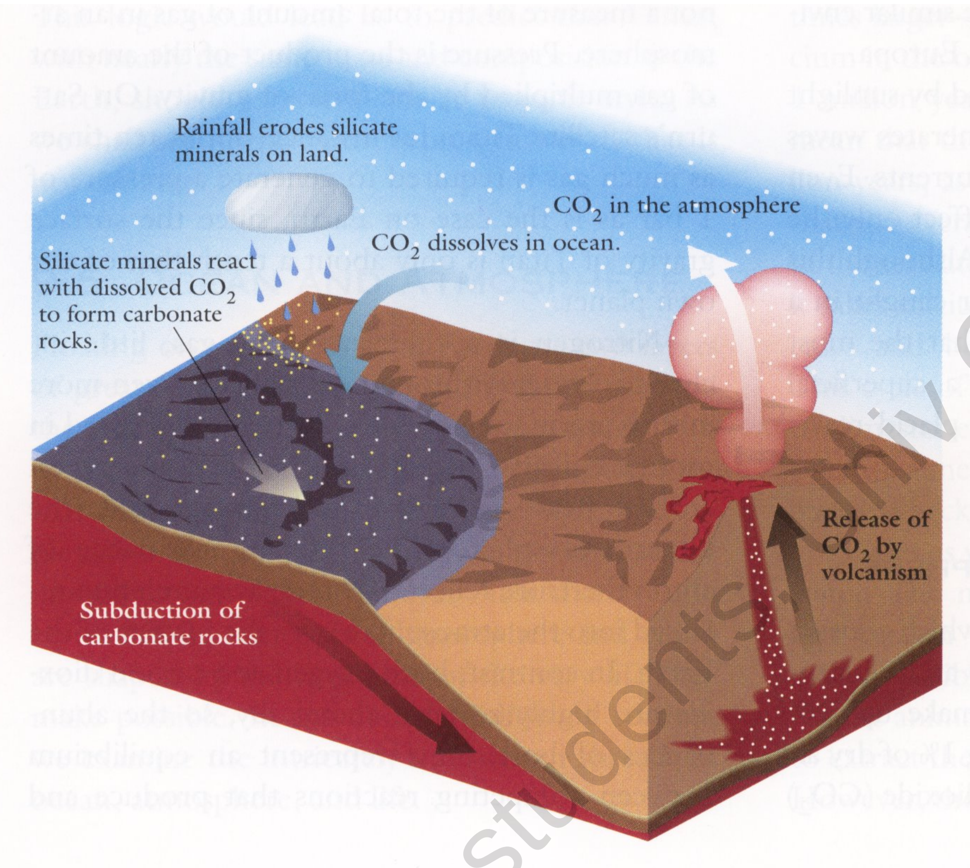

Figure 1.14. On Earth, CO_2 is washed out of the atmosphere and into the ocean by rain. In the ocean, resolved CO_2 reacts with eroded silicates from the continents, forming carbonate rocks sedimenting at the ocean floor. As the ocean floor slides in under the continental crust, the carbonate rocks are melted, and the CO_2 re-emitted into the atmosphere through volcanic eruptions. When this cyclic process stopped at Mars, CO_2 must have disappeared from the atmosphere and the surface cooled. But why are there seemingly no carbonate rocks on the present surface of Mars?

On top of this, Mars vicinity to the asteroid belt would make collisions with larger pieces of asteroids more frequent than on Earth. The crater densities of large craters on Mars are approximately a factor 2 higher than the corresponding crater densities on the Moon (Fig. 1.3), due to the Martian vicinity to the asteroid belt. It is estimated that even today impacts violent enough to expel material entirely from the martian gravitational field takes place with only a few million years interval. Approximately 1 ton of this material hit the Earth annually, as discussed in more detail above. Another difficulty for higher life forms on Mars, is the bombardment of solar UV radiation directly onto the surface (due to the thin atmosphere) and the exposure to solar wind and cosmic radiation particles due to the lack of a magnetic field. Both of these exposures are damaging to all lifeforms, but in particular to higher lifeforms. We will discuss in more detail later in this chapter how it will affect the possibilities for potentail human colonisation of Mars. The relatively early (compared to Earth) thinning of the atmosphere allowed a large amount of solar UV radiation to reach the surface of Mars already early in its evolution, and the rapid (compared to Earth) interior cooling gave rise to decreasing interior convection motion and corresponding early decreasing strength of the magnetic field, which gave access for cosmic radiation to the martian surface early in Mars geological evolution, too. Measurements of magnetization of the surface, shows that Mars once did have a strong magnetic field, but that it disappeared quite rapidly, and today is dominated by localized surface fields rather than a general planetary dipole field as on Earth.

All in all it seems we are forced to conclude that microbial life might have originated on Mars, but that the conditions for microbial life at, or close to, the surface today are difficult, and that higher lifeforms are unlikely to ever have had a chance to establish themselves anywhere on Mars. We have seen that much of this is not actually because Mars is too far from the Sun, but rather because it is too small (too low gravity to keep the atmosphere over biological timescales, too large surface-to-volume ratio to keep a continental drift, too quick cooling to keep a sufficient global magnetic field, etc). These are some of the reasons why we were putting a lower mass limit to the habitable zone in the corresponding figure in the chapter about exoplanets. It is not obvious at all where this mass-limit is, but everything indicate that Mars' mass (i.e. 1/10 $M_\oplus$) is too small to sustain life – at least higher life forms over evolutionary timescales.

1.4 The icy moons

1.4.1 Europa

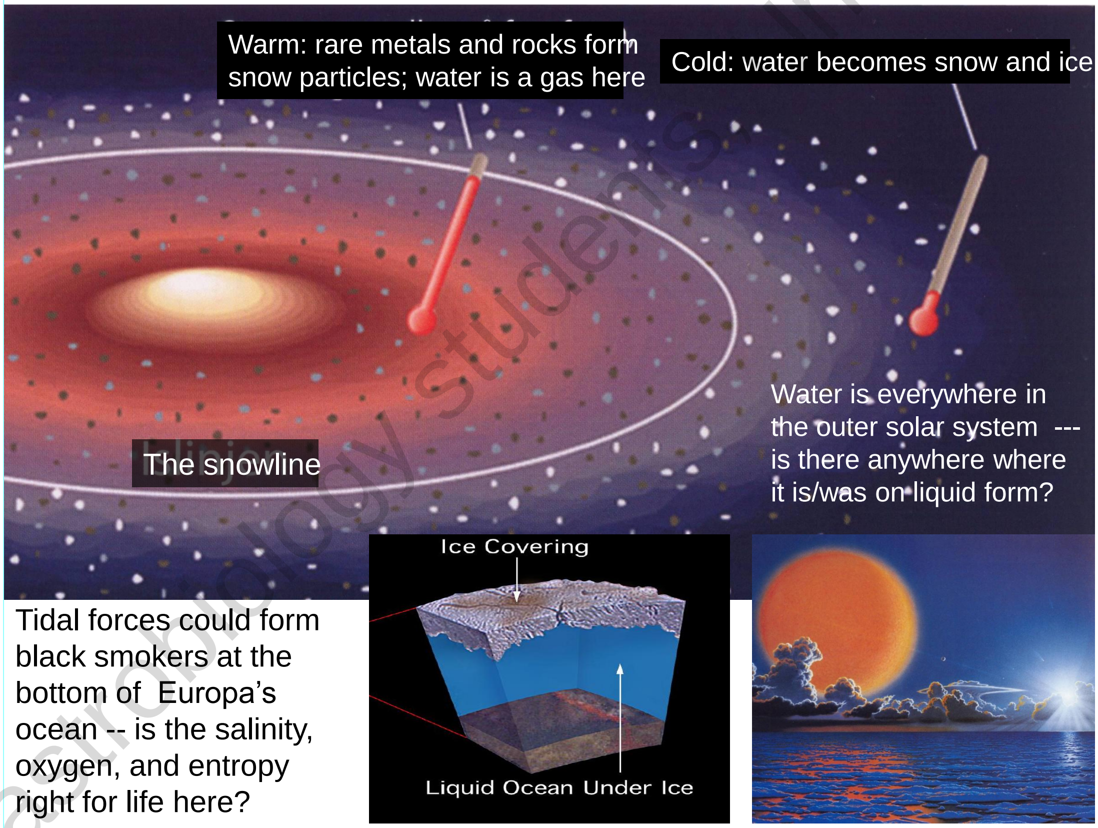

Figure 1.15. H_2O is one of the most abundant molecules in the Universe. In the proto solar nebula it condensed in the form of snow and ice in the outer parts of the disk (beyond "the snow line" approximately where Jupiter is today). Out there it was included in the solid bodies that became the comets, planets, and moons. In the inner parts of the nebula it was too hot for H_2O to condense, so it stayed as gas in the nebula, and the inner (terrestrial) planets and their moons became dry solid bodies of rocks and metals, initially without water. The solar system is therefore teamed with H_2O in the outer parts, but it is still an open question where and if it exist in the most valuable form for life: as liquid water. Some of the outer moons may be the most likely places to find liquid water, and Europa may house the largest ocean in the solar system, deep beneath its icy crust.

Europa is the second of the four large Galilean satellites orbiting Jupiter. The three inner of the Galilean satellites orbit in a mutual resonance pattern. Each 7.2 days the lager moon Ganymede orbits Jupiter once, lunar sized Europa orbits twice, while the innermost of the three, Io, orbits Jupiter four times. This resonance pattern of the three large moons force them into elliptic orbits which, combined with their spin-locked rotational periods caused by Jupiter, create huge tidal forces. The energy released by these forces is enough to make Io the most volcanic active object in the solar system, and enough to melt the deeper layers of the Europan ice-crust. With an average density of 3 g/cm^3, Europa is predominantly a rocky object. However, detailed mapping of the gravitational field from spacecraft flybys (mainly the many Galileo space probe orbits during the period 1995 to 2000) indicated that the outer 80 to 170 km of Europa is H_2O. The gravity data themselves cannot tell whether the H_2O is in the form of water or ice. With a surface temperature of 110 K near the equator and 50 K near the poles, obviously the surface is ice, but a liquid interior ocean may start as close as 10 km beneath the surface. At first glance the satellite photos of Europa's surface looks like Arctic pack ice on Earth. Patterns on separated pieces of surface clearly show that they have floated or been pushed away from one another, but comparison between the detailed Galileo photos and the 20 years older Voyager flyby photos has not revealed any movement of surface pieces, and also no movement of the general surface pattern compared to the orbit-synchronous rotation of the interior of Europa (to an accuracy of a relative rotation of the interior and surface of 1 rotation per 10,000 year) that one would expect if the ice-crust was only a thin layer. Instead of liquid water the cracks between the surface plates must consist of warmer, lighter and more soft ice which is very slowly being pushed up in-between colder surface ice plates, much like the ocean floor spread of the tectonic plates on the Earth (though no corresponding subduction zones have been identified on Europa). Numerical models suggest that the ice-crust is between 10 and 30 km thick, which would imply a liquid interior ocean of a depth somewhere between 50 and 160 km – maybe the largest ocean in the solar system.

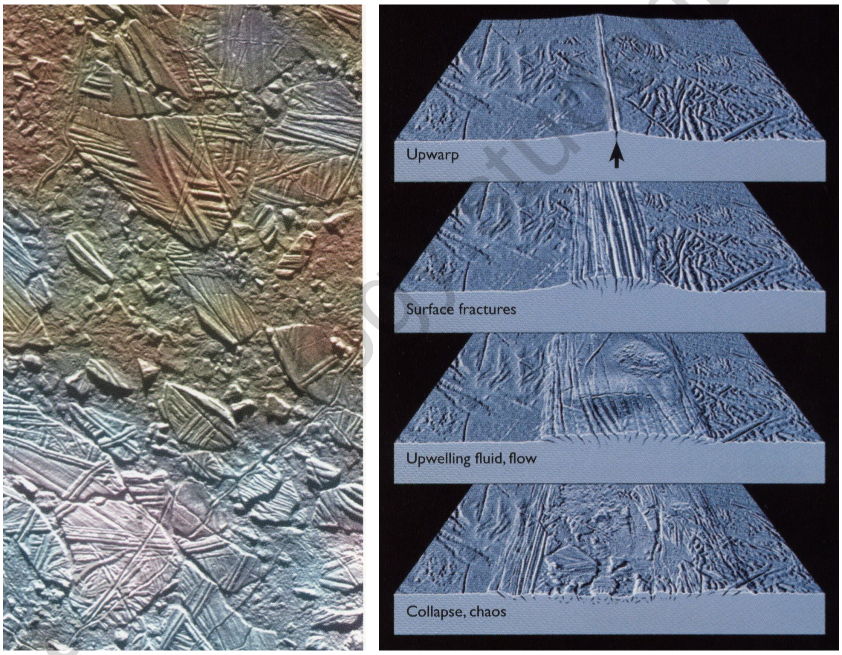

Figure 1.16. The "pack ice" on Europa's surface is formed when lighter (warmer and more soft) ice from deeper down is pressed up through the cold, dense surface-ice.

The colour of the material in the cracks is slightly dark-reddish, and spectroscopic measurements from the Galileo probe indicates that it might be the ice of salt-water, based on a solution of magnesium sulfate. Also measurements of changes in Jupiter's strong magnetic field around Europa, can be understood as caused by the salinity of a deep ocean.

The driving force of the cracks and dome-like uplifts is obviously the heath from the tidal force. How the tidal energy is exactly transformed to melting the ice is not known, but one possibility could be under-water volcanoes and black-smokers-like phenomena. On Earth, the energy from such events would be a driving force for colonies of living organisms, even at very deep and dark ocean spots. However, the life that thrive on such places on Earth did not arise there. It is dependent on oxygen, indicating that it arose on the surface and only later adopted to deep sea niches. It depends on the convective mixing of small amounts of oxygen from the atmosphere into the depths.

If the ocean on Europa had an ice-crust from the beginning, or from very shortly after the formation of Europa, then virtually no UV radiation would ever have reached the ocean, and no bond-breaking solar photons would ever have created even the most diluted primordial soup out of the Europan ocean, from which the first microscopic organisms could rise. However, the heat from the strong (red and infrared) radiation from Jupiter during its initial contraction phase, might have kept the surface of Europa liquid for an extended period. During this period the solar UV flux on the ocean will have been

$$f_{\odot,\mathrm{Europa}} \approx f_{\odot}/25 \tag{1.3}$$

With a depth d 50 times larger than in the oceans of the Earth, the density of organic material will eventually have been more than 1000 times smaller than on Earth, and if also the time before the ice-crust closed the access to the water was shorter than the time before ozone blocked the UV radiation of Earth, it will have lowered the concentration of the Europan primordial soup correspondingly according to the calculations in previous chapters. Furthermore, of course, the lack of an (substantial) atmosphere lowered the number of (CH_4 or CO_2) molecules available to break down by the UV radiation to produce free carbon atoms that could form complex organic molecules. In total it would seems that the odds for spontaneous arise of life on Europa must have been vanishingly small compared to Earth, in-spite of the existence of a substantial ocean of liquid water.

The only thing that immediately seems to could have changed this picture, is if even more complex organic molecules than those found in carbonaceous chondrites are present in comets, and comets played a dominant role for the origin of life compared to spontaneous generation in the primitive ocean. Cometary impacts must have been more abundant in the outer part of the solar systems where the comets are formed than in the terrestrial region, while collisions with carbonaceous chondrites will have been almost absent on Europa (because the source of all meteorites is the asteroid belt which is inside Jupiter's orbit). Model estimates show that even today cometary impacts that give rise to craters of 10 km diameter or larger will happen on Europa as often as once every 1.5 Myr (which with existing crater counts give an estimate of the age of the Europan surface of $\approx$ 30 Myr; considering uncertainties in the theory gives ages between 10 and 250 Myr).

The only way to get an answer to the question of possible life on Europa seems to be to prepare a robotic mission to Europa with a robotic submarine that could melt its way through the ice-crust and perform in situ studies of the subsurface ocean. Preparatory experiments are now being initiated for the subsurface Lake Vostok on Antarctica, which probably is the closest we come on Earth to Europan conditions. A major concern for the Lake Vostok investigation is that any such robotic mission is almost impossible to make completely sterile. On Europa, however, the radiation caused by Jupiter's strong magnetic field will most likely sterilize any space probe with destination to Europa, so that most Earth-borne passengers are likely to be killed before landing on Europa. If the first Europan robotic submarine would carry micro-organisms from Earth into the Europan sub-surface ocean, they could meet a huge amount of accumulated organic nutrient from cometary impacts over billion of years, if there has never been any Europan organisms to consume them. The likelihood of the microorganisms from Earth to thrive and multiply in the ocean is then not small. Could it, thus, be that our coming mission to Europa will be the first event to add the aspect of reproduction to Gaia that will eventually make us understand Gaia more as a living entity than we use to?

1.4.2 Enceladus

Several moons in the solar system have substantial resources of interior water, either in the form of ice, and possibly in the form of a liquid sub-surface ocean. Three moons have been seen to have active volcanic-like eruptions (Io, Triton, and Enceladus). However, Enceladus is the only one of the three to have both known surface eruptions and very likely also a liquid subsurface ocean, i.e. to have Geyser-like hot-water explosions. This makes Enceladus a particular interesting place in the search for possible extraterrestrial habitates in our solar system.

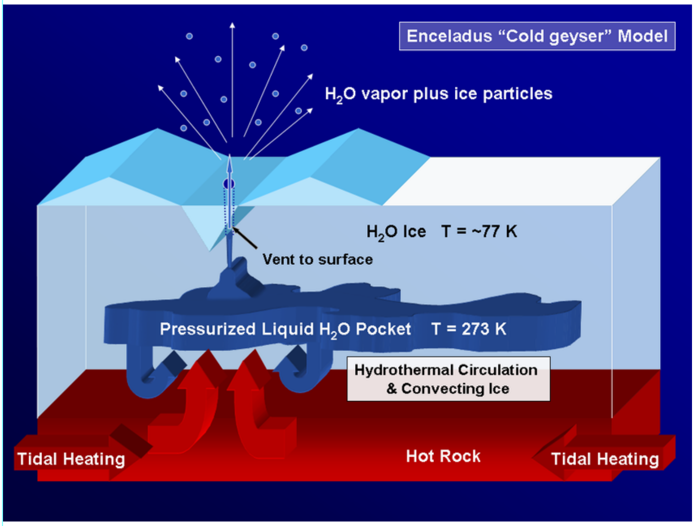

Figure 1.17. A model for the geysers in the Enceladus south polar region: Tidal forces heats the large rocky interior of Enceladus. The heath is focused toward pockets of liquid sub-surface water. When pressure due to the heating of the water builds up to become high enough, it creates a vent in the overlaying ice-crust somewhere along the "tiger-stripes", and a plume erupts, consisting of vapour and ice particles mixed with hydro-carbons (possibly from the sub-surface ocean). Some of the material will condense to solid particles and fall along the ridges of the tiger-stripes, while other material will be expelled into space and form the E-ring around Saturn.

The images from the Voyager passages in 1980 and 1981 clearly show that Enceladus orbits inside the densest region of the outermost Saturn ring known at that time (the E-ring), but it was not possible from the Voyager images to directly show that Enceladus was the source of the ring-material. This could only be established from sequences of more detailed studies from 2004 and onwards, where the Saturn-orbiting Cassini satellite revealed that large amounts of material were streaming out from the south-polar region of Enceladus, directly feeding the E-ring. The findings from the Cassini flybys showed that the erupting material was water, and that the water-plumes contained traces of hydrocarbon molecules. This urged for closer-up looks, and succeeding flybys between 2008 and 2010 were therefore planned to make the satellite pass as close as only 25 km above Enceladus' surface.

A Geyser-like phenomenon on an icy object is called a cryovolcano; it is literally an icy volcano. The ice-volcanic melt is called cryomagma. It is expelled as liquid or as vapour, which soon after the eruption will condense to solids at the typically very low surrounding temperature. Indirect evidences suggest that cryo-volcanism is or have been active on several other objects in the outer solar system, including the Kuiperbelt object Quaoar, Pluto's moon Charon, Uranus' moon Miranda, the Jupiter moons Europa and Ganymedes, and the Saturn moon Titan, but Enceladus is so far the only place outside the Earth where direct evidences for present-day water-eruptions have been confirmed. The energy for a cryo-volcanic eruptions can come from tidal heating (as on Enceladus), by subsurface greenhouse heating (as suspected on Triton), by radioactive heating or any other source of energy that could heat the subsurface layers sufficiently to form explosive eruptions of subsurface material. On Earth, geyser-eruptions are caused by contact between water and hot magma near the surface.

Enceladus is in a 1:2 mean motion resonance with the 10 times more massive moon Dione. This forces Enceladus into an elliptic orbit that together with its tidally locked synchronous rotation is the main source of tidal heat (just like the tidal heating of Io and other of the inner Jupiter-moons). The details about how the tidal heating could be transfered from the interior to boil the water is unknown. Fig. 1.17 shows one model of

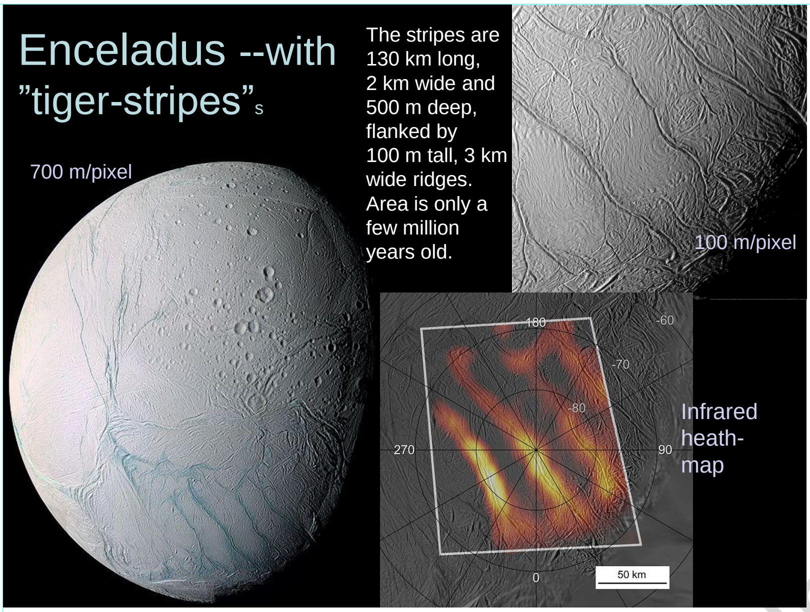

Figure 1.18. The volcanic eruptions occur at seemingly random places along the so-called ”tiger-stripes” in the south-polar region of Enceladus.

how it could take place. The density of Enceladus is relatively high compared to other outer moons, indicating that it must have a relatively large iron-rock core. The tidal heating is therefore believed to be dissipated in the rocky material, which somehow heats pockets of ice and water in the layers above. Numerical models indicate that tidal forces would have been able to keep the core temperature as high as 1000 K still today.

There are no obvious volcanic ”mountains” or other large structures visible at the registered places of the eruptions, but instead the eruptions seems to occur at random places along the so-called ”tiger-stripes” (Fig. 1.18) which are characteristic lines shaping the south-polar region. The stripes are typically 130 km long, 2 km wide, and 500 m deep, flanked by 100 m high, 3 km wide ridges. The low crater density in the ”tiger-area” indicate that the visible surface is only a few million years old in this region of Enceladus. Infrared images show elevated surface temperatures along all the 4 major ”tiger-stripes”.

1.4.3 Titan

Titan, the largest moon of Saturn, is perhaps the most intriguing astrobiological laboratory in our solar system. Due to the cold temperatures ($\approx$ 94 K) and lack of liquid water, it is not generally believed to be an environment conductive for stable Earth-like life. However, Titan does appear to have a complex carbon cycle involving methane clouds, photolytically produced hydrocarbons, and precipitation capable of eroding the hard, mostly water-ice surface. Its 1.5 bar atmosphere is dominated by nitrogen (1.4 bar) with CH_4 the second most abundant constituent. The production of such high surface pressure (1.5 bar) by a moon with a surface gravity of only 1.352 m/s^2 (0.14 g) requires a substantially more dense and larger atmosphere than the Earth’s atmosphere. At the distance of almost 1000 km from Titan’s surface Cassini was force to use thrusters to correct its path in order to compensate for the drag by Titan’s atmosphere. A human being would be able to fly by own force with attached bird-like wings in the dense, low-gravity environment of Titan’s atmosphere. A suggested future ESA-NASA mission to Titan is planned to bring a balloon with a payload of instruments in an attached gondola, which will be able to float in Titan’s atmosphere for years after its possible arrival around 2030.

Prior to the Cassini-Huygens mission to Titan in 2005 it was speculated that large dark areas seen in high-resolution infrared images from the ground, were methane oceans. The estimated (and now confirmed) surface temperature of 94 K ($-179°$C) is so close to the triple point of methane that one could expect methane to exist in either gaseous, liquid or solid form, or possible in all three forms side by side just like water does on Earth. Images from Cassini-Huygens showed that the dark areas weren't oceans, but rather large regions of dark material covered with windblown sand dunes up to 300 m high. The sand is most likely organic, tar-like solids that have snowed out of the atmosphere. Solar UV photolysis irreversibly destroys CH_4, producing an organic haze layer that may then precipitate hydrocarbons to the surface.

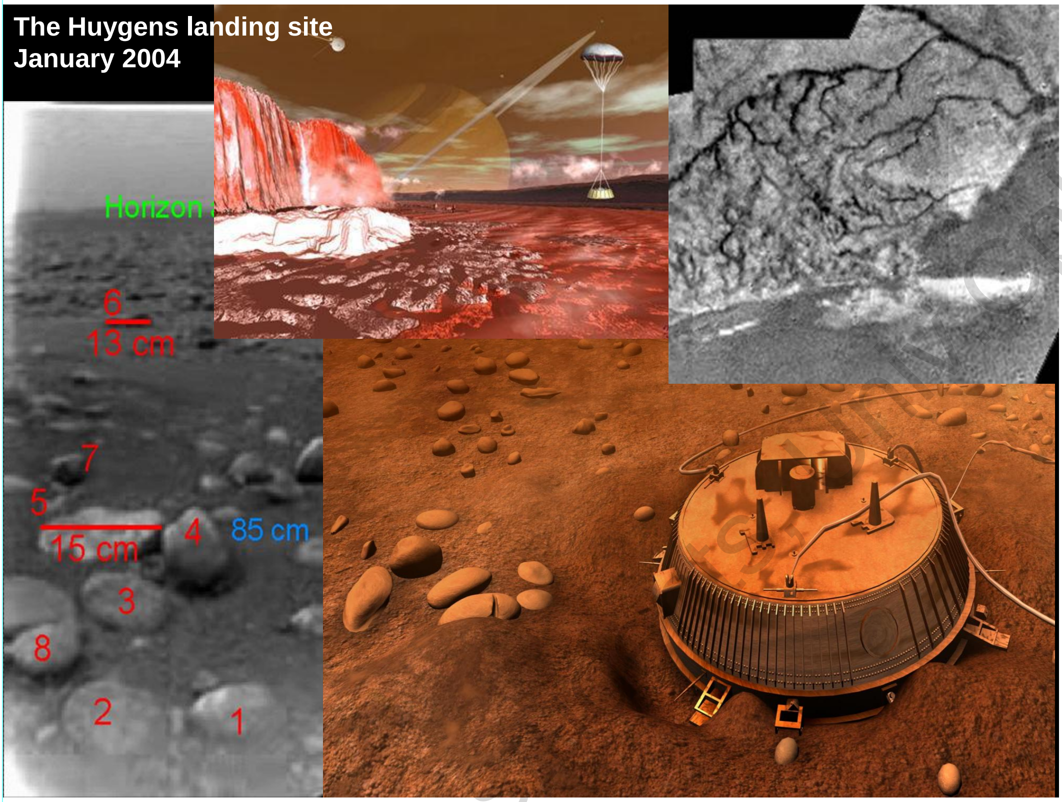

Figure 1.19. The surface features around the Huygens landing site. Upper right and left: The landing area photographed during the descend (right), and an artist impression of the existence of methane "waterfall", lake, clouds and iceberg next to the landing place. Lower panels: Photo of the "water-rock-iceblocks" in front of the lander, and artist visualization of the landscape seen on the photo.

Although Cassini-Huygens didn't find oceans of methane, it did identify a large complex of methane lakes, particularly in the northern hemisphere, and some of them are large enough to contain more "natural gas reservoirs" than 10 times all the Earth's known fossil hydrocarbon reservoirs combined. The persistence of relatively large CH_4 abundances in Titan's atmosphere cannot be explained as an equilibrium with the surface reservoirs of liquid methane. On Earth the large (although modest compared with Titan's abundances) amount of free methane in the atmosphere is attributed to a constant biological contribution. Although this possibility cannot a priory be ruled out on Titan, a more likely explanation for the pumping of free methane into Titan's atmosphere is cryovolcanic activity (just like on Enceladus, but here expelling methane). In order for this to be the explanation for the excess (and out of equilibrium) amount of atmospheric methane on Titan, huge resources of liquid methane must exist as sub-surface methane oceans that can feed the cryovolcanoes.

Results from the Huygens probe imager revealed dramatic dendritic networks, which supports the idea of a possible methane-based meteorological cycle that may have eroded them. Upon landing, the Huygens probe returned images of what appear to be water-ice rocks, and data from the surface penetration instrument may indicate a thin crust covering a material similar to wet clay. Analysis from the gas chromatograph/mass spectrometer confirm that CH_4 is, indeed, one of the major components in this material. The soil seen in between the water-ice-rocks in the Huygens image from the surface of Titan, is believed to be organic compounds precipitated from the atmosphere. All in all, the Cassini-Huygens probes delivered strong support

Composition of Titan's Atmosphere

Major constituent		*Percent*
Nitrogen	(N_2)	82 – 99
Methane	(CH_4)	1 – 6
(Argon?)	(Ar)	<1 – 6
Minor constituent		***Parts per million***
Hydrogen	(H_2)	2,000
Hydrocarbons		
Ethane	(C_2H_6)	20
Acetylene	(C_2H_2)	4
Ethylene	(C_2H_4)	1
Propane	(C_3H_8)	1
Methylacetylene	(C_3H_4)	0.03
Diacetylene	(C_4H_2)	0.02
Nitrogen compounds		
Hydrogen cyanide	(HCN)	1
Cyanogen	(C_2N_2)	0.02
Cyanoacetylene	(HC_3N)	0.03
Acetonitrile	(CH_3CN)	0.003
Dicyanoacetylene	(C_4N_2)	condensed
Oxygen compounds		
Carbon monoxide	(CO)	50
Carbon dioxide	(CO_2)	0.01

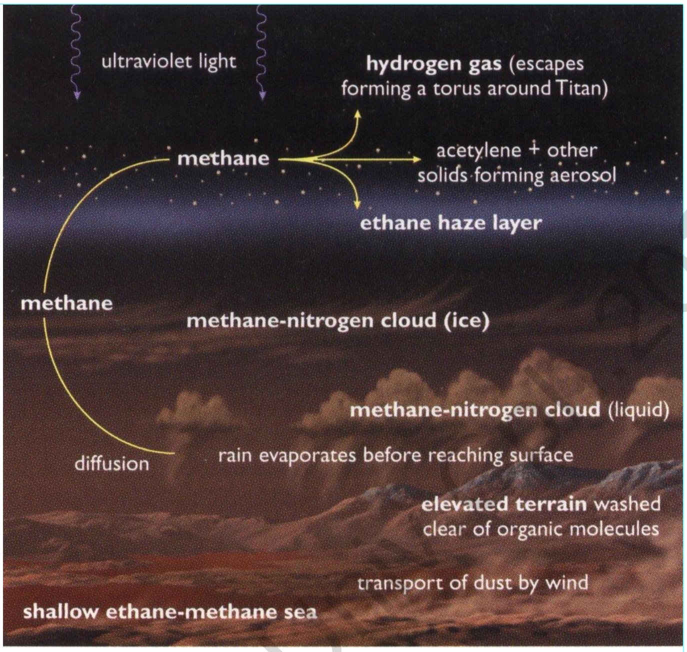

Figure 1.20. The chemical composition and possible weather dynamics of Titan's atmosphere.

for believing that Titan has a weathering system analogue to our meteorology, where methane plays a role analogue to the role of water on Earth.

The atmosphere has only one Hadley cell, due to the slow (tidally locked) rotation of Titan, circulating hot air from the summer hemisphere through the stratosphere, causing methane rain or snow to fall on the winter hemisphere. Cold air is then circulated from the winter pole back to the summer pole at lower atmospheric heights. Cassini visited Saturn-Titan during the 12 year long southern summer period (lasting until 2010), and considerably more (methane) lakes were observed in the northern (winter) hemisphere than in the southern hemisphere.

Titan must be loaded with complex organic molecules at its surface. The molecules are created by solar UV radiation and energetic particles catalyzing reaction chains between methane and nitrogen and its reactants. Laboratory simulations of the atmospheric conditions at Titan has lead to a number of reactions of methane and nitrogen via acetylene (C_2H_2), HCN, and corresponding ions to form rings of benzene and other complex organic molecules of up to hundreds of atomic mass units large, commonly referred to as tholins. Tholin is described, by the scientists that produced it, as resembling a mixture of small plastic dust and rancid butter. It is speculated that tholin particles are responsible for the orange colour of the upper atmosphere of Titan, and it may rain down through the atmosphere to finally form a gooey, sticky layer covering part of Titan's surface.

If the Earth's early atmosphere was rich in methane (as assumed in several laboratory experiments, including the famous Urey-Miller experiment) similar material will have formed from the solar UV radiation interacting with atmospheric methane and nitrogen on early Earth. This make many scientists see Titan as a laboratory for early Earth-like life, or at least for the creation of the early building blocks for life's origin on Earth. On Earth, however, the produced "glue" will have been floating in water, which we believe to have been crucial for the formation and development of the first organic cells. During cometary impacts on Titan, lakes of water will form at the surface, and calculations show that they will stay liquid for centuries after the impacts. The extreme low surface temperature will soon cover them with an isolating ice layer, which may even assure the existence of near-surface, sub-surface liquid water lakes for millenia. Since the surface will have been loaded with extremely high concentrations of organics build up in between the cometary impacts, some theories speculate that life would form so rapidly under these conditions that Earth-like water-based

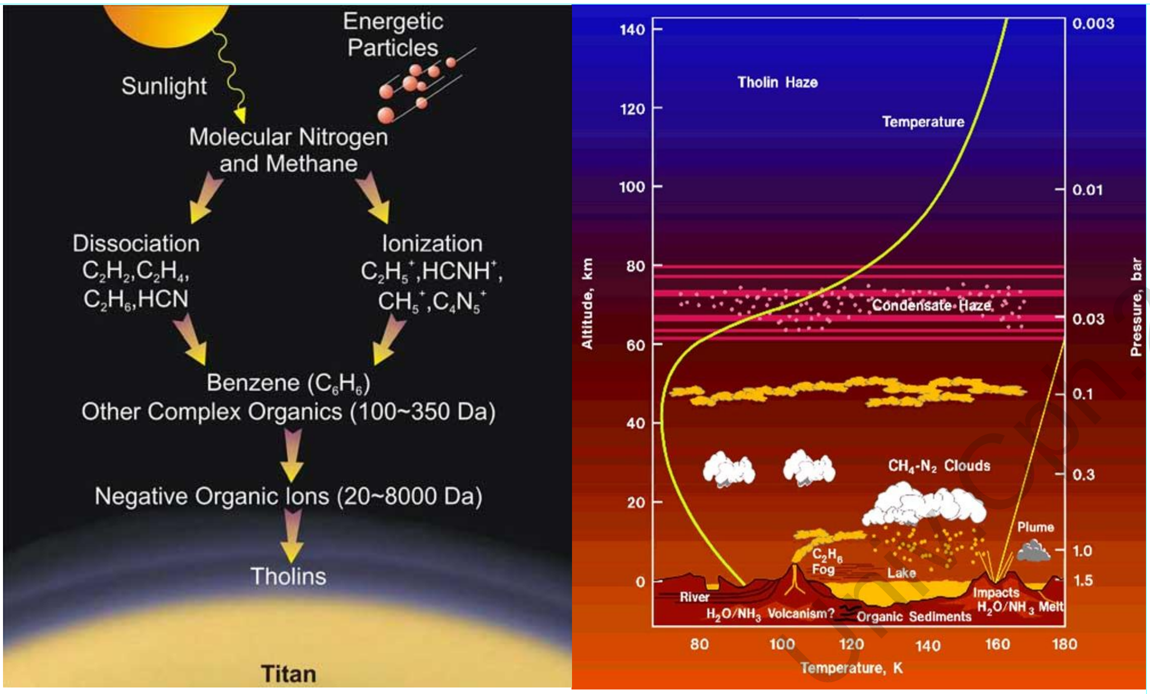

Figure 1.21. The formation of tholins in Titan's atmosphere.

life could arise spontaneously each time a comet hit the titanic surface.

Just like on Venus, the atmosphere has a strong global wind speed which rotates the whole atmosphere with high speed around the solid planet. Also the solid crust seems to rotate relative the interior, which is seen as an indication of the existence of subsurface liquid (probably water) on a global scale. Oceans of water cannot exist permanently at small sub-surface depths at the low temperatures of Titan. Instead it is speculated that a solution of water and ammonia (a "cooling liquid" which freezes considerably below 0°C) forms a large sub-surface ocean at a depth of no less than 200 km below Titan's surface. Prior to Cassini it was generally expected that all possible surface features would relate to impacts, but the discovery of such features as a 130 km long, 1.5 km high and 30 km broad mountain chain in the southern hemisphere indicate otherwise. It seems to consist of water-ice pressed up through tectonic activity to form an ice-mountain chain, covered in a thick layer of methane snow.

If all these preliminary results from the Cassini mission, will be confirmed by future studies, Titan is in many ways the most Earth-like body in our solar system, with tectonic activity, volcanoes, oceans of (sub-surface) water, a global wind system, and an ongoing meteorological recycling system of gas, liquid, and solid (methane) between the atmosphere and an active surface crust.

1.5 Alternative life forms?

Life as we know it on Earth has water as its main ingredient, and we are therefore often putting equality sign between the search for liquid water outside Earth and the search for extraterrestrial life. The biological cell in any living organism on Earth requires water inside each cell and water outside each cell, as explained in more detail in a previous chapter. Without liquid water, life as we know it on Earth cannot exist. In a first attempt to identify exoplanets potentially suitable for the existence of life, we therefore defined the habitable zone as the distance from the star where an Earth-like planet would have a surface temperature that would allow the existence of liquid water. This definition in itself is not very straightforward to use on existing data of exoplanets, since we generally only know the mass (or size) and the orbit of the planet, and not whether it is Earth-like. For an exoplanet to have the surface temperature the Earth would have in its place, would require that the exoplanet had the same albedo, the same atmospheric greenhouse heating, and approximately the same rotation period as the Earth. This would probably almost never be the case, and in particular the habitable zone would around most stars be so close to the star that planets there would be tidally locked (i.e. always face the same side toward the star). And even for exoplanets around solar-like stars, identical to Earth in mass, atmospheric composition, albedo and rotation period, we could far from be sure that the right conditions were present for the arise and development of life. For example, we would in general not know whether an "Earth-like" exoplanet had water at its surface, since (as we discussed in detail in a previous chapter) it is far from trivial to figure out how the water came even to the Earth itself, and basically unknown whether water is the rule or the exception on Earth-like exoplanets. One would therefore naturally ask two questions: Could life be based on something other than water as the solvent of its cells? Could we look for life in a more general way than looking for water?

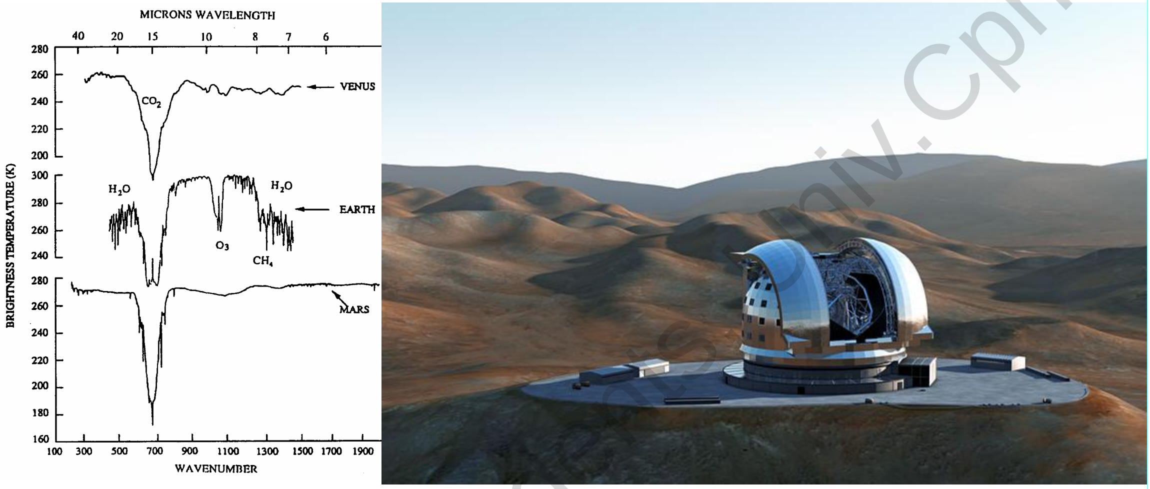

Figure 1.22. The giant E-ELT telescope will be able to obtain spectra of Earth-like exoplanets orbiting nearby solar-like stars, and detect whether the atmosphere is in chemical or biological equilibrium.

In the now canceled space missions Darwin and TPF, as well as the next-generation giant ground based telescopes, one is trying to look for life without looking for water or any other pre-specified chemistry. If one is able to identify planets those atmosphere on large scale is out of chemical equilibrium, one would conclude that it could well be due to the existence of living organisms. We already discussed in a previous chapter that the near infrared spectrum of the Earth's atmosphere, show large scale signs of the co-existence of carbon-dioxide, methane, oxygen, and water (CO_2, CH_4, O_3, and H_2O). These 4 molecules are out of chemical equilibrium with one another an can only exist if something is constantly pumping these molecules into the atmosphere in large amounts. In order to get the spectrum in Fig. 1.22 (and a static abundance of 1.7 ppm CH_4 in the atmosphere), something will need to continuously pump 10^9 tons of CH_4 into the atmosphere per year. Without life's persisting pumping of CH_4 and O_2 into the atmosphere, we would quickly end up with a chemical equilibrium in the atmosphere, and between the atmosphere and the surface, leaving CO_2 as the only major component of the atmosphere, as is the case on Venus and Mars. Since the existence of in particular oxygen and methane in the atmosphere is due to the existence of life (planets and bacteria) on Earth, we would say that the Earth's atmosphere is in biological equilibrium, meaning that the atmospheric chemistry, the biological respiration process, and the surface reactions on Earth, are in dynamic equilibrium.

There is a simple correspondence between the theoretically obtainable resolution in an image and the size of the mirror of the telescope collecting the light for the image, which is given by

$$\Theta" = \lambda(\mu)/4D(m) \tag{1.4}$$

where Θ" is the diffraction limit, i.e. the minimum distance (in arc-seconds) between two point sources of light, of equal intensity, that is resolvable with a telescope that has a mirror of diameter $D(m)$ in meters, observing at wavelengths $\lambda(\mu)$ in micro meters. If the two light sources are of unequal brightness, it requires a larger separation, and/or a larger mirror. A mirror of approximately 40 m in diameter is the minimum size mirror that will be able to separate the light from a Solar-type star from the light of an Earth-like exoplanet orbiting in an Earth-like orbit around some of the nearest stars. This size telescope has therefore been a science-driver for the construction of the European ELT telescope (Fig. 1.22) which will most likely have its first light as the world's largest telescope around 2024. If the first spectra show an Earth-like atmospheric composition, we can be reasonable sure that photosynthesizing organisms inhabit the planet together with methane producing micro-organisms and living creatures that has evolved to take advantage of the most efficient metabolism (the oxygen respiration) that we know of. It is not unlikely we will see such a spectrum. One the other hand, if we keep seeing only CO_2 in the spectra when we manage to identify and observe more and more Earth-like exoplanets, we have all reasons to feel more and more lonely (or privileged?), in terms of life in the universe.

Alternatively, and very likely, we may of course find something that is neither Earth-like nor Venus/Mars-like. But the observational set-up is still able to identify whether it is biology or chemistry we see in action. Any composition that is out of chemical equilibrium will show us that low entropy is building up on the planet, and in our attempt to identify what we should mean by life as a concept, we concluded in a previous chapter that local decrease of the entropy is a necessary and central ingredient.

In order to speculate what we might expect, if we should identify signs of non-Earth-like life in the spectra of Earth-like exoplanets, Titan is maybe the most inspiring environment. Titan's atmosphere is dominated by N_2 (as Earth's atmosphere), methane, and other hydrocarbon compounds, and the surface has lakes of methane and a weather with rain and snow of methane, tholin and possible other hydrocarbon compounds. Tholin is not resolvable in liquid methane, but it is resolvable in water. Water exist in large amounts on Titan, because Titan formed far outside the snow-line, but because of its low surface temperature, water on the surface exist only in the form of ice as hard as steel. Much of what looks like rocks on the images from the Huygens space probe (Fig. 1.19) from the surface of Titan, is in fact clumps of water-ice. But planets and moons in the outer solar system, including Titan, are regularly hit by impacting comets. A comet hitting Titans surface would immediately melt, and possible frozen water-ice in the surface region would melt too. If the impacting speed is low enough, most of the melted comet (i.e. liquid water) would form a lake of fresh water from the comet and possible water from the surface of Titan too. Within minutes the surface of the lake would start freezing, and the icecover would allow the lake to exist on the surface for millenniums. Tholin is known to spontaneously form amino acids when resolved in water (or in ammonia), and with the extremely high concentrations of organic material at the surface of Titan (and in comets), one could speculate whether life not too different from life on Earth would arise spontaneously each time a comet impacted Titan. If the comet hit a region that had preciously been melted, one could even speculate whether dormant micro-organisms from the previous period of liquid water could come to life again, and maybe evolve a step further for the next short period of life-given conditions in the lake.

In the methane lakes themselves (at a temperature of $\sim$90 K), large organic molecules are generally poorly resolvable, so life as we know it from Earth is unlikely to have arisen in the methane lakes, in spite of the large concentration of organic material in the atmosphere and the surface of Titan. Large silicate molecules are, however, easily resolvable in cold liquid methane, and if the long debated silicate-based life forms could really arise, develop, and thrive, Titan's methane lakes would be an obvious place to look for them.

Far below Titan's surface an ocean of liquid water is likely to exist, just as in most other larger moons in the outer solar system. Measurements from the Cassini/Huygens mission has shown that the crust is rotating relative to the core, which requires a liquid somewhere between the core and the crust. However, Titan is too cold for a pure water ocean (there are no large tidal forces to heat the interior as in Europa an Enceladus), so the ocean must contain large amounts of freezing additives, most likely amonia. Some models predict that Titan under a 25 km thick ice-crust has a few hundred km deep ocean of a 30% ammonia solution in water. Such a solution is what we would term a very strong bacteria killing cleaning detergent, if we were asking for it in a supermarket on Earth. This is certainly not an ocean where we would look for Earth-like organisms. But if most of the oxygen atoms in the biological molecules in Earth-like life was substituted with nitrogen, such "ammono-life" might thrive well in Titan's underground oceans. In the model in Fig. 1.21 it is speculated that volcanoes of H_2O and NH_3 could even exist, adding to the weather. Titan could obviously be the most interesting place to look for life as we don't know it from Earth.

Whatever the chemical basis of any life forms we can think of might be, it will need to locally lower its entropy content. In order to do so, it needs to intake low entropy energy from the surroundings, perform some kind of metabolism, and finally get rid of high entropy reactants to the surroundings. Titan might have all the ingredients for such processes, and Fig. 1.21 illustrate one way of capturing low entropy sunlight together with solar UV radiation and cosmic particles of high enough energies to break simple molecular carbon bonds, to build up some more complex (i.e. low-entropy) material, which may or may not be of potential relevance for some kind of biology. The Jupiter-moon Europa, on the other hand, do not have access to such low entropy energy, because of the ice-shield and lacking atmosphere, and no energy source with low enough entropy (high enough energy per particle) to break simple molecular carbon bonds in order to create free carbon to form complex carbon molecules of. On Mars it has been speculated that small isolated pockets of methane could be due to lifeforms living under the surface of these regions of Mars. Neither on Mars nor on Europa do we, however, see the large scale non-chemical equilibrium that we might expect, and will be looking for in the first spectra of Earth-like exoplanets we will be able to obtain less than 10 years ahead.

1.6 Are there intelligent beings outside our solar system?

In the chapter about exoplanets, we discussed shortly Frank Drake's equation for estimating the number of extraterrestrial civilizations. We will begin this section with repeating it in the context of the descriptions from the previous chapters:

$$N = R_* \cdot f_p \cdot n_e \cdot f_l \cdot f_i \cdot f_c \cdot L \tag{1.5}$$

N is the number of civilizations in our Galaxy whom we in principle can communicate with. R_* is the number of stars born in the Galaxy per time unit, f_p is the fraction of these which will have planetary systems, n_e is the number of Earth-like planets (planets suitable for life) per planetary system, f_l is the fraction of Earth-like planets where life actually originates (i.e., the probability that life will develop once the physical conditions are as on Earth), f_i is the probability of life to develop into intelligent beings once it has originated, f_c is the probability of intelligent life to develop equipment which allows interstellar communication, and finally L is the life-time of such communicating civilizations.

The difficulty is obviously not in solving the equation, but in estimating the factors. Only R_* and f_p are now reasonably well known, but we have certainly progressed well on the other factors too since Drake first wrote his equation in 1962. If 10^{11} stars have originated during 10^{10} years (almost all stars ever born in our Galaxy still exist as stars), obviously on average 10 new stars are born per year in our Galaxy, so we conclude that $R_* \approx$10/year. We concluded, in the chapter about exoplanets, that $\sim$5% of solar type stars in the solar neighbourhood have giant planets inside the snow-line, and that much fewer have Jupiter-Saturn-like planets in Jupiter-Saturn-like orbits. We also concluded that the environment of globular clusters, as well as the environment of regions closer to the Galactic centre than the solar neighbourhood, gave rise to a lower frequency of hot-jupiters than our solar neighbourhood, and we saw that hot-jupiters are much more abundant around high-metallicity stars than around low-metallicity stars. We also saw evidence that there are about 50 times more Earth-sized exoplanets than Jupiter sized, so we will expect that $f_p \approx 1$. We noticed that there are more planets than stars in our Galaxy, and that statistically there are $\sim$10 billion Earth-sized planets in the habitable zones, where the temperature could be right for possible water to exist in liquid form at the planetary surface. This could tempt us to conclude that $n_e \approx$1/10, but we need a working definition of what is meant by "Earth-like" before we can really give a reasonable estimate of the value of n_e. Should it for example just be a planet in the habitable zone, or should it be a planet which actually has liquid water at the surface, or should it (in the other direction) just be a planet (or a moon) which potentially could inhabit some kind of life we can envision (should then Europa- and Titan-like moons be called Earth-like in this understanding). We therefore conclude that it is too early to say anything significant about the abundance of habitable exoplanets – not because we do not know with reasonable accuracy (compared to several of the other terms in the equation) how many exoplanets that are in the habitable zone, but rather because we do not know what we mean by n_e. We could be tempted to claim that $n_e \approx$1/10, but it could be larger (if we include planets other than Earth-sized planets in the habitable zone), but it could in principle also be as small as 10^{-11}.

In the chapter about life on Earth, we investigated possible scenarios for the origin of life on Earth. We concluded that the building blocks for life (i.e., the large molecules that are part of life's proteins and genetic material) could have come from space (with comets or carbonaceous chondrites), but they could also have originated on Earth. We will probably need space missions to other planets and moons in our solar system in order to figure out whether places where conditions were relatively similar to ours for a period of time also gave rise to life (which may then still exist there or have left fossil traces). Only then will we be able to give some kind of an estimate of f_l in Drake's equation. We touched upon the relative importance of upright walking, tool-making, and intelligence for human development of civilization – $f_i \cdot f_c$. We also touched upon the role and consequences of cometary (or asteroid) impacts on Earth. On one hand they might have delivered the water to the Earth (and maybe even important seeds for life) and on the other hand they may have been (and still be) destructive for life through the destructive force of their impacts. Hence, it is still almost impossible to say anything qualified about f_i and f_c. Finally, the term L may be astrophysical or biological – it may be limited by the time intervals between two life/civilization-destructive cometary/asteroidal impacts (or any other cosmic catastrophes such as gamma-ray bursts or nearby supernovae) or L may be limited by a destructive biological event (a destructive virus, the evolution toward less intelligent humans, a self-destructive war, etc). It could be tempting to conclude that $L >$10,000 years.

1.7 Why have we not got contact with ET yet ?

1.7.1 Can they hear us ?

A shortcut to solving Drake's equation would be to obtain contact with an intelligent extraterrestrial civilization. The most realistic way we can today envision interstellar communication is by means of radio signals, although recent discussions and the so-called breakthrough-initiative have focused also on the possibility of communication with laser beams, and even by means of space ships flying to the nearest exoplanets. As opposed to most other electro-magnetic signals, radio waves (and well thosen laser signals) can go relatively unhindered through the entire Galaxy; the interstellar medium is quite transparent to radio frequencies.

In radio astronomy it is customary to express the power,P_n, of the signal in terms of the antenna temperature, T, which is the temperature the antenna should have in order to generate a noise signal of the same power as the measured signal,

$$P_n = kTB \tag{1.6}$$

where k = 1.3806$\times 10^{-23}$ J/K is Boltzmann's constant, and B is the bandwidth. If the bandwidth is given in hertz (= s^{-1}), the power will be measured in watt = joule per second. The more frequencies we let into the detector at the same time, the stronger the signal, and the warmer the antenna the more radiation it will emit.

Obviously it is not just the antenna that emits "noise" but also the ground and all other surrounding sources will emit noise that one will want to eliminate in order to "hear" the signal. When we look through interstellar space, the two dominant electro-magnetic noise sources are quantum noise and synchrotron radiation. Synchrotron radiation is emitted from charged particles spiraling in the magnetic fields of the Sun and Earth. While the synchrotron noise increase toward long wavelengths (low frequencies), quantum noise increase toward short wavelengths (high frequencies), and together they therefore have a minimum of noise a medium wavelengths. This defines the region where we will expect any radio communicating civilization to want to communicate over interstellar distances, and it spans the region from say ν = 1 to 30 GHz ($\sim \lambda$ = 30 to 1 cm). On top of this our own atmosphere has a minimum emittance around λ = 10 cm, with rapidly increasing noise toward shorter wavelengths, so radio communicating civilizations living on an Earth-like planet would probably prefer the radio-window from say 1 GHz to 8 GHz ($\sim \lambda$ = 30 to 4 cm).

Within this window there are two pronounced atomic and molecular transitions, namely the orto-para transition at 21 cm ($\sim$1.4GHz) in atomic hydrogen (i.e. the transition between the electron spinning parallel to anti-parallel with the nucleus in neutral hydrogen) and a transition in the hydroxyl (i.e. OH) molecule at 18 cm ($\sim$1.7GHz). Any civilization developing knowledge about the surrounding Galactic space, and engaging in how to communicate with other civilizations throughout the Galaxy we share, must at some

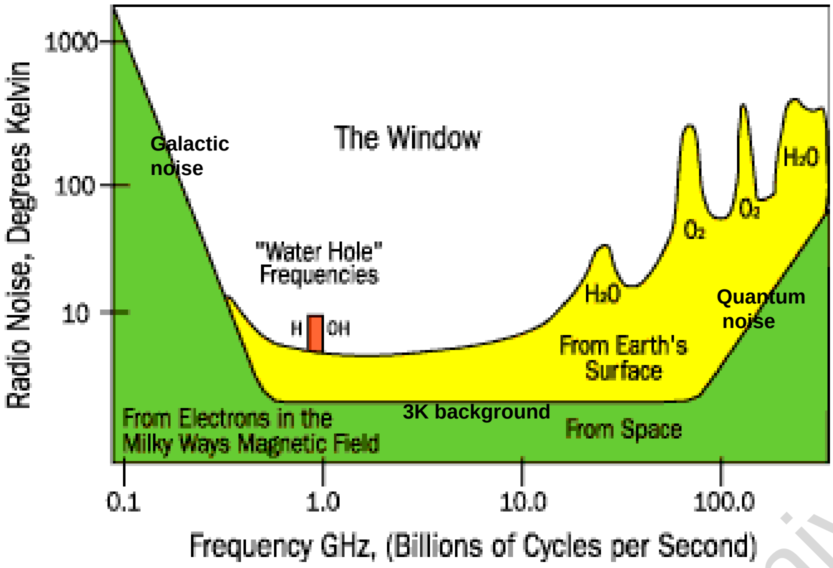

Figure 1.23. The ”waterhole” from 1.4 to 1.7 GHz may be the most likely frequency interval to establish radio contact with extraterrestrial, technologically advanced civilizations.

stage in their development have studied the interstellar medium and realized, just like us, that the medium manifest its large scale structure most clearly at wavelengths around 21 cm, and they must have realized that wavelengths around 18 cm impose important and easy obtainable additional information about its structure and chemistry. In order for them to have gained any knowledge about interstellar communication, they must therefore at some stage have build instruments that can operate (receive and emit radio waves) in this interval of the electro-magnetic spectrum. If we should guess at any wavelength interval that a galactic civilization could find natural for interstellar communication, it would therefore be the region between λ = 18 and 21 cm (ν = 1.4 to 1.7GHz). In popular terms the interval is often called ”the waterhole”; not because it has anything to do with water, but because it is between H and OH (adding to H_2O), because it defines a (region near to the) minimum in interstellar noise, because waterhole in its literal meaning is the place where animals gather and in English slang is the local place where humans exchange ideas. Here we may hope for our first interstellar exchange of ideas with our Galactic peers.

Even the best radio telescope will at the best chosen frequency obviously have to beat the natural minimum in background noise in order for its signal to be heard over interstellar distances. At its minimum the noise consist of the background thermal emission of space itself (the cosmic microwave background radiation from the Big Bang or 3-K radiation, which obviously add an antenna temperature of 3 K) plus a Kelvin or two from Earth’s atmosphere. On top of the natural radio noise, also the electronics in the receiver will obviously add some thermal noise to any transmission. Even though this radiation has been decreased to perfection during the latest decades, even this has a lower limit of a few kelvins set by the second law of thermodynamics. State-off-the-art of todays radio telescopes is able to limit the total background noise (including the cosmic microwave background, the ground around the antenna, the receiver electronics, etc) to 15 K. Any signal we should be able to receive from our cosmic brethrens, must therefore reach us at a power corresponding to at least an antenna temperature of 15 K (or close to this) to be extractable from the noise, and any signal we might want to sent into interstellar space to be received by extraterrestrials, must also have a strength that allow it to be received with an antenna temperature of the order of no less than 15 K. Are we able to do that?

Since the power is directly proportional to the bandwidth, it obviously requires less energy to send a signal of low bandwidth. However, the lower the bandwidth, the less information is it possible to include into the message. Normal speech requires ∼2.5 kHz, and a standard television signal occupies ∼4.5 mHz. If we are happy with a minimal information content of say 5 bits per second, it requires a bandwidth of $\sim$2.5 Hz. The noise power of a good radio telescope on Earth (or at any other typical place in the Galaxy) would according to Eq. 1.6 and the above description of the natural background noise be

$$P_n = 1.38 \times 10^{-23} \times 15 \times 2.5 = 5.2 \times 10^{-22} \text{ watt} \tag{1.7}$$

If we want to sent such a signal omnidirectionally (i.e. in all directions) into space and want that a potential receiver within say 100 light years ($\approx 10^{18}$ meters), and we for simplicity assume that the receiving antenna has an effective receiving area of 1 m^2, it requires an input energy of

$$P_n = 5.2 \times 10^{-22} \times 4\pi \times (10^{18})^2 = 6.5 \times 10^{15} \text{ watt} \tag{1.8}$$

This is far more than the total present-day global electricity production. Our civilization will therefore not be able to send even such a simple general omnidirectional radio signal about our existence that a civilisation similar to us would be able to hear as anything above the natural background noise. To send a TV signal with a million times broader band width would require a million times more energy in order to be heard as something above the natural noise by a civilisation at a planet around a star 100 light years away.

The opening angle of a beam transmitted or received by a parabolic dish antenna is proportional to the diameter of the antenna. We have above assumed that the alien civilization used an antenna with an effective area of 1 m^2. Such an antenna would have an opening angle of 11 degrees, or cover 0.3% of the sky, and the receivers would, hence, need to constantly scan through a few hundred places on the sky if they had no idea where to look for us. The receiving efficiency could of course be increased if the alien civilization used a larger radio telescope, but their search area of the sky would be correspondingly smaller. A 1 km^2 area telescope, as the one we are presently planning to construct on Earth, would require from us a million times weaker transmission power (6.5×10^9 watts) which would in principle be doable by a coordinated effort of our civilization, but the alien civilizations receivers would have a beam opening a million times smaller, and hence would need to scan a million times more places of the sky to discover our signal if they didn't know where to look. The conclusion is therefore that omnidirectional information about our existence is basically impossible.

Fortunately, we saw in the chapter about exoplanets, that the traces of life in the form of low-entropy atmospheric signatures (a spectrum showing non-equilibrium chemistry) will be observable from ground-based telescopes within a few years, for stars within 100 light years distance. I.e. that we have a chance of within a few years being able to point out which stars are orbited by planets with life, and hence potentially intelligent civilizations. Once we know where to aim our radio signals we have a much better chance of being heard; or in the equivalent but more cautious approach: to receive a possible intelligent signal. Listening toward TRAPPIST-1 and Proxima Centauri has allready begun. Assume for a moment that we have constructed the planned 1 km^2 array telescope (www.skatelescope.org/), and that it is build from parabolic dish antennas. Then the beam will transmit toward 3×10^{-9} of the full sky. We will therefore need 3×10^{-9} lower power than estimated above for a omnidirectional beaming, or just 20 watt of transmission power in order to be heard just above the noise level of a corresponding 1 km^2 receiver on another planet within 100 light years away. Of course they will also need to know that we are here, because their beam size is also so little that they will have a very small chance of pointing at us if they point at random in the sky. The conclusion is therefore that civilizations that are beyond the point where they have build and operated ELT-like search instruments for extraterrestrial life, will be able to communicate with one another over hundreds of light years. If we invest a transmitting beam power of 20 million watt instead of 20 watt, we will reach 1000 times further out in space, covering in principle all of our Galaxy. Hence the distance in our Galaxy within which we in principle would be able to communicate with civilizations similar to ours, is no longer set by our ability to send radio signals that can be heard (or to hear theirs), but by our ability to pinpoint where to aim the signals. This ability is at present (i.e., will be within a decade or two, because we are not quite at this capacity yet with any telescopes we have build) limited to of the order of 100 light years distance, but other civilizations that have been able to map the planets with low-entropy atmospheres out to larger distance from where they live (possible to all distances throughout the Galaxy), will at present be able to sent us signals that we soon (whenever soon we will eventually have a 1 km^2 radio telescope at disposal) will be able to receive.

Of course electro-magnetic radiation travels only with the speed of light, so we have amble time to think about what we will respond to a civilization that is possible beaming to us from the other end of the Galaxy. Our atmosphere have been out of equilibrium for more than 1 billion years, so they may have been beaming at us for quite some time, but we are almost ready now to receive the signal. Maybe we are even ready to respond to it (?), but it will take another 200,000 years at the worst before we have a chance to know what they think about the answer we sent.

1.7.2 Can we reach them?

Four ”postcards” and two ”LP-records” have been sent into interstellar space with a message about our existence, on board the Pioner (10 and 11) and Voyager (1 and 2) space ships. At a speed of 3 AU per year

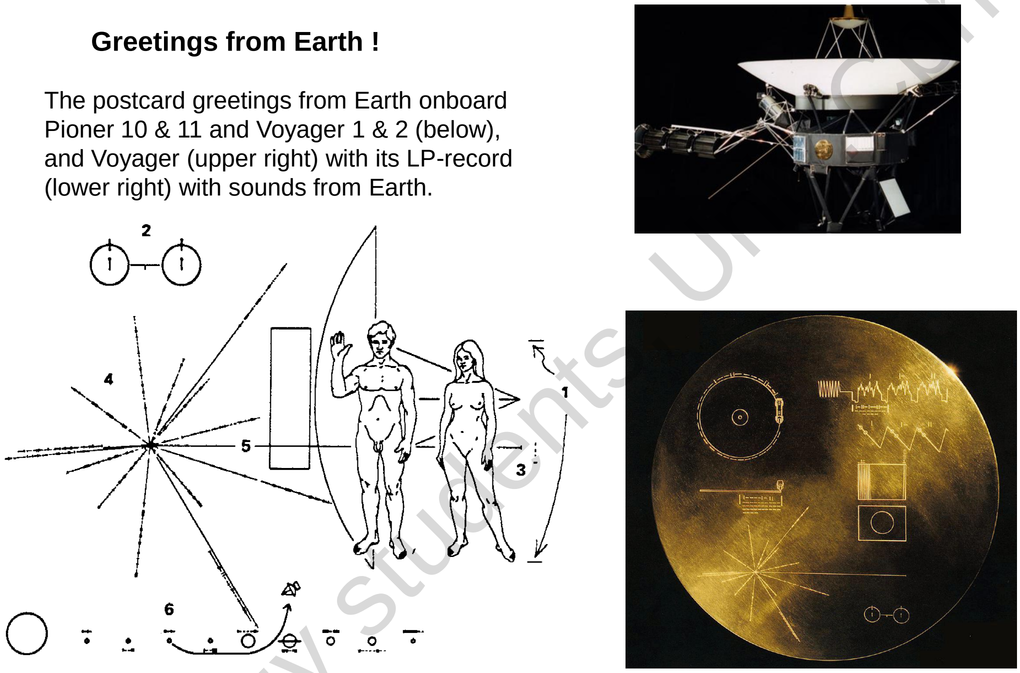

Figure 1.24. Four ”postcards” and two ”LP-records” are on their way into interstellar space with greetings and information from Earth.

they are now passing through the heliosheat, the region that seperates the solar system from the surrounding interstellar space. The heliosheat is expected to end with a bow-shock which Voyager is expected to pass through in a few years, after which time it is in interstellar space. In 296,000 years it will pass Sirius (although at a distance of 4,6 light years). Nobody expect that they will ever reach anyone out there, and the 4 satellites are not aiming anywhere specific in our Galaxy. The postcards and the records are probably better understood as an exercise in trying to think through how we could inform an alien civilisation something meaningful about us.

The Breakthrough initiative was initiated by Russian internet- and investment billionire Yuri Milner, now involving other billionaires too and a board of Steven Hawking, Martin Rees, Geoff Marcy, Ann Druyan and Frank Drake. The project has committed ∼100 million $ to the breakthrough listen SETI project and ∼10 billion $ to the breakthrough starshot project. The initiative also includes an annual prize in physics and in medicine/biology of 3 M$ (twice the size of the Nobel prize) given since 2012 to 70 scientists and groups, based on a public internet nomination followed by a peer review selection by previous laurates.

The breakthrough listen project will receive 100 million US$ during the next 10 years to search for ETs. 1/3 is for buying observing time at large radio telescopes and Lick Observatory (visual signals from ET laser transmissions); 1/3 is for developing hardware, and 1/3 for a employing a staff of astronomers. The project

Figure 1.25. The breakthrough project will focus on intensive search for extraterrestrial radio signals (upper left), construction of thousands of nano-satellites to fly to exoplanets. Lower right panel shows Yuri Milner and Stephen Hawking at the announcing meeting at the Royal Society in London.

will search automatically over all frequencies in a wide range of the GHz region, searching the nearest 1 million stars, 100 galaxies, the the MilkyWay plane. The 10GB data obtained per second, and the associated software, will be public domain, using also SETI@home. There is a 1 million $ prize for best digital message representing humanity, decodable for ET -– to potentially be broadcasted after public global discussions of the ethics.

Finally, the breakthrough starshot project is a 10 billion $ project to launch 1000 cm-sized spacecrafts, the StarChips, each with 4x4m light sail, to Alpha Centauri by 2036 at a speed of 1/5 C, reaching the destination by $\sim$ 2060. An array of GW lasers on the ground will accelerate the satellites 100 km/s/s for 10 minutes, using about a TJ to launch each satellites, or the total output of a large nuclear power station for a week.

1.7.3 Are we the only technological civilization in the Galaxy?

So why have we not heard from ET yet? – the explanation is simple enough: we have no clue yet where to direct our radio telescopes, and our telescopes are on the lower limit in size to receive the signals we might expect. We are almost at the threshold that will make it possible for us to overcome this obstacle, and we may well see it passed within our lifetime. However, there is a completely other consideration that make many scientists think that even if we build large enough radio telescopes, and even if we find planets out of chemical equilibrium caused by biological activity at their surfaces, then there will be no-one to communicate with.

The argument was first raised by the Italian-American physicist Enrico Fermi, and is now called the Fermi paradox. Fermi reasoned as follow:
It took Homo Sapiens of the order of 100,000 years to spread to all corners of the globe. Whenever people reached a new destination, it was a question of relatively short time before some of the settlers found a challenge in settling the next nearby virgin land. Genetic studies of the human mitochondria showed how and when it happened, as is sketched in Fig. 1.26. It is about 500 years ago Europeans (re-)colonized the American continent. It took about 500 years from then until USA landed the first human on the nearest other body in space. It will take less than 100 years before nations of the Earth will establish the first permanent

colonies on the Moon and on Mars. People on Mars (and perhaps nations on Earth) will soon after colonize the martian moons and near-by asteroids, and soon some of the settlements will find perhaps the Jupiter moons near enough and challenging enough to establish the first outposts and settlements there too. From then on the step to the rest of the solar system is small. Fermi argued that it is obviously a build in human nature to seek the challenges in finding new areas to live, disregarding the difficulties. Most likely there were more death-casualties during the first trans-Atlantic expansion than there will ever be during the coming transports to Mars. Maybe this urge of expansion (and challenges) is even a necessary feature of the species that will eventually dominate the planet where it lives. If this is correctly seen, then there will be a day where humans from one of the posts in our solar system will overcome the challenge of traveling to another solar system. Perhaps it will be hundred years ahead, perhaps it will not be earlier than 1000 years from now, but if Fermi is correct it will happen one day. In principle we know of the technology that is required to travel to nearby habitable exoplanets, and the travel time is such that young people embarking on the travel might experience the arrival. Once humans have spread and settled through most of our solar system (and even more so once they have settled exoplanets around nearby stars) it is very difficult to envision a catastrophe that could ever extinct all of the populations. This day is very close from an evolutionary point of view, and once it has happened is is hard to see how humanity could perish from the Galaxy again.

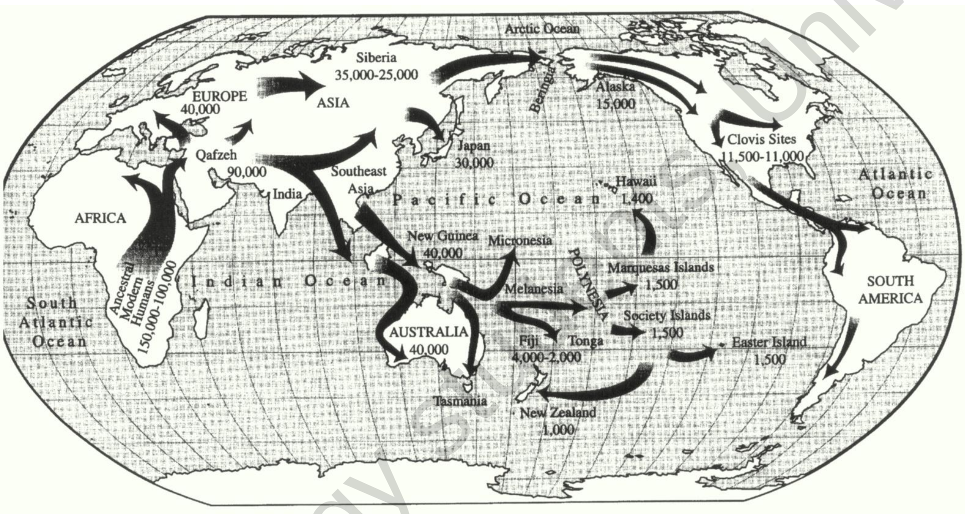

Figure 1.26. The human mitochondria genom project has revealed the route and timescales on which Homo sapiens populated the globe.

The Fermi paradox is now the conclusion from the above considerations: Some stars, that in principle could have formed exoplanets quite identical to our Earth, will have formed several billion years before ours. If on just one of these planets a species with the same expansive and technological ambitions and abilities as ours had formed, their civilization could have developed to our present level billions of years ago. If there is no unknown physical limitation that will make human interstellar travel even in principle impossible (which is quite hard to envision what should be based on our present knowledge), then the humans from our solar system will establish their existence on nearby exoplanets within say a few thousand years. If their lives and basic nature continue to be as their biological ancestors, some of them will soon find similar challenges in moving to nearby other stellar systems. Even if we assumed that it took 5000 years from arrival to the time the first group of settler-descendants wanted to move elsewhere, it is an exponential expansion, and humans would have settled every habitable solar system around every single star in the entire Galaxy within 50 million years – less than a percentage of the time that have passed from the arise of the potentially first civilization in our Galaxy to the formation of the Earth. Why are they not already here?

Fermi's conclusion was that there has never existed any other civilization in the entire Galaxy with the technological and expansive capacities as ours, otherwise they would already be here. Therefore obviously there are no-one to send radio signals to either and no-one to listen for.

On Earth itself, the life time of an average species may be of the order of 10^6 years, and there may be of the order of 10^8 different species living on the planet today. This implies that there may have existed as many different species on Eath as there are stars in our Galaxy, 10^{11} in total, and only one of them became space faring and radio communicating. It is such a species we are looking for in the SETI project.

1.7.4 Is *now* the right time to contact ET?

4. ARGUMENTS FOR REPLY

4.1. Distance insulates us

Even with speed-of-light electromagnetic communication, the consequences of reply would not be felt for a very long time. The greater the round trip time, the greater the insulation. The more distant the point of origin to the signal, the less immediate the consequences of a reply. Arguments for reply should be strengthened in direct proportion to the distance.

4.2. Physical contact with ETI is unlikely

Physical contact is immensely more difficult than electromagnetic contact. We would be physically safe. Again, as distance increases this argument would become more credible [2].

4.3. We cannot hide

For nearly a century we have emitted low-power electromagnetic radiation into space. For the past two decades our strong planetary and military radars have emitted very powerful signals into the interstellar medium. If we have not already been discovered, eventually we will be discovered.

4.4. Delayed reply would deny us information that may be important or even essential for our survival

We are in a period of planetary adolescence. We are a threat to ourselves. It may be possible to learn how to manage our hostilities and our environment by communicating with others who have more maturity as a civilization.

4.5. A general knowledge gain is made possible by replying

If the interaction with ETI can be established, knowledge that could cure disease, extend life and revolutionize science and philosophy may be waiting for us. The knowledge gained could be beyond our ability to even imagine.

4.6. The sender is likely to be benevolent

In the words of Arthur C. Clark, 'As our own civilization is in the process of proving, one cannot have superior science and inferior morals. The combination is unstable and self-destroying' [3].

5. ARGUMENTS AGAINST REPLY

Several arguments could be used to counter the above list of pro-reply statements. They are as follows:

5.1. Unknown technologies for conquering distance may exist

We are just beginning our history as a space-faring civilization. We should not assume that the technologies we know are the only technologies that exist. Superluminal (faster-than light) technologies cannot be ruled out [4].

5.2. Physical contact is more probable than currently thought

Our primitive science and technology underestimates the possibilities of physical travel on an interstellar level. Physical contact is more probable than we now believe.

5.3. Our electromagnetic emissions are difficult to detect at great distances. We may be able to hide

The vast majority of signals emitted from Earth are very weak and hard to detect at great distances. The more powerful signals that have been emitted in recent decades are fewer and for the most part not designed to carry information. We may still be safe from detection.

5.4. Delayed reply would give us time to listen for additional signals and perhaps learn more about ETI. This would allow us time to develop technologies that would give us more security

We should not rush to reply. Delay and passive monitoring of the emissions of ETI would allow us to perhaps receive additional signals that would give us more information about the nature of the sender.

If we were to wait for decades or even centuries before replying, we would have time to improve our own technologies, making us more secure as a species.

5.5. The general knowledge acquired by replying could be dangerous

The cultural, theological, and philosophical knowledge obtained from ETI could weaken and perhaps destroy human allegiance to existing institutions. The culture shock could destroy major elements of our social order [5].

5.6. The likelihood of a malicious sender cannot be ruled out

Science has demonstrated that evolution is blind and brutal. Evolution moves towards the direction that gives a species survival advantage. Since evolution is not a benevolent enterprise, there is little reason to assume a benevolent extraterrestrial civilization.

Figure 1.27. Are these the right security considerations concerning contact with extraterrestrial civilizations ?

The Earth was created 4.6 billion years ago. Try to scale this time down to one year, just to understand this tremendous time span inside your mind. Then the Earth was created on January the 1st. During the beginning of March the crust of the Earth had solidified, water had appeared on the solid surface, and the first living organisms arose in the ocean soon after. The first eukaryotic cell – the kind of cell our own body and all other animals and plants are made of – appeared on the globe around September 1st. The first animals left the ocean and invaded the land on December 1st. The first mammals appeared on December 15. The last common forefather of humans and the other great apes passed away in the African jungle on December 31 at half past 2 in the afternoon. Homo erectus produced his fantastic stone axis New Year's evening from approximately 8 to 10 pm. The first Homo sapiens arrived on stage on the African Savanna at 11:37 pm, and they left Africa for the first time on 11:49 pm – as what we today would have called an extinction endangered species! On 11:58 pm they had spread all over the globe, and on 11:59:57 those that traveled west out of Africa meet for the first time the descendants of those that traveled east out of Africa. Shortly after, 11:59:59 New Year's evening, they constructed the first radio. At this moment, on New Year's evening at the stroke of 24, we point our most modern radio telescopes toward space, and listen if there are other civilizations on other planets somewhere out there far away, who have also constructed a radio, and are now also in this single second of their evolution, or somewhere in their following minutes, where they try to send us a message. If they had existed a few hours more than the initial first radiotransmitting second we are now in, then they would already be here, according to the Fermi paradox. If technological civilisations usually die out in less than a second after their first radio transmission, then we have to be lucky to hear them. We will soon obtain the first high-resolution spectra of Earth-like habitable exoplanets. Will we see traces of only primitive bacterial-like lifeforms, will we see oxygen from photosynthetic life, will we see abundance of technologically produced gasses in their atmospheres, or will we see the scarring remnants of all-destructive biological and nuclear warfare everywhere where intelligent species can potentially have existed?

If we truly believe that civilizations of our kind have only an "evolutionary second" to exist in, or are so rare that the N in the Drake equation (Eq. 1.5) becomes infinitely small for reasons of any of the other terms, then there is no reason to take the effort to keep listening for signals from extraterrestrial civilizations; we will never receive a message, and if we receive one anyway there is no point in responding, because they are allready long gone anyway.

If, on the other hand, we believe that it is possible that there are peacefull, loving, intelligent aliens like ourselves somewhere out there in space that want to communicate with us, then we need to think well in time about what we want to do if we one day receive a signal. Receiving a message from a civilization far superior to ours may have such a tremendous and irreversible effect on our societies and civilization that we can hardly imagine. All experience from our own globe seems to indicate that sudden contact between civilizations at slightly different stage of development has always lead to mind-blowing consequences – usually not optimal ones for the technologically least developed, so it might be wise for us to prepare ourselves a bit. The consequences of contact with a civilization that has a billion years longer evolutionary timescale behind it than ours is beyond imagination and beyond conclusions that can be drawn from extrapolation of present experience of contact between human civilizations, or even conclusions drawn from the past contact between the civilizations of Homo Sapiens and those of Homo Neanderthal and Homo Erectus. It is a very serious issue that urgently needs international considerations, as we rapidly increase our ability to receive as well as send signals to and from alien civilisations. When TV journalists in soap-programs and in order to become popular promote the viewers to send photos of themselves to the TV station for transmission toward the nearest habitable exoplanets, we can only comfort ourselves with knowing that modern antennas of TV stations are to small to raise the signal transmitting the images above the cosmic background noise. This technological security can howerver not be expected to last long, and I include in Fig. 1.27 some security considerations that were published in Acta Astronomica some years ago.

Take a moment to think about what and who you think should answer to ET on your behalf if we get contact one day – and consider carefully whether you think it would be wisest to answer or rather just to pretend that we are not here. It may be the biggest decision Homo sapiens have had to take ever since some of our forefathers decided to leave Africa $\sim$100,000 years ago.

1.8 How did we become (so) human?

The theory of evolution is often simplified in one sentence, *"survival of the fittest"*. *Survival* has a clear cut meaning, but what does it mean to be *"the fittest"*? To fit what, when, how and for whom? In Darwin's original idea, it meant something like "managing well in a given environment, in the sense of competing successfully for the food and multiplying more than competing species". The organism that was best fit to the environment at a given place at a given time would get most outcome, and for some time therefore the population of its species would expand, untill some new organism or some new variant of the same organism became more fit, usually because the environment would change. It is nothing abrupt and dramatic, but rather *"a peaceful competition in a changing environment over long timescales"* that to many people feels "justified, fair and natural", even though most people don't like, or even ignore, that it also means that we as humans must expect to be here on Earth for a few hundred thousand years and then give place for something new. A combination of this traditional Darwinism with more genetic theory (including those originally developed by Mendle's experiments with plants) is often called neo-Darwinism. With an even deeper understanding of the role of the genes themselves, focus shifted from the fittest organisms to the fittest gene. This latter view include in particular the central and influential work in the 1960'ies by William D. Hamilton, later popularized and brought further most successfully by Richard Dawkins with his book ”The selfish gene” from 1976. In this interpretation it is not the individuals as such that compete, but the genes directly. It became a shift from the selfish organism to the selfish gene, survival of the fittest gene rather than the fittest organism. For many it feels as a depressive and more machine-like view that the gene itself could strive to become most fit and not the organism. It could look like that we ourselves do not have a free will, but are just slaves to our genes manipulating us around through its own independent plans. Personally, however, I rather think of it as modern psycology; the more we understand about why we react in certain ways, the more free will we have to chose whether to react that ”automatic way” or differently. All other animals on this planet are bound to react as dictated by their *selfish genes*, while humans as the first species on Earth has become capable to take control over not only the immediate surroundings, but also over ourselves and our behaviour. How did it happen? ..and again the question: ”is it completely unique to us and on this planet?” or is it just the unavoidable way evolution always goes on all habitable planets?

During the 4.6 billion years existence of our own solar system it happened only once, and only at one place. Today we are, as the only organism in the solar system, in the middle of not only the genetic evolution all species are a product of, but evolving based on three evolutionary instruments, the gene that has guided all evolution since billions of years, and probably determines our feelings and instincts, *the meme* that has uniquely affected our own evolution in the latest few thousand of years, and the scientific approach that has allowed us during the latest few hundred of years to understand it all and decide how we want our evolution to proceed. We are the only species we know of that has ever been affected by more than one kind of evolution. In this chapter we will dive a bit deeper into looking at what an enormous change this is, and what it may mean for the development of us, for life, for our planet, and maybe even for the universe.

The genetic meaning of ”the fittest” may be expressed as rutely as ”the population of organisms that under given environmental conditions is capable of multiplying fastest, and hence compete most effeciently for the available resources.” For higher lifeforms (but not for simple organisms such as a bacteria) the success requires at least three things: 1) The individual needs to be good at gathering and digesting neutritious food, 2) it needs to be good at attracting healthy partners to mate with, and 3) it needs to live long enough to benefit from the ability of 1) and 2). In Darwin’s ”survival of the fittest”, the goal, or ”meaning”, of (higher) life is to maximize 1)+2)+3).

Point 1 and 3 are simple. The lion with the sharpest teeth is the best killer. Those lions that had sharpest teeth, the most sensitive smelling sense, sharp eyes sitting in the front of the head for sterioscopic view and distance determination to the pray, got most food. Therefore they (or the genes expressing these features) survived better and therefore multiplied fastest (i.e. got more outcome), so that with time the species of lions had (genes coding for) sharp teeth (and good smelling and stereoscopic vision etc). Those gezellas that ran fastest (and had a good hearing, good wide-field view in all directins to best spot the predators in time) survived the lions best, so they lived longer and got more kids, and therfore the species of gazellas became fast runners (with sharp hearing etc). Those hunter-gathers that were able to hunt most efficiently gathered most meet, and those that were best at finding things to gather, gathered most. They survived best, therefore humans became skillful hunters-gathers. Humans became better hunters than lions and survived in a larger variety of environments. Therefore there are more humans than there are lions. From the gene’s perspective, the best gene survived and reproduced itself by means of a suitable organism (sometimes called the genes-eye perspective or the selfish gene). From the organism’s point of view (the more traditional Darwinistic view) the best organisms survived by means of evolving the best genes. Some aspects of evolution are best explained by the first point of view (for example human altruism, lemmings or bees collective suicides), but from both points of view 1), 2), 3) describre the fittest and the ”meaning of life”.

Before diving deeper into point 2) we will take a de-tour to look at whether it is really the organism or the gene that is the selfish driver of the evolution (and whether it is actually ”selfish” at all). This may at first seem just an academic discussion, because after all the organisms evolve independent of whether we claim that it is the gene or the organism that sits in the driver’s seat. But if it is the organism as such, we will get more and more fit organisms, while if it is the gene we will get more and more successful features. The gene(s) for a large and colourful peacock tail is inconvinient for the peacock because it makes it more difficult to escape the predators, but if the gene for beautiful peacocktail is able to promote itself, time will produce more and more beautiful peacocktails. A more difficult example is the genes for altruism. The everyday meaning of the word *altruism* is ”an action that will benefit the receiver on the cost of a disadvantage for the giver”. In a biological sense it means ”an act that will make it more likely for the receiver to produce more kids (i.e. multiply the receiver’s genes) on the cost of the giver getting less kids (i.e. relatively reducing the giver’s genes)”. This makes no sense from a traditional Darwinistic point of view, and Darwin was aware that there was no room for such a behaviour in his original theory. An altruistic person would get fewer offspring than a non-altruistic person, therefore the altruistic gene would spread less than the genes of the non altruistic person, and hence the gene coding for altruism, and hence the consept of altruism, would automatically die out.

It became William D. Hamilton that in the 1960’ies layed down the mathematical foundation for what is now the accepted solution (and furthered in what was later called the selfish gene theory). It can be expressed in the simple formula

$$C < r \times B \tag{1.9}$$

C represent the cost for the giver, *B* is the benefit for the receiver, and *r* is a number between 0 and 1 that states the genetic relation between the giver and the receiver. It is of course the definition of the factors that is the difficulty, and we will not go into details with it, only state that the relation factor *r* is one for one-egged twins, 0.5 between parents and children, and approaching 0 between complete strangers (e.g. a spider and an individual human). *B* in biological terms is a measure of how many more kids it is likely that the altruistic act will bless the reciver with. In order that it shall work in pracis, the giver must be able to spot in a fraction of a second what a likely value of *B* for the receiver is (and what the values of *C* and *r* are). If it is a question of, for example, rescuing another person from drowning with the risk of drowning yourself, then you cannot pick out your calculator first to calculate *C* and *B* and fill in a questionary to determine *r*, but you must act on intuition (i.e. genetic experience). If the receiver has large head compared to the body and large eyes, it is probably a child if it is a human. It will therefore have a long life in front that can give rise to many children. *B* will be large. If the giver is old and have already had the children expected in its lifetime then *C* in terms of the likely future genetic multiplication is low. Therfore a grown up person would immediately jump into the water to save the child. Since it is based on genetic intuition (”feelings”), we would reach the same reaction toward anything that had the same tricker signals as the look of a human child, and we would rather save a cute looking animal (i.e. large eyes, large head compared to the body, etc) than an ogly looking monster, even though we could have no a priory knowledge about which one is ”best”. We will react the same way in potential encounters with aliens, and that’s why an ”invasion-from-Mars” film would be no block buster if the invaders looked like ET, but would sell better if the invaders looked like overgrown spiders.

Hamilton’s rule is so much incoorporated into our every day moral and laws that we might expect that any other advanced civilisation in the universe formed by the principles of evolution must follow the same moral codex as we and have the same basic law system – at least ”primitive civilisation” still ruled by the laws of evolution. Most people would find it natural if you spend 1,000 Euro on something one of your family members are in need of, but would grate it as generous if you spend 100 Euro as a gift to support a random stranger in the same need. Most people expect that your children inherit after you, and most societies have a minimum percentage that your children by law has to inherit, and if you have no children, relatives with the largest value of r will automatically by law inherit you. These sets of morals and laws were already established at the savannah. In fact they were in place already long before that, since we see also Chimpanzees and even insects following them. Most pronounced is so-called eusocial insects such as bees. A bee hive is a comunity of thousands of siblings, most of which are not reproductive. The non-reproductive bees (”the workers”) will per instinct gladly attack potential predators (such as you and me) in swarms, if the intruder comes too close to the hive. Their ”C”-value is infinitely small and they therefore instinctly gladly stick you even if it cost them their lives to do so, if just ”B” is a bit above zero. These are basic genetic laws, set to promote specific genes that multiply well, and they follow Hamilton’s rule.

Humans do follow Hamilton’s rule, but we also react more complex that just that. We would gladly spend a little time helping a tourist to find the way, while a Chimpanzee would be reluctant to show a chimp from the competing group the rute to the banan plant. We are not in general aggresive toward strangers, and it even touch our hearts to see people help strangers ”for no reason”, so from a genetic point of view, people with helpful emotions must have had an evolutionary advantage; the genes promoting helpful nature must have been more fit for human evolution than the opposite, otherwise they would have died out under evolutionary pressure. This is what psycologists call *reciprocal altruism*. If I help you at some time when you are in need, I expect that you help me at some other time when I am in need. But it comes in a package with feelings of sticing together if both fulfill the rules (”friendship”) and emotions of offence and angriness if you cheat and don’t help when I am in need (”disapointment”), and bad consciesness if I know I didn’t live up to these genetic rules. Otherwise the genes promoting altruism toward non-relatives would die out. They would not be able to compete with those genes promoting the behaviour of ”free-riders” (”cheaters” or ”selfish organisms” in the language of traditional Darwinism). It is as the saying ”you cheat me once, blame on you; you cheat me twice, blame on me”. Honey bees, fish, and seaguls don’t have this behaviour, but Chimpanzees do. It requires a certain memory of the behaviour of individual fellow organisms to work, but once you are there it is beneficial for the evolution of your species. It is good to have 3 friends you trust if you are to hunt a buffalo, and the meet you gather is enough to help you getting more friends if you know how to share it with people you can trust. If evolutionary theory is as fundamental as the laws of gravity, friendship and collaboration is necessary to reach the state of a technological civilisation, and potential alien visitors will understand the consept of friendship and cherish reciprocal altruism. Arthur Clarck once said that ”any alien civilisation we may get into contact with will necessarily be friendly; those that were not friendly by nature will long ago have destroied themselves”. We may from the above discussion suspect that they will never even reach the state of advanced technological civilisation without being friendly, and may speculate whether we ourselves are dangerously mid-way inbetween, but certainly the selfish-gene version of Darwin’s evolutionary theory tells us that there is a strong and natural connection between the different levels of peacefullness among the gorillas, the chimpanzees and our own species and the level of evolutionary success – and there is nothing particular ”selfish” in that, rather the contrary.

Point 2 in the meaning of ”being fit” above is even more tricky. The strongest male gorilla is able to gather more food than weaker male gorillas, which in itself makes him stronger and gives him a longer life to mate and spread his genes. But it also makes him more attractive as mating partner, because mating with exactly him will make it more likely for the kids to survive because they statistically will also be strong. Hence the female gorillas that are good at spotting and appreciate his qualities will spread their own genes optimally too. He will gather gorilla females around him, and he will get more ofspring in a double sense. It is not important in this strategy for the females to be large and strong, and it is not important for the females whom of them appears most attractive for the male gorilla, their productivity is limited by the time they will have to carry the fetus, and it is probably not only us that think they look rather similar to one another. From an evolutionary point of view, they just need to be attracted to the strong alpha-males so that they are not tempted to run away and spread weaker gorilla’s genes, which is simple because strong males will be able to better provide security for them and their outcome, and to chase away the weaker males. In this picture the best surviving gorilla groups will consist of a large gorilla male and a harem of smaller gorilla females, with a strategy for point 1), 2) and 3) fitting better into the environment than potential other gorilla groups without this strategy. They fit into Darwin’s original picture of the ”selfish individual” (as well as ”the selfish gene”); all are happy in an evolutionary sense. Of course on a presonal level all may not be happy, e.g. the physically weaker males may feel neglected, the females may feel surpressed by the alfa-male, and the silver-back may feel over-loaded with work and stress, but ”natural selection” is not about happiness but about ”fittnes” which we have defined as the ability of the species (or genes or individuals or groups of members in a species) to produce most surviving offspring in competition with other species (or genes or individuals or groups within the species).

Both chimpanzees and bonobos multiply based on another genetic strategy. One could feel tempted to be ruthless enough to say that nature had chosen to experiment with different strategies among otherwise similar species, and let the fittest move forward to the next experiment. For chimps and bonobos the most active will have the largest chance to increase its genes in the genepoole. Nature has therfore (with time and genetic competition) equiped chimp and bonobo males with enormous tecticles compared to the gorilla males (with humans in between). During observations of chimps in the wild, females have been observed to have up to 50 intercourses per day with most of the group’s male chimps. This strategy is not likely to further any particular chimp’s genes, but to assure the group an optimal number of outcomes. Our slightly more distant cussin, the lesser ape Gibbon, has thosen the strategy to form pair-bonds and stay together in couples during lifetime. As a result, the males and females of chimps, bonobos, and gibbons, are of approximately the same size and strength, as opposed to gorillas. Evolution has chosen to diversify its strategies of survival, and to let time determine which of the genetic strategies were most successful, if any of them.

The deer male with the largest antlers is best capable of fighting away the other deer males. Therefore he will have most hins around him, he will get most outcome, thereby multiplying the genes for large antles down through the generations, and the population of surviving deers will have males with large antlers. The antles by themselves are, however, not useful for gathering food (as with the gorilla muscles), so the strategy is good for point 2, but not for point 1 and 3. The deer would be better off from having more muscles (maybe in the legs to run fast) or a long neck to reach leaves in larger parts of the trees, but evolution is not about logic but about random experiments, and he has no such choice; it's the gene that determines the strategy. The large antles are not *per se* an advantage for the survival of the individual, but they are useful for the species of deers with large antles to survive – i.e. for the gene of large antles to survive. Random mutations are experiments with new genes leading to new forms of organisms. Those genes that give rise to organisms that survive best in the given environemnts will be transfered in largest amount to the following generations, and their expression will dominate the species. The reason deers with big antles do better than deers with smaller antles is point 2) rather than point 1), and it may be because large antles intemidate, scare or simply beat down deers with smaller antlers, promoting at the same time as a side effect genes for fighting and violence in best fitting male deers, hence with time making the population of deers consist of relatively agressive and risktaking males and relatively peaceful hins.

The peacock has it even more inconvinient as far as concerns survival of the individual. The excess energy necessary to produce a beautiful and colourful tail (as well as large antlers) require a strong and healthy male. By showing the most beautiful tail, the male therefore signals "I'm a powerfull male full of excess energy, my genes will make it more likely for my offspring to survive". It is another way to intimidate the surrounding peacocks, but with our own sense of impression of beauty in the colorfull tail we could easily think he has touched the hearts of the peahens by posessing his beautiful tail – it may be seen as a peacock beauty contest for the females; a difficult-to-cheat signal of healthy genes. Whichever way, the result is of course the same; the male with the most impresive tail dominates the party. However, a big tail makes it unnecessarily difficult for the owner to maneuvre. For an animal that is a prey this seems an unpractical 1)+2)+3) strategy, but it underlies the complexity of point 2, and stress the fact that survival of the fittest is not a question of survival of the individual. It is not even the species that get fittest under the evolutionary pressure. In that case peahens would have developed to be attracted by something of more practical importance than a colourfull tail, and peacocks would have been scared away by for example other peacocks superior ability to escape predators. But evolution is not a rational discussion, but an ongoing random experiment. In the case of the peacock tail it is the experiment with a (selfish) gene for big peacock tail that is moving forward and only time will show whether it by and large was a good strategy or not.

What about humans? What kind of experiments are our selfish genes doing with us? Is/was our genetic evolution guided by "the peacock's tail" or "the lion's teeth"? The answer is that we show a funy mixture of both, plus some surprising additions that made us take a crucial direction different from any other animal we know of. Point 1) and 3) are simple. We are doing relatively well, although we are for the moment somewhat out of tune with evolution, but it will always be like that (for all species). The genes that at a given point in time has succeded in spreading most are those that were best fit to the *past* environment, and when the environment change they will always "live in the past". We (and all other species) will have instincts and feelings that are guided by fitting best to the range of environments our forefathers lived in, not those we live in. For a large fraction of our past (but not all of it) our forefathers lived on the African savannah, and our feelings and instinctive reactions are therfore well adapted to the life on the African savannah, but right now we "live in Candyland", not on the savannah. We are eating as much sugar as we can get around to because it was rare at the savannah (point 1), and we are shortening our lifes and life-quality (point 1 and 3) by inhaling tobaco-smoke because it is fun or may have peacock-tale like effects. This is what evolutionary psycologists call evolutionary mismach. We are afraid of spiders and large carnevors because they were dangeous for our forefathers in the jungle and at the savannah, while it would make more sense to be afraid of hamburgers and cars because they kill far more people today than carnevors. When our genes competed about being the most fit, hunger was an ever-present life threat, so today our genes tell us to eat almost all we can get to (in particular sugar, salt and starchy carbohydrates that were in need on the savannah). For the first time in human history it is much more likely to die from overeating than from hunger, but we behave as if the opposite was the case, because an accelerating change in the environment we chose to live in brings us further and further out in evolutionary mismatch with our feelings.

All animals necessarily live in evolutionary mismatch, and the faster the changes are (in environment or in the abilities of the given species) the larger the evolutionary mismatch. Species that do not develop and always stay in the same habitat will always live ”in balance with nature” (just like humans would have done if we had been limited to a few thousand individuals and hadn't left the African savannah), but they are unlikely to survive changing conditions. The huge expansion in number of humans during the latest 100,000 years has increased the human genepoole accordingly. Together with the diversity of environments populated by humans, this has accelerated the natural evolution of the human species to a level that has probably never existed before, in particular concerning disease resistance, diet, and nervous system function. The arctic regions have probably been populated for only 15,000 years or so. Yet, eskimoes already have a larger volume to surface ratio than the rest of the human population, beneficial for survival in cold climate. Argiculture has only been widespread for 5,000 years, and yet populations with roots in argiculture (e.g. Europeans) have genetic lactose tolerence (making it possible to benefit from neutritious animal milk) and better genetic ability to digest starchy (argicultural) food. Decendents of humans that migrated away from equatorial regions 50,000 years ago developed light skinn to more efficiently form vitamin D from smaller amounts of UV light (but afro-americans didn't develop light skin yet after few hundred years in more northern regions, and duch-africaans didn't develop dark skinn yet after few hundred years in Africa). Modern humans have a much better imune system against a large number of illnesses, in particular among people with large genetic mixing in same temperate regions.

The above examples seems to indicate that genetic evolution would ease many of contempory problems or evolutionary mismatches over timescales not of hundreds or years and not of hundred of thousands of years but inbetween, of orders of thusands to tens of thousands of years. We would expect that genetic evolution would solve the obisity epedemic and stop people from smoking in a few thousand years from now. Any other animal species would solve such problems over such timescales. But none of us would find it ”natural” for us to accept 10,000 years timescales to solve obisity issues, tobaco adiction, polution issues, climate challenges, nuclear weapon armments etc. Why not? How did humans develop so radically different from any other animal on the globe? The answer is simple: *accumulated knowledge*. The issue is not how are we doing it, the issue is how did it happen? How did blind evolution manage to develop a species that understood how it happened? It happened once in our solar system, and only ”yesterday” in evolutionary or geological terms, after almost 4 billion years of blind random genetic experimenting. Is it normal? can we expect it has happened on millions of habitable exoplanets throughout our Galaxy? Why did dinosaurs not develop any kind of accumulated knowledge? Why did they not prepare for the meteorite strike? They had more than 100 million years to figure out that it would come sooner or later, and yet it came as a surprise that destroyed their entire species after more than 100 million years of superior ruling of the Earth.

We are not the first species that have been good at point 1) and 3). Sharks, dinosaurs, rats, seaguls and coli bacteria are (/were) all excellent at pont 1) and 3), yet none of them has ever transmitted a radiosignal toward another civilisation in our Galaxy, and they show no understanding of how nature function around them, other than what the genes have forced them to. Could it really be that the more complex point 2) was crucial in our success? Maybe, and in that case probably in combination with our forefathers almost mystriously silly choice of walking on two legs in spite of our anatomically constructed adaptation to four. We have developed a strong and mutual pickiness in mate selection which is unusual for most other mamales, and it might have created new possibilities for our species, and new challenges as well. The male peacock that presented the most impressive tail and scared most other males away had free access to the female participants. The roles of both sexes are simple and well defined. He is off to the next party as soon as the job of spreading his genes is done. There is no need for the peahens to impress him, and he will not participate further to the family after the first act. The gorilla strategy is essentially equally simple. For both species the genetic goal of the population is to multiply the strongest genes in the poole as much and quickly as possible in the competition with other species, as it is for us and which we have been very successful at based on the human strategy, and which during our short evolution happened to have become more complex. Psycologists holds for example that a deep masculine human voice, bulky muscles, and bushy beards all have an intimidating effect on surrounding men, and that statistics show no clear attraction to these features from human females. Therefore they are most likely the equivalent of the peacock's tail or the deer's antlers intimidating effect on surrounding males, and not the equivalent of the lions teeth (or the aspects of the peacock-tail that potentially impressed the peahens). They only assure attention from more women at the party by intemidating the other males, but this is less important for humans than the impressive tail is for a peacock, because the human strategy is aimed at a smaller number of children than the peacock's, and a larger degree of survival of each individual child. Why did it develop this way? The human male does not need to attract all the women at the party, but will have a better strategy in aiming at a few that he judges have a good ability to take care of his outcome. Therefore evolution selected those men that had an ability to spot and attract the most healty young women in their reproductive age (it is also quite uncommon for our species that there is a female reproductive limited period, this is for example not the case for chimpanzees), and he learned it by simple tricks. A high degree of symmetry in face and body is usually associated with healty genes, longlivity etc (just as is the large antles of a male deer). They are those features that even today are tested in beauty contests of all cultures, and men all over the globe are glued to the TV screen as did it concern life and death, as it originally did. There is nothing objectively more beautifull in the winner than those features men already selected for at the savannah in Africa, but our genes combined with modern technology makes us in a stage of an evolutionary mismatch, also at this point, just like the excessive suger-intake is an evolutionary mismatch.

The offspring also needs to survive, and here the evolutionary mismatch presents us for another challenge and opportunity, which the gorillas live without, because its changing tiny cultural evolution goes as slow as its genetic evolution, and therefore has almost no evolutionary mismatch. Or in other words because its evolution is determined by its genes and not by culture. For the uhr-man at the savannah, the best strategy to spread his genes would be to provide well for the woman that caried his child long enough that the child grew up to independence and hence ability to multiply his genes further. But in order to not risk all eggs in one basket in a dangerous environment, the optimum strategy would be to provide also for a few others, the amount depending on how skillful a provider he was. Fitness selected the men that was best at this double strategy, and it followed us long into modern history with the challenges it established. The human male known in this way to have had most children is Moroccan emperor Ismail the Bloodthirsty, resigning 1672 - 1727, who is said to have had 888 children, but remark that the present-day most provideable men, as for example Jeff Bezos or Bill Gates, in spite of their ability to provide a fine living for hundrets if not thousands of wifes, all have had of the order of only one wife throughout a lifetime, and quite few children, illustrating an important non-genetic shift in human evolutionary strategy.

For the male and female peas (peacock and peahen) it is the peacock's genetic excess energy to produce a large tail that determined the genetic evolution of the species. All the hens were equal, and all their genes were secured by mating with the male with the most impressive tail. The male and female evolutionary stragtegy were identical and simple as for the gorillas. For humans the optimal strategy however developed different for the male and female, because of crucial seemingly random choices our forefathers took 4 or 5 million years ago. For men the double-appearance on the stage would have become the optimal genetic strategy, the one "the selfish gene" would chose, but for women the optimum strategy to spread her genes became to include the ability to select and attract men that would be skillful and careful providers over long timescales and willing to share with her and her child. The most optimal combination of resourceful and powerful men such as Ismail the Bloodthirsty combined with care, or in modern terms e.g. Bill Gates or Donald Trump, depending on the eyes that judge. Women that developed the ability to attracht such men would have larger chance for themselves and their children to survive, and evolution would therfore automatically select for these genetic characteristics, and we will still inherit them from our parents just simply because most humans inherited those, because they survived best in the ancient genetic competition. For a woman on the savannah the optimal genetic strategy would be to attract such a man and then scare away ("be jealous toward") other women that potentially would be too closely associated to the man of their select, in order to assure that not too much of the providing capacity would go off-house. For a man the optimum strategy would be to assure that no other genes than his took resources from his providing capacity. Agresiveness would therefore become strongest among men (just as for deers, peacocks, and gorillas), and men's jealousy would lean toward intolerence toward too much in-house visits, while to much out of the house attention would dominate the female agenda. We may recognize it among other evolutionary mismatches we have to struggle with because of our fast and successful evolution. This mutual importance of picking the optimal partner (although with different understanding and focus on what is meant by ”optimal”) urged for a mutual pickiness in human mate selection that is not normal for mamales and which strongly has shaped our cultural and technological evolution, and therefore in the end also our ability to send a radiosignal looking for peers in the vasteness of cosmos, and to one day soon colonize other worlds and maybe develop into a Galactic ”super-civilization”.

Of course evolutionary strategies are not conscious choices. It would be a good enough rational selfish-gene strategy for a man to become a fertility doctor and exchange all the delivered sperm with his own (as actually happened in one known sad example) and a good enough rational selfish-gene strategy for a woman to make the husband think he was the father of her children, but the selfish geene is exactly the question of our instincts and feelings and not cool calculations (so genetic studies consistently show that more than 99% of men that think they are fathers of their children actually also are so). The development in our mating strategy has been essential for our ability as a species (and therefore our survival and multiplication). It has assured a relatively peaceful co-existence and a possibility to form much larger groups of collaborating individuals than any other animals are able to organize. Take just a moment to think about the fact that you every day encounter more other human beings unknown to you than any hunter-gather on the savannah would have met during his entire lifetime. Yet you do not on a daily basis get into a fight with the other passengers in the bus or fellow shoppers in the supermarked, even though you have to share resources with all of them and they potentially could be compatitors of the spreading of your genes. How has it developed this way? Is it something obvious that will happen to all organisms at some state of evolution, and hence something we will find on all habitable exoplanets throughout the Galaxy? Is it just a question of time before polar bears will also build supermarkeds with pre-cooked fish, or gorillas opening all sort of candy stores in the jungle and the alpha-males stop fighting one another? Or was it something unique to humans, and what is it then that took only us along this path of evolution?

For the human females it became important that ”the male stayed after the party”, and the social appearance became more important for genetic succes in human societies than in chimp societies. Human females therefore have a larger challenge and interest toward picking ”the best” male than a female chimp has, so she needed to develop alertness toward other features than strength and deep voice, and she needed to develop an imaginative awareness of her appearnace in the eyes of the males in order to compete efficiently in the reproductive succes, to a degree that was irellevant for female chimps and peahens. It would be unusual for a modern human female to turn up at a party in old warn-out clothes, unwashed, with un-brushed teeth and greasy hair. For a peahen it would be irellevant. These differences are likely to have their roots in our upright walking, and they have had enormous implication for our evolution, and probably even for our chances for surviving as a species long into the future. As described above, the upright walking and the large brain, made it necessary for evolution to find a compromise between the size of women’s pelvis and the duration of the pregnancy. Too large pelvis would make it impossibly to walk on two legs, too small would make it impossible for a large-brained child to pass through the birth chanel. A full blown pregnancy lasting untill the child was as developed and able to manage as well as new-borns of most other animals would grow the head to be too big to pass the birth chanel. A too short pregnancy, allowing as easy a birth as for most other animals, would in most cases lead to the dead of the child within the first few weeks after birth. Evolution obviously favoured those individuals with the length of pregnancy that was optimal for the combined survival chance for the mother and the child, which is a relatively short pregnancy of 9 months and a functional but volunerably and not fully developed new-born child. As a consequence, the human mother and child are more dependent on a prolonged period of care than mothers and children of other species. Human females that are attracted to, and able to pick, males that are more peaceful and have a larger tendency of carying will therefore survive better, which will create an evolutionary pressure toward a more carying and peacefull species. In some sense one could say that it was the anatomically need for the women to be more picky in the choice of males with new social features that accelerated an evolutionary pressure on more peaceful and carying men, and hence a more peaceful society (than e.g. that of chimpanzes or gorillas or earlier hominid species). More peacefull co-existence will in turn make it possible to form larger groups, which opened up for a new evolutionary development relatively unique to humans: the ability to form larger societies without always fighting, and therefore the creation of excess resources to do other things than just surviving. It is this feature that has brought us onto the doorstep of colonizing the Galaxy – as the first species to do so ever – according to the paradox so famously expressed by Enrico Fermi. We can for sure become even more peaceful in the future, to an enormous further benefit for all of us, but it was the mysterious decision for upright walking that began the journey, and it had much wider consequences than a better view over the savanah. Is it cosmically ”normal”?

1.9 The colonisation of Mars

When the Apollo project was on its hights in the mid and late 1960es, about 400,000 people were associated to the project. The missions were meant to continue with manned flyby of Venus and eventually manned missions to land on Mars. However, the cost of the Vietnam war, the following oil crise, and the associated change in economic and political climate, brought the project to an abrubt end with Apollo 17 in 1972. The plan for manned missions to Mars was not taken up again before the Bush administration's Space Exploration Initiative (SEI) in 1989, but the following 3 months study reached the conclusion that the project would cost 450 billion\$, approximately 5 times the total cost of the Apollo project, or 10% of the cost of the second invasion of Irak, and the project was immediately abandoned. However, the cost estimate was based on Verner Von Braun's ”science fiction like” vision about how to fly to Mars, beginning with establishing an Earth orbiting space port (and assembling space station) feeded with construction materials from a factory on the Moon, eventually leading to building and launching a huge interplanetary cruiser from the space port to Mars. The folowing two decades saw more private initiatives entering the development of a consept of manned missions to Mars at a much lower price than calculated in the SEI report, together with a return to a stronger international competition. The Chinese space program had already in 2013 its Chang'e 3 rover driving over the lunar landscape, the Indian, Japaneese, and ESA space programs have in total had several satellites orbiting and impacting the Moon and returning valuable new data. Many countries and private organisations have the ambitions to not only explore the Moon further, but also Mars. In 2014 India was the fifth nation (after USA, Sovjet Union, ESA and Japan) to attempt a mission to Mars, and the first to succeed in their first attempt, and even at the lowest budget seen so far. The chinese space organisation aim at its first sample return mission in 2030 and manned missions in 2036. Arab Emirates plan for a Mars orbiter by 2020, India to bring a Mars lander and rover in 2022, Japan to bring an orbiter in 2020 and 2024, private USA organisation Space X to land people on Mars by late 2020ies and Dutch private Mars One to establish a human colony during early 2030ies. Russia envisions its first manned landing on Mars by 2030, while NASA is more cautios with expressing that its first astronauts will walk on Mars some time during the 2030ies.

In 2007 Google supported the X-Prize initiative with 30 million \$ in reward to the first private organisation that before 31st of March 2018 would be able to land an unmanned spaceship on the Moon, move around at least 500 m in the terran, and transmit back to Earth viedos and photos from the lunar surface. By the end of 2017 there were 5 organisations still competing to winn the prize. All of them, however, stating that they will not be able to meet the deadline. The X-Prize initiative will continue in one or another form without Google also after 31/3/2018, and whichever way the competition will end, it has been an inspiring eyeopener, contributing to promoting private iniviatives and organisations to do what only nations were previously able to do in space, and contributed to accelerate the development that will eventually make most of the people living on Earth today see humanity transform into a multiplanetary species.

Present day proposals have pointed out that launching a rocket from the surface of the Earth directly to Mars, without the need for a space station and a lunar base, is less energy consuming than launching a spaceship to the Moon, and can be done for less than a tenth of the price of the SEI project. Already in 1925 the German mathematician Walter Hohmann showed that the cheapest flight to Mars (in terms of energy consumption, and hence also in terms of cost) was along what we now call the Hohmann transfer orbit, which is one half of an elipse with perihel at Earth's orbit and aphel at Mars' orbit. Such a transfer would launch the spaceship in the direction of Earth's own movement and reach Mars parallel with Mars' orbit. The beginning velocity of the spaceship would in this way be the departure velocity from Earth delivered by the rocket, which is typically around 5 km/s, plus the huge orbital velocity of Earth itself (30 km/s), which is added ”for free” in the right direction of the flight. With the aphel of the elipse at Mars orbit it is assured that the spaceship moves parallel with Mars on arrival, thereby having the smallest possible relative speed and allowing simple capture by Mars' gravitational field. Apollo flew the 400,000 km to the Moon in 3 days with an average speed of 1.5 km/s. The same average speed along a Hohmann orbit to Mars would benefit from the initial 30 km/s added from Earth's movement, therefore going basically 20 times faster and do the 1,000 times longer flight distance to Mars in only 3x1000/20 = 150 days. Remark that the lunar flight cannot take advantage of the Earth's movement, because the flight is internal within the Earth-Moon system. Launching the rocket from a base near equator can add the Earth's rotational speed to the initial speed, but this is only 0.4 km/s, and is mainly important for launch into Earth orbit or to the Moon.

1.9.1 the fuel and the flight

A more detailed calculation, taking into account the decelleration along the Hohmann transit etc, show the needed departure velocity to be 3.3 km/s, the transit time to Mars to be 254 days, and the configuration being such that if the spaceship miss target (or the landing on Mars needs to be abandoned for some reason) the Hohmann orbit would automatically bring the spaceship back to rendevous with Earth after two full orbits, or approximately 3 years of traveling. Such an orbit would make the spaceship rendevou with Mars at very low relative velocity, and be optimal for cargo delivering to the Mars habitates. Using a bit more fuel to get a departure velocity of 5.1 km/s into a near Hohmann transfer with aphel a bit beyond Mars could, however, bring the astronauts to Mars in only 180 days (hence reduce the radiation dose and the time in zero gravity), and bring the return time in case of failure of landing down to a total travel time of 2 years. This orbit would make the spacecraft cross Mars' orbit almost parallel with Mars' movement at a velocity of 21 km/s, which compared with Mars velocity of 24 km/s would result in a comfortable relative velocity of 3 km/s. If Apollo had approached the Moon with 3 km/s it would have lost target or crash landed; it just simply wouldn't have been able to break enough for a soft landing. The larger gravity of Mars, and the possibility of aerobreaking in Mars' thinn atmosphere, however, makes 3 km/s slow enough for secure landing. This transfer is therefore what many people today call "the direct flight to Mars", and it might be the preferred transfer for a manned mission. Because of the very little amount of fuel needed for the landing (which could eventually be almost zero if parachutes and aerobreaking are optimized) and the possibility of taking advantage of Earth's orbital velocity in getting the space ship entering its orbit at high speed, the funy paradox arises that it is actually cheaper (fuel wise) to sent a space probe to Mars than to the Moon. A Saturn V rocket would have been able to send 25 tons to the surface of Mars along a fast track Hohmann orbit allready in 1972.

Aiming at the spaceship going back to Earth along the same fast Hohmann transfer will require the crew to wait 550 days on Mars, in order to again fly along an elipse reaching its perihel at the same time the Earth is there. The total time of such a mission would therefore be a trip of 2x180+550 = 910 days = 2.5 years, leaving amble time to explore the surface of Mars during the 550 days = 1.5 years stay. Such a trip would expose the astronauts to a total radiation dose of 50 rem, which would give an astronaut in the mid-thirties 1% extra risk of dying from cancer later in life due to the mission. The radiation exposure at the bottom of the martian atmosphere is 4 times smaller than in free space, giving the 550 days stay at Mars a dose equal to 140 days travel time, or one month each way if it is assumed that the astronauts work and sleep inside fully protected habitates half of their time on Mars. In terms of radiation risk, a faster trip with less or close to zero exploration time at the martian surface is therefore only an advantage if it can reduce the time in space with more than a month each way. The 450 billion$ original project aimed at the shortest possible total mission time (640 days) in an attempt to lower the radiation risk. However, even though it reduced the total time of the trip, it increased the travel time in interplanetary space from 360 to 600 days, with the result that the total radiation exposure would have been substantially higher, not lower, than along a Hohmann orbit, on top of having costed 10 times more, have given the astronauts only 2 weeks of exploration time on the martian surface, instead of 550 days, and more than 600 days in zero gravity space travel as opposed to 360 days in "the Mars direct" journey. For all these reasons modern designs of a Mars exploration expedition has a minimum of 2.5 years of duration, with most of the time at the Martian surface itself. It is in some sense tragic to think that the needed 5.1 km/s departure velocity for a Mars-direct flight could have been achived from ground already in the Apollo area with one of the complete and ready Saturn V launch vehicle that was build but abandoned and never used because of the abrupt canceling of the project.

For optimal use of the ability to land 25 tons on the surface of Mars per flight, one would like that only a minimum of this cargo is fuel to bring the astronauts back to Earth. For this reason various consepts of in situ fuel production on Mars have been investigated, as well as the posibility of maximum in situ production of oxygen and water for the stay. Cynical souls have concluded that without such in-situ fuel production, it is 10 times cheaper to send astronauts on a one-way ticket than on a return ticket. Two promising possibilities are production of methane and methanol, followed by burning of methane with oxygen for rocket fuel and methanol with oxygen for car transportation on Mars' surface. Hydrogen (for the methane and methanol production) might be the most difficult component to produce on Mars, and in the beginning it may therefore be necessary to bring it from Earth. Since hydrogen is, however, only 5% of the total weight of the fuel, reducing the fuel transport to be only liquid hydrogen would be a gain of a factor 20 in weight compared to bringing it all. Mars' atmosphere is 95% CO_2, so the needed carbon and oxygen is conviniently available, provided sufficiently low energy consuming and efficient machinery can be constructed to take advantage of the existing atmospheric gas.

If one fills a container with activated carbon or zeolite, it will as a "sponge" by itself condense 20% its weight of CO_2 out of the atmospheres at the martian night time temperatures (-90°C). Next day one can let it evaporate off into a container when the temperature on Mars increases to say +10°C. By compressing the CO_2 in the container to 7 bar, it will become liquid. Left over dust will then sediment to the bottom, and nitrogen and argon will evaporate. The nitrogen/argon has essential use as buffer gas in breating and combustion. The purified CO_2 can now be burned with hydrogen to give the needed methane and oxygen, through the exothermic (i.e. giving energy from the reaction and therefore running by itself) Sabatier reaction,

$$CO_2 + 4H_2 \rightarrow CH_4 + 2H_2O \tag{1.10}$$

Water is then split by electrolysis reaction as

$$2H_2O \rightarrow 2H_2 + O_2 \tag{1.11}$$

and the H_2 is recycled back into the Sabatier reaction. We see that the outcome is that

$$CO_2 + 2H_2 \rightarrow CH_4 + O_2 \tag{1.12}$$

For most purposes one would want a higher ratio of oxygen to methane than the 1:1 ratio reaction 1.12 will provide, both for a more efficient burning and for extra oxygen for breating. One could in that case split some of the atmospheric CO_2 directly to produce extra oxygen and CO. The produced CO could in the beginning be vented into the atmosphere (as a polutant that will find its reactions with surface minerals), but in long term CO would be a very useful gas in itself for the first martian industrial production schemes. In long term one could also produe extra oxygen for breating from greenhouse food production. As on Earth, the carbon bound in the food would then react with the oxygen we breath, and finally vent as CO_2 back into the martian atmosphere. One could of course also bring extra water in to react in 1.11, either by freezing water out of the atmosphere (the thinn martian atmosphere is saturated in water from evaporation of the polar caps) or extracting water directly from underground ice resovoirs. This way of producing the needed oxygen would also add free hydrogen that could reduce or eliminate the amount of liquid hydrogen needed to be brought from Earth. The final burning of methane (or methanol) with oxygen for rockets or surface vehichles would leave no other reactants than the original water and CO_2 going back into the systm.

1.9.2 Can human beings survive the radiation on Mars?

A challenge for the first expeditions to Mars will be how to protect the astronauts against the higher level of radiation they will be exposed to on Mars and during the flight between the Earth and Mars. Some people have been of the opinion that the radiation alone would make colonization of Mars impossible, while others are of the opinion that apropriate precautions will reduce the radiation issue to a minor and solvable problem. The astronauts and colonists will for sure be exposed to a higher level of cosmic rays and solar wind particles than on Earth. Exactly how much higher, how to best shield from it, and how much it actually affects the human body, are quite uncertain. On Earth, the atmosphere and the magnetic field shilds us from both types of radiation.

Some basics about radioactivity and ionizing radiation – rem, rad, sievers, gray, curie, becquerel

Historically radioactivity has been measured in a number of different units based on various issues related to the radioactive decay. At the original discussions of radioactivity the central issue was how much radiation was created by various materials, and during a meeting in 1910, when the knowledge of the nature of radioactivity was still limited, it was decided to define the basic unit of the radioactive *activity* as the radiation coming from 1 g of radium (^{266}Ra), and call it *curie* (with the symbol Ci), and recommended to use the symbol A to describe activity in mathematical formulas. Later the activity was specified more precisely as 1 Ci = 3.7×10^{10} decays per second. In 1975 it was decided that it would be most desireable to be able to relate the measure of activity directly to the SI system, and the most logical unit of (radio)activity would then be 1 decay per second, which was called a *becquerel* (with the symbol Bq). A less logical SI unit for activity had already been introduced in 1946 with the unit name a rutherford (symbol Rd) defined as 1 million decays per second (i.e., today 1 Rd = 10^6 Bq). Although the units curie and rutherford are still used, activity of the radioactive process is today officially, and most logically, measured in units of 1 Bq = 1/(3.7×10^{10}) Ci = 2.7×10^{-11} Ci.

The natural continous lifelong activity level of a human body is approximately 7400 decays per second, i.e. 7400 Bq or 2×10^{-7} Ci, due to naturally occuring radioactive elements in our body tissue, dominated by $\sim$0.02 grams of radioactive potasium (^{40}K) arising from an average total potasium abundance in our body construction of 2.5g per kg human body. Also minor species such as for example ^{14}C contribute to the continous radioactivity of our body. On top of this comes a temporary activity from the food that pass through our body. For example a 150g banana typically contain 0.5g of potasium and 15g of carbon, giving rise to a total of 20 Bq of which 75% comes from the ^{40}K and 25% from ^{14}C. I find it fascinating to think that when I eat a banana I will be exposed to 5 radioactive decays per second in my body, which eventually are caused by supernovae far away in the Galaxy that are responsible for the continous production of ^{14}C in Earth's atmosphere that is part of nature's construction of neutriceous food I can intake to maintain my body in order to be me. Because it has always been like this, evolution has developed a healty cell repair system in our body that is only active if it is continously exposed to a small amount of radioactivity. One would obviously like to know how long time it would take evolution to develop a repair system for higher forms of organisms that could sustain a higher amount of radiation than our own body is now tuned to. Lower life forms seems to be able to adopt to higher radiation levels rather quickly, such as has made it possibly for radioactivity tolerant becteria to now live inside the core of nuclear power stations. One could speculate whether we could help such a development for higher life forms too, and whether the gain of knowledge we will get from the first Martian colonists in this respect could help not only them to better survive the Martina radiation environment, but also humans back on Earth to better fight for example radiation caused cancer and other issues related to more efficient cell repair and longer lifespan. For illustrative purposes it has been suggested to introduce the popular and approximate unit BED (a banana equivalent dose). People living 10 USmiles (16 km) from the Three Mile Island reactor, received during 1979 (in connection with the largest nuclear reactor accident having occured in the western world) 800 BED, or the equivalent of having stood right next to the shelf of bananas in a large supermarked during the full year. A CT scanner delivers 70,000 BED, and the total natural background radiation is $\sim$100 BED.

For a given kind of radiation, the activity level obviously plays an important role for its potential effect on a living organism. However, the effects also depends on which kind of particles are emitted, as well as how their energy is absorbed by our body tissue. The activity (measured in units of Ci or Bq) is therefore only part of the measure of the danger of being exposed to radioactivity, and also radiation from other sources than radioactive decays can affect the human tissue. Therefore it is more relevant to talk about *ionizing radiation* than radioacvtivity, i.e. particles or photons with high enough energy to ionize a given material, and to measure the amount of energy associated with the absorption of this radiation. This is called the dose of ionizing radiation, and it is measured in the units called gray, sievert, rad and rem.

Gray (Gy) is a derived SI unit, defined as the amount of energy absorbed per unit absorbing material, such that 1 Gy = 1 J/kg. The corresponding unit in the cgs system is called rad. 1 rad = 1 erg/g is hence the amount of energy in units of erg that is absorbed per g of absorbing material. Gray and rad are therefore neither a measure of activity nor a measure of the flux (i.e. joule/m^2 or erg/cm^2) of radiation hitting the material, but a measaure of the absorbed energy such a radiation impose to the absorbing material. 1 rad of ionizing radiation is therfore the amount of energy absorbed per gram of radiated material from particles or photons that have high enough energy to ionize atoms in the radiated material when absorbed.

Such radiation is obviously damaging for our tissue, but the amount of gray or ram are not uniquely describing the radiative damage in our boddy, since the damage also depends on such details as how deep the radiation penetrates into our body and how much energy is released per particle or photon of the radiation. Therefore each type of radiation is empirically associated with a so-called quality factor, described by the dimensionless number Q (also some times called the radiation weight factor W_{r} or the relative biological effectiveness RBE). The damage caused to our body from different types of ionizing radiation is normalized to the damage that is caused by X-ray radiation (in many languages, including Danish, called Röngten radiation), and is therefore measured in the unit rem – "röngten equivalent man". 1 rem is therefore the absorbed energy per g of absorbing material that would cause the equivalent damage to our body as 1 rad of X-ray radiation. The corresponding SI unit is called sievert (with the symbol Sv). Remark that both sivert and gray (and rem and rad) are measured in energy per absorbing mass unit, but they describe two very different physical things. Gy and ram describe how much energy is physically absorbed, while Sv and rem describe how much damage such energy absorption cause to our body. Eventually one would of course rather than Sv and rem introduce a quantity defined in units of the probablilty of dying from the radiation or the probability of developing fatal cancer within a certain number of years etc, but our knowledge about radiation effects on the human body is not yet there, so for the moment the best we can do is to use emperical scaling laws that express how dangerous absorbed energy from various kinds of radiation is to our health. For this purpose it is believed that ionizing proton radiation is twice as dangerous as X-rays. Since Q=1 per definition for X-rays, empirically Q=2 for protons. Similarly, it is believed that radiation with alpha particles (the nucleus of He) has Q=20, neutrons somewhere inbetween depending on their energy, while HZE (high (H) atomic number (Z) and energy (E)) ions have 20 or potentially more, and electrons (β-radiation), muons and γ-rays all have Q=1.

The energy physically absorbed is called the (physical) dose and is often expressed with the symbol D, while the damage to our tissue by absorbing a certain amount of energy per kg of our body weight is called the equivalent dose and often expressed by the letter H, such that

$$H[Sv] = Q \times D[Gy] \tag{1.13}$$

Equation 1.13 indicates that there is a linear relation between the physical radiation dose and the damage it represents to our tissue, represented for example as the risk of developing cancer due to the radiation. Obviously the reality is more complicated. For very low levels of radiation there seems to be a treshold below which no damage occures. For very high levels of radiation the damage is obviously independent of the dose (the patient dies in any case). The range of radiation doses where Eq. 1.13 is valid is called the linear non-treshold regiem. If one adds the assumption that the radiation damage has equal probability to manifest itself at any time after the exposed dose, then we will say that Eq. 1.13 express a linear stochastic non-treshold relation and we can use it for estimating the cancer risk per year as function of physical radiation dose. Nevertheless, we also know that an *acute dose* (i.e. a dose of radiation coming during a short time interval) is more damaging than a similar *chronic dose* (i.e. the same amount of sieverts, but distributed over long time, such that the cell repair system can manage to repair the cells along the process). Therefore Eq. 1.13 applies best to chronic radiation of relatively modest doses, and this is exactly what we expect the astronauts on Mars, and in particular the settlers, will be exposed to, which gives us some reason for confidence that we may be able to estimate the risk the Mars colony will be exposed to due to radiation.

1.9.3 how much radiation will the Mars astronauts be exposed to?

All previous astronauts have caried a personal dose meter that has measured how much radiation they have accumulated during the flight, and assesments of radiation dose and damage is published with regular interval both by national and international radiation monitoring organisations. In spite of this, it is not easy to get a clear picture of how much radiation an astronaut would be exposed to on a trip to Mars, and the oppinions are widely divergent. One of the reasons is that the many astronauts that have been on ISS are partly inside the Earth's protecting magnetic field. Assesment of the risk the radiation imposes is obviously even more uncertain; only few astronauts have been in space long enough to compare with the length of a Mars mission, and only very few astronauts have been in space long enough ago to asses the long term effects. Statistics from radiation exposure from human sources are often acute dose exposure (Hiroshima, Nagasaki, Chernobyl, CT scanning, etc) and/or related to ionizing radiation of a different character than what the astronauts will meet in space and on the surface of Mars.

One of the few relevant and most cited studies today comes from the Radiation Assesment Detector (RAD) onboard the Curiosity rover that landed on Mars in 2012. It measured an average ionizing radiation dose of 0.48 mGy/day during the cruise from Earth to Mars, and 0.21 mGy/day on the Martian surface. It was estimated that the quality factor of the radiation in space was 3.8 and at the martian surface was 3.0. Taken straight on, a 2×180 days cruise back and forth plus 550 days on the surface would then correspond to 2 x 180 x 0.48 x 3.8 + 550 x 0.21 x 3.0 = 650 mSv + 345 mSv = 1000 mSv. This is exactly the ESA (and most other space agencies) so called astronaut career standard limit (i.e. maximum allowed lifetime equivalent dose exposure).

The above numbers supports the idea that it would have been safe and possible to fly Apollo 18 on a direct trip to Mars. The lifting power was in place and the astronauts would not have surpassed the astronaut career limit of radiation. It also indicates that a relatively unshielded flight is the most problematic part of the trip (from a radiation point of wiev), and that it therefore is more important to get quickly to Mars than to limit the exploration time on the surface, and therefore that the direct flight to Mars would have been more safe than the minimal mission time aimed for in the expensive SEI. Already the thinn existing martian atmosphere reduce the effective dose rate with a factor of 3 compared to a cruse in a 10g/cm^2 shielded spaceship, which is estimated to have been the effective shielding of the RAD during cruise, and which is only slightly above the shielding value in the Apollo capsules when they flew to the Moon.

The above number would indicate that longer trips than a single scientific mission to Mars would be impossible. One should, however, note that the RAD was basically unshielded during operation on the martian surface, and to large extend during the flight too, and therefore a simple sumation of the RAD doses would give an unrealistically high estimate of the dose that astronauts would be exposed to. It was estimated that the (accidential) shielding due to surrounding material of the packed configuarion during the cruise has been $\sim$10g/cm^2. A reasonable construction of a cruise ship for human transportation to Mars would certainly aim to shield the passengers better than the accidential packing of the RAD. It would be reasonable to envision a shielding of no less than 20 g/cm^2 in regular space and 40 g/cm^2 in the sleeping regions (which would also serve as shelter regions during solar storms). Lets for a moment assume that a 20g/cm^2 shielding in the spaceship offers the same reduction in effective dose as the 20g/cm^2 shielding the martian atmosphere is offering, and let us assume a 40g/cm^2 shielding offers a further reduction of an additional factor of 3 over the 20g/cm^2 shielding, and that the pasengers are half of the time in regular space and half of the time in the sleeping regions. Then the total dose during a 180 days cruse to Mars would be limited to 90 x 0.6 mSv + 90 x 0.2 mSv = 70 mSv each way (which would be the total cruise for a colonisation). On the surface one may envision the colonists being indoor 16 hours a day in a completely radiation shielded area, and outdoor 8 hours a day. Adding a 10 g/cm^2 insulation most of the time when operating outdoor (e.g. a 4 mm aluminium roof of a car) would reduce the outdoor equivalent dose rate to 0.4 mSv/day (assuming the same linear relation as above). A 550 days expedition on the surface of Mars would then reduce to (2/3) x 550 x 0 + (1/3) x 550 x 0.4 = 70 mSv. Adding the cruise back and forth would result in a total radiation dose of 210 mSv for the exploration mission (as opposed to the 1000 mSv estimated above). A 50 years colonisation mission would impose a 45 mSv per year (following the above estimate) x 50 years = 2,200 mSv, plus 70 mSv dose during the one-way flight, totalling a lifetime dose of 2.3 Sv.

1.9.4 a simple working model for the radiation dose through the atmosphere

To get a better estimate of the radiation dose expectancy and how to best control it, one could try to model the shielding as function of column mass (e.g. g per cm^2) in various materials for various kinds of ionizing radiations, including SEP (solar energetic particles) and GCR (galactic cosmic rays), penerating an atmosphere dominated by CO_2 (as on Mars) and N_2 + O_2 (as on Earth) and scale it to measured dose values.

SEP comes basically in two forms, soft SEPs dominated by particles with energies below $\sim$150 MeV and hard SEP dominated by a more high-energy particle distribution. Galactic cosmic rays (GCR) is a stream of high energy particles from supernova explosions throughout the Galaxy, and is dominated by $\sim$90% protons, 10% helium and a small amount of heavy nuclei and electrons. Their flux is a factor $\sim$2 higher during solar minimum than during solar maximum, due to shielding of the inner solar system by the solar magnetic field. GCR have a maximum flux in the GeV range, and therefore penetrates deep into the atmosphere of Mars (and Earth) and even a couple of meters into the martian regolith. The flux at the sea level on Earth is $\sim$10,000 particles per m^2 per second. The soft SEP is basically completely blocked by even the thinn martian atmosphere, while the hard SEP (associated with the most violent solar eruptions) partly penetrates the Martian atmosphere (and to a smaler extend occasionally the Earth's atmosphere too).

Table 1.1 lists the measured equivalent dose rates at various hights in Earth's atmosphere, as well as the two RAD measurements from Mars. If the radiation had been composed of one kind of particles at one energy, the absorption would have followed a simple exponential law. This is what is called Beer-Lamberts law in chemistry or Lamberts law in astronomy and optics, which states that when light shines through a gas the same percentage of the light is always absorbed when it pass through the same amount of gas, independent on the intensity of the light. Translated into a dose H of radiation it would mean that

$$H(h)/H_0 = e^{-h/\tau} \tag{1.14}$$

stating that the radiation dose, $H(h)$, after a shielding h relative to the unshielded dose, H_0, is an exponentially decreasing function of shielding with a "shielding constant" τ (which is what in radioactivity is related to the half-life and in optics and astronomy is the definition of the optical depth). The shielding h is here the number of g/cm^2 of shielding material. If we as a first approximation assume that it doesn't matter what the shielding material is made of and what kind of particles at which energies the ionizing radiation consiste of (SEP, GCR, etc; this is what we in astronomy would call a grey atmosphere), then a fit to the measured doses at given hights on Earth (or different shieldings on Mars measured by RAD) could give us the value of τ, and hence make us estimate the dose at any choice of shielding. Note that for the Earth h is not the physical height, but the number of grams of absorbing atmosphere per cm^2 (i.e. the gas-column mass) above a given location (i.e. above a given physical height r in the atmosphere). This column height is conviniently calculated as function of physical height by noticing that the density of the atmosphere drops exponentially with the scale height, $H_{\rm scl}$, such that also the column density, $h(r)$, drops exponentially with height, r, above the sea level, r_0,

$$h(r)/h(r_0) = e^{-r/H_{scl}} \tag{1.15}$$

For the Earth, the scale height is 5.6 km, and therefore the column mass (i.e. the shielding) from cosmic radiation in the city of La Paz at the hight of 3.9 km above sea level is 100 x e$^{-3.9/5.6}$ = 49.8% of the shielding at sea level. In other words, people in La Paz have half of the mass of the Earth's atmosphere above them and half of it below them and are therefore shielded e times less from cosmic radiation than people living at sea level are.

place	meassured dose ratio	column difference	derived τ
Denver to sea	2.3	250	300
La Paz to Denver	3.4	250	205
La Paz to sea	7.8	500	245
RAD surface to cruise	3.0	10	9.1

Table 1.1. Measured ratio of equivalent dose rates (column 2) on Earth and on Mars when moving from one place to another mentioned in column 1, estimated differnces in shielding column mass (column 3) and derived value from an exponential model (column 4).

Table 1.1 list the relative column mass (normalized to 1 kg/cm^2, corresponding to 1 Bar, at sea level) calculated in this way, and the corresponding estimated τ value calculated from Eq. 1.14 and measured values of $H(h)/H_0$. For Mars $H(h)$ is known for two values of h, namely h=10g/cm^2 (aluminum shielding of RAD) and h=20g/cm^2 (CO_2 atmosphere) and the estimated τ is calculated from the measured dose ratio 0.6/1.8 = $e^{-10/\tau}$ => τ = 9.1 g/cm^2. Obviously the τ value is very different for the two planets, based on the sparse and inhomogenous measurements on Mars, reflecting both that different kinds of radiation reach the bottom of the martian atmosphere compared to the Earth, and reflecting also the fact that the two Mars values represent respectively shielding from (mainly) aluminium and CO_2 gas which is probably not directly comparable. The bottom line is that we should expect at least as much (and most likely much more) reduction in surface radiation on Mars as on Earth as function of added shielding mass to the naturally occuring present atmosphere, because the τ value representing the Mars atmosphere in its present condition and with the present level of SEP and GCR, is much smaller than that of the Earth and the kind of SEP and GCR that penetrates the Earth's N_2-O_2 atmosphere.

The GCR effective dose at sea level on Earth is ~0.0007 mSv/day and in the crusing hights of commercial airplanes (~9 km, corresponding to a hight where 20% of Earth's atmosphere is above and 80% is below) is ~0.1 mSv/day.

1.9.5 How damaging is the radiation dose of a trip to Mars?

It is often assumed that a dose of 1 Sv imposed within the non-treshold linear stochastic regiem implies a 5% added risk of radiation induced cancer during a lifetime. If this is correct, one can use Eq. 1.13 to estimate the risk associated with other travel and shielding scenaria. For example the 2.6 Sv estimated above for a 50 years living on Mars, woulod add 2.6 x 5 = 13% risk of developing cancer during a lifetime, which is similar to the added lung cancer risk from smoking. The corresponding numbers for a scientific expedition was estimated above to be 210 mSv, making the expedition well within the 1 Sv accepted astronaut career limit, or adding 1% cancer risk. In the often cited book "The case for Mars" by Robert Zubrin, he reach the more optimistic number of added cancer risk by setlers on Mars given in Tab. 1.2, where he also gives the estimated distribution between the different cancer forms. His estimates are the risk of developing cancer within 30 years for a female astronaut landing on Mars at age 30. The main reason that the table is gender dependent is that men have almost zero risk of developing radiation induced breast cancer, and therefore have a correspondingly smaller risk of developing cancer due to a stay on Mars than women in otherwise similar conditions. The risk of cancer development (within a certain number of years) is larger if the traveler settle on Mars at an earlier age (in particular if under 10 years of age because children in general are much more susceptible to radiation damage due to the faster cell growth), and correspondingly smaller if setteling at a higher age (for the same reason).

The conclusion from the above rough estimates seems to be that great care has to be taken to develop optimal shielding, in particular for the travel if the goal is a scientific exploration mission and in particular for the living quarters and the transport veihecles if the goal it colonisation. But it also shows that radiation is not a show stopper for neither scientific manned expeditions nor for regular colonisation. It is likely that we will not get much improved estimates of the added radiation induced cancer risk connected to Mars colonization and exploration before humans in statistically significant numbers have lived on Mars for an extensive period of time. We just simply lack knowledge of the effect of chronic radiation of the level and type the colonists will be exposed to on Mars, and it is not really obvious that a linear extrapolation into the regiem is meaningful, in particular considering our limited knowledge of the distribution of types of ionizing radiation on the martian surface.

Our knowledge of radiation in the dose range the martian colonists will be exposed to comes mainly from studies of the victims of the Hiroshima and Nagasaki asults and from the the Chernobyl accident. In both cases, however, the victims received acute doses, which is known to have very different effects on the body than chronic doses of the same size. In particular we have no knowledge about how the colonisation doses on Mars will affect our repair mecanism. We only know that a lower radiation dose makes the repair mechanism less efficient. It is probably more likely that a long term modestly increased radiation level will increase the activity of our cell repair system than reducing it, and we could even hope that the knowledge the coming martians will obtain about how the human cell repair mechanism is affected by modestly increased chronic radiation levels will contribute not only to their own possibility to live healtier and longer on Mars, but also contribute to improve our understanding of the possibilities we have to fight cancer here back on Earth.

type of cancer	percentage fatal
leukemia	0.30%
breast	0.45%
lung	0.40%
stomac, colon, etc	0.30%
bone	0.06%
all other	0.30%
total	1.81%

Table 1.2. Estimated accumulated risk of getting cancer within 30 years after being exposed to 100 rem chronic radiation dose.

1.9.6 Can the Martians claim independence; will it be beneficials for the Earthlings?

In order for the colonists to colonize Mars, i.e. form a new and independent society, they must be able to become self-sustainable, or produce material, ideas and/or services they can trade with Earth. The first aim for a colony far from Earth would be to produce in situ the material that will be too heavy to transport from Earth. This involves to produce their own food and their own infrastructure, including houses, vehicles, and fuel. In all these respects Mars is very hospitable. The ground is suitable for growing a large variety of plants and corps, and maybe most important: Mars has plenty of easily available of the most important element – carbon. The Moon in crontrast has basically none, and even the Earth is substantially depleted in carbon relative to its cosmic abundance. Carbon is per definition essential in all organic material, including fuel and food. It can be developed into structures that are lighter than iron and hard as steel (as is well known in modern leight weight substitutes for steel in cars and airplanes), and it forms the backboone of all types of plastic, from thinn transperant foils, to heavy duty car and construction parts, kitchen plates, plastic bags, lubricates, furnitures, clothes, milk bottles, etc. Mars' atmospheric mass is 25×10^{15} kg, compared to Earth's 5×10^{18} kg. Mars atmosphere consist of 95% CO_2 in contrast to Earth's 0.04%, so that the total CO_2 mass in Mars' and Earth's atmospheres are 2.5×10^{16} and 2.0×10^{15} kg, respectively. Even after adding the terrestrial methane and other carbon bearing trace gasses, there is still a total of 10 times more carbon in Mars' atmosphere than in Earth's. Carbon and oxygen are the most easily available elements for helping the martians getting selfsustained. We have already discussed above how the available CO_2 (with the addition of hydrogen, which eventually can be extracted from the water ice which is abundant in the martian sub-surface layers) can be transformed into fuel for transportation and heating and for the jorney back to Earth. Some of the processes would produce as a by-product poisenous CO gas, for example oxygen production from $2CO_2 \rightarrow 2CO + O_2$. However, CO is poisenous because it is very reactive, which therefore also makes it a potential reactant in many useful industrial processes. If the CO is vented into the atmosphere, it will react uncontrolled with the soil throughout the martian surface. If on the other hand the hot CO "waste" gas is let into a controlled reaction with soils similar to those measured at the Viking lander places (40% SiO_2 and 20% Fe_2O_3) it will transform into pure SiO_2 (useful e.g. for glass production) and pure iron by the exothermic reaction

$$3CO + Fe_2O_3 \rightarrow 2Fe + 3CO_2 \tag{1.16}$$

An even more important reaction involving CO is

$$2CO + 4H_2 \rightarrow C_2H_4 + 2H_2O \tag{1.17}$$

Ethylene (C_2H_4) is a key to a full scale martian plastic industry.

A very important plastic from the very start could be socalled Kevlar which has comparable strength to steel, but can be produced in big transparent foils that could serve as the dome construction of the first larger pressurized greenhouses and habitats. The martian soil is well porous due to aoneas of dust storms, and it contain most of the crop fertilizing elements in larger amounts than soil on Earth. Vegetables and crops from

Earth are likely to grow well in presurized greenhouses on Mars. A 50 meter diameter 0.2 mm thick Kevlar dome would weigh one tonne, and would be strong enough to enclose a 50 mbar (0.7 psi) atmosphere, which is sufficient for plant growth. Lifting the pressure to 350 mbar (5 psi) would be sufficient for humans to walk inside the greenhouse without space suit, and at 500 mbar (with the right atmospheric composition and temperature) conditions would be as in the city of La Paz. 350 mbar is sufficient for polinating bees to fly around in the martian 1/3 Earth-gravity greenhouse environment – as if they were at home on Earth (again of course provided good enough atmospheric composition and temperature were provided). It is not well know what the optimal abundance of CO_2 in the atmosphere is, for Earth-like plants and human beings being able to breath it, but a new greenhouse may start out with air consisting of the maximum allowed amount of compressed martian CO_2 atmosphere for the first plants to convert into oxygen. If the air composed 7 mbar CO_2 (i.e. the original martian atmosphere, or twenty times the CO_2 partial pressure on Earth) augmented with nitrogen and oxygen for the rest, plants are already likely to grow several times faster than on Earth, and humans would be able to breath the air. Plants on Earth are generally deprived in CO_2 due to the very low CO_2 content in Earth's natural atmosphere. Greenhouses for tomatoe production on Earth are, for example, often artifically CO_2 enhanced to get the plants grow faster.

Fast growing crops could convert basically all the CO_2 in such domes into oxygen on timescales of weeks to months, hereby allowing a fully sustainable balance with the connected human habitates on ”a deeper ecologic level” than ever seen on Earth. It may even be that the first ”export” the martians could provide to Earth was knowledge and inspiration about how to connect the CO_2 outlet from conventional fosile power stations on Earth to high-technology high-production greenhouses that produce fast-growing high-quality food together with converting the CO_2 back into oxygen. The non-eatable (for humans) parts of the plants could (more efficiently than the standard procedure on Earth of plowing them into the grown) be used as basic food for example in ”farms” of Tilapia fish (cold bloded herbivores; recall that herbivores means feeding on plants as opposed to carnevores feeding on meet, e.g., ”predatory carnivores and their herbivore prey”). Experiments on Earth inspired by preparation of the colonisation of Mars has shown that the non-digestible plant waste could also conviniently be used as the soil for fast-growing protein-rich mushroms which were shown to convert 70% of undigestable plant remnants into eatable proteins, and growing well in warm areas with basically no sunlight. Such areas on Mars could for example be the underground tunnels between habitates or between habitates and farmland. One could even envision Earth's population benefitting from this offspring of the Mars colonisation research to find better and more ecological use of the argicultural plant waste on Earth to produce protein rich food that can grow at places where no other food can easily grow.

Patented 19th century American inventions changed Europe, and later the rest of the world, and made America rich and independent. In the same way, inventions produced as a matter of necessity by a frontier culture on Mars will change the civilisations that participate in the Mars-adventure, and later the rest of the Earth too, while it will make the martian society rich, self-sustained and independent. It would have been hard to guess in the 18th century that the inventions and products that would contribute most to make America rich, was not an improved steam engine, but rather abundant resources of fosile fuel, development of new forms of transportation, invention of fast food, electronics, military equipment, and foremost of all probably a new political structure that attracted hard working and inventive minds. Future generations are likely to be equally surprised about what precisely made the martian society prosperous, and how it affected life on Earth. Today we can only guess.

1.9.7 The need for a society to have available energy sources.

One of the stepping-stone challenges the martian society will be blessed with, is that here are no fosile fuels on Mars. Development of efficient, large-scale alternative energy sources are therefore (even) more urgently required on Mars than on Earth. The most powerful energy source we know of is fusion of hydrogen to helium. The fundamental reason behind this fact lies in the curve of binding energies that was introduced earlier in connection with the description of stellar evolution. On Earth the development of fusion technology is mainly halted by lack of investment, due to (too) easily available fosile fuels. We have already seen in detail in the chapter about stellar evolution how the Sun converts hydrogen to helium, known in astronomy as the pp-reaction. More than 99% of the reaction runs along the so-called ppI chain,

$$p^+ + p^+ \rightarrow D + e^+ + \nu_e \quad (1.18)$$
$$D + p^+ \rightarrow {}^3He + \gamma \quad (1.19)$$
$${}^3He +^3 He \rightarrow {}^4He + 2p^+ \quad (1.20)$$

However, the first reaction (1.18) is so slow that it only works in the Sun because there are imense amounts of hydrogen, and millions of years, available for the reaction. In a fusion reactor one would necessarily have to skip the first step and find available deuterium (D = ^{2}H) and/or helium-3 (^{3}He). If the development of fusion technology on Earth continues with the speed it has done for the last 50 years, it is likely that the first fusion reactor wil be developed by martians, and pieces of the technology itself as well as the raw material (D and/or ^{3}He) might be one of many potential martian export edventures to Earth that will make the martian society rich at the same time that it will benefit the Earth.

In fact, present day experimental fusion reactors do not really copy any of the PPI reactions, but rather either run via

$$D + T \rightarrow {}^4He + n + 17.6MeV \quad (1.21)$$

and could in principle also run via

$$D +^3 He \rightarrow {}^4He +^1 H + 18MeV \quad (1.22)$$

and only maybe one day via Eq. 1.20. However, the reaction cross sections in Eq. 1.22 and 1.20 are smaller than the one in Eq. 1.21. Therefore there is a large risk that low-probability side chains to Eq. 1.22 and 1.20 will lead to tritium formation that will then quickly use the deuterium up via Eq. 1.21 because of its higher reaction rate – it is not *only* because of lacking funding that the attempts to construct a commercial fusion reactor has so far taken more than 60 years of experiments; the details of the technique are really difficult.

On Earth one out of every 6,000 hydrogen atoms in ordinary water (and any other hydrogen containing molecule) will be D (more precisely D/H in standard ocean water is $1.56\,10^{-4}$). The technical effort necessary to seperate the rare deuterium from ordinary water brings it up to a commercial price tag of about 10,000$/kg. This is not a show stopper in itself since a kg of D will produce electricity worth in the surroundings of 5 million$. The Martians, however, already here has an advantage in not having spoilingly easy access to fosile fuel, and at the same time having a need to produce hydrogen for industry (Eq. 1.10, 1.16, 1.17), oxygen for breathing, and water for drinking (and many other purposes). All this will force them to electrolyse several tonnes of water per person per year. During this process it is easy, and at a low extra cost, to seperate D from H. Hence the water production can as a natural byproduct create of the order of a kg of deuterium per person per year, at an export value to Earth of 10,000$ or ready to create 5 million$ worth of energy per person on Mars itself once they master the fusion technology. In fact both fision and fusion reactors will strongly benefit from access to deuterium. The fusion reactor for the reason outlined above, and the fision reactor because it can run on the simpler "ordinary" uranium instead of enriched uranium if deuterium rich water can be used as moderator instead of ordinary water.

One kg of fuel in either reaction 1.21 of 1.22 will produce E[per kg] = E[per molecule]/N[kg per molecule] = 18MeV/4m$_u$ = 3× 10^{27} MeV = 10^8 kWh. There are approximately $1.6\,10^{-4}$ kg HDO per kg sea water, and since the total mass of the Earth's oceans is $1.6\,10^{21}$kg, then there are a total of approximately (2/19) × $1.6\,10^{-4}$ × $1.6\,10^{21}$ = $2.5\,10^{16}$ kg deuterium bound in Earth's oceans, or $\approx 10^{24}$ kWh worth of energy if the ocean's content of deuterium is burned in fusion reactors. The worlds total energy consumption today is around 10^{14} kWh per year, and therefore the deuterium could provide us with energy at the present day consumption level for approximately 10 billion years into the future if sufficient amounts of either T or ^{3}He could be made available (or reaction 1.19 and 1.20 could be made runing in a reactor).

At the present state of development, the deuterium-tritium reaction has a disadvantage compared to the polution-free expectation that exist to fusion. Free neutrons are radioactive in themselves (as we discussed in connection with the Big Bang nucleosynthesis, n $\rightarrow$ p + e$^-$ + ν with a halflife of $\sim$ 10 min). This radioactivity is technicallly difficult to get completely rid of. The most likely solution today seems to be to use a wall of lithium as moderator for the high energetic neutrons. The energy the neutrons display into the moderator wall is used to run a steamturbine which produce the outcome electricity, and at the same time the reaction of the neutrons with the lithium results in the production of tritium, which goes back into the reaction 1.21 in a closed loop, such that the only fuel needed for the reactor is deuterium. Technically it is, however, very difficult not to have the neutrons react with some other atoms as well, and this will eventuallt produce some radioactive waste. At present it seems that the amount of radioactive waste can be limited to $\sim$ 1,000 time less than in a fission reactor, but it is technically difficult to get it completely down to zero.

Reaction 1.22 (and 1.20) has the advantage of producing no radioactive waste products, but the disadvantage that ^{3}He is much less abundant in nature than D. ^{3}He is absent on Earth (unless it is produced in reaction 1.19), but available on the Moon, and this fact has contributed to some kind of new space race to get back to landings on the Moon, and it has been an official goal in the Chinese lunar space program to be able to mine ^{3}He from the Moon and bring it back to Earth to provide the whole Earth with polution-free energy. It will, however, not be trivial to cover the Earth's energy comsumption with fusion from reaction 1.22. It is not known how large the ^{3}He resources on the Moon actually are, but a likely estimate is that the upper 10 cm of lunar regolith all over the Moon is enriched with 4 ppb (i.e. $4\,10^{-9}$) ^{3}He. The lunar surface covers $4\,\pi(1.7\,10^6\mathrm{m})^2 = 3.4\,10^{13}\mathrm{m}^2$. If we set the density to 2.5 tons/m^2, then there are M[lunar ^{3}He] = $4\,10^{-9}\,2.5\,0.1\,3.4\,10^{13}\,10^3 = 3.4\,10^7$ kg ^{3}He on the Moon, which from reaction 1.22 could provide 10^{15}kWh of energy, or 10 years of the Earth's present energy consumption. The reason that the media (and the space agencies) some times talk about that there are ^{3}He available on the Moon for 10,000 years of Earth's energy consumption, is based on a probably quite unrealistic assumption that the ^{3}He layers stretches a few meters down into the soil rather than a few cm (increasing the resources with a factor 100), and the optimistic estimate that the abundance might be as high as 50 ppb instead of the 4 ppb assumed here (increasing the estimates with an additional factor of 10). If the latter estimates should turn out to be correct, it would require processing $7\,10^{13}$ tons of lunar regolith and changing the upper 10 meters of the entire lunar surface into a gravel pit to extract the fuel.

The reason that the lunar regolith, and not the terrestrial regolith, contain ^{3}He is that that it comes from the solar wind. On Earth the magnetic field and the atmosphere prevents the solar wind from colliding with the Earth's surface. On any airless body without an atmosphere (and with no geological activity to reshape the surface), the solar wind will have hit the regolith with a speed of hundreds of km/sec during billions of years and slowly have build up a small abundance of ^{3}He into the mineralogical structure. On the Moon lava outflow 3.5 billion years ago filled the low lying mare areas with dense basalt layers, dominated by iron and titanium rich minerals such as ilmenite (in proportion $FeTiO_3$). Following impacts have spread pieces of ilmenite from the mare region to all over the Moon. When high speed solar wind particles slam into the surface layers of the Moon, they will penetrate into the first few mm of whatever minerals exist on the surface, but in most minerals the solar wind hydrogen and helium will difuse out of the rocks quite quickly again and flow back into space again. The crystal structure of ilmenite is however such that the penetrating atoms will be locked into the mineral for billions of years. In this way the lunar surface has slowly been enriched in solar wind He which has a ^{3}He/^{4}He ratio of 2×10^{-6}. Meteoritic bombardment is continously transforming the lunar surface into the dust, soil, gravel and larger pieces which together is called the regolith. It is believed that the regolith is 10 to 15m thick in the highlands and 4 to 5 meters thick in the denser and younger mare regions. During impacts some of the material will form molten droplets which will again solidify into small glass inclusions. The more glas inclusions, the older (i.e. more bombarded) the regolith is. One usually use the term that the regolith is more mature. On the Moon the most mature regolith contain the highest concentration of ^{3}He, but the ^{3}He is not included into the glas, it is just an expression of that the mature regolith has been bombarded for the longest time (with micro meteorites as well as with solar wind particles). The ^{3}He is concentrated to the surface layers of the ilmenite grains. It is not obvious how deep the mature regolith layer is. It may be a few cm and it may be a few meters. The length of time lunar ^{3}He could potentially support human energy consumption depends critically on the depth of the mature regolith layer.

The regolith layer on the two Martian moons is estimated to be of the order 50m thick, and also some of the asteroids seems to have substantial regolith layers. The largest of the martian moons is circling only 7,000 km above the martian surface, and even the asteroids in the asteroid belt (and maybe somewhat surprisingly also the Moon) are accessible at lower energy cost from Mars than from Earth.

Even if the reaction 1.22 may not contribute to solving the Earth's energy demands, it still has the advantage that it, at least in principle, would be possible to make rocket motors running on ^{3}He (with a power per kg fuel several thousands times larger than ordinary chemical fuel). This may not be uninteresting for a civilisation that in the future might have its main income in the trading with Earth from mining asteroids and sell the products to Earth – and maybe to sell it throughout the rest of the solar system one day, too.

1.9.8 The rocket propulsion.

The basic principle of a rocket is to use the principle of momentum conservation (action = re-action), i.e. that the achived momentum of the rocket, ΔV M$_d$, equals the momentum of the exhaust gas,

$$\Delta V M_d = m_f v_{\rm ex} \tag{1.23}$$

where M_d is the dry mass of the rocket (i.e. the empty rocket, M_r, plus the payload, M_p), ΔV is the obtained flight velocity, m$_f$ is the mass of the fuel, and v$_{\rm ex}$ is the velocity with which the motor is able to expell the exhaust from burning the fuel. The higher the speed of the exhaust is, the higher is the produced momentum of the system, and the lower the mass of the empty rocket is compared to the mass of the payload, the larger fraction of the produced momentum can go into momentum of the payload (i.e., the more payload can the rocket lift).

The force $F_{\rm thrust}$ the motor is exerting on the rocket is called the thrust, and it is the product of the mass of the exhaust, $\dot{m}$, expelled per time unit, times the velocity, $v_{\rm ex}$ with which the exhaust gas flows (e.g. in units kg/s x m/s = kg m s^{-2} = N)

$$F_{\rm thrust} = v_{\rm ex} \times \dot{m} \tag{1.24}$$

which is another way of calculating the final change in velocity exerted on the rocket,

$$\Delta V = \int \frac{F_{\rm thrust}}{M_{tot}(t)} dt = \int \frac{v_{\rm ex}\dot{m}}{M_{tot}(t)} dt \tag{1.25}$$

It can be shown that these relations leads to the so-called rocket equation

$$\Delta V/v_{\rm ex} = ln[(M_d + m_f(t=0))/M_d] \tag{1.26}$$

which clearly illustrates how the obtained velocity of a rocket of a given dry mass (M$_d$) and total mass ($M_{\rm tot}(t=0)$ = M$_d$+m$_f(t=0)$) is directly proportional to the exhaust velocity. However, it further shows that the actual velocity with which a rocket is able to sent a given payload on its way, grows faster than lienarly with the exhaust velocity, because the amount of fuel (m$_f(t=0)$), and hence also the mass of the empty rocket (M$_r$), can be made smaler if the energy available per mass unit of fuel goes up, and the fraction of the total mass of the system that can go into payload therfore grows faster than immediately seen from Eq. 1.26. In praxis one would usually be able to construct rockets such that $\Delta V/v_{\rm ex} \approx 2$. Now, typical exhaust velocities obtainable with chemical fuels (as e.g. methane-oxygen fuel) is $\sim$ 3,000 m/s. With D-^{3}He fuel on the other hand one can obtain exhaust velocities of $\sim$15 million m/s, which could in principle allow rockets to fly at velocities of 10% of the speed of light. Although such speeds are theoretical limits, it is obvious that development of D-^{3}He fusion has large potential for a space faring civilisation. A similar benefit cannot be obtained from the deuterium-tritium fusion, because the energy in Eq. 1.21 is released in the form of neutrons which are complicated to direct through a nocle, while the charge protons and alpha particles in the 1.22 reaction can be immediately directed through the nocle via a magnetic field.

In summary of the discussion of nuclear fusion, the Martian civilisation will have a large potential and insentive for developing nuclear fusion, maybe by use of Eq. 1.21 for surface activities and from Eq. 1.22 for rocketing, to the benefit for the martian economy and activity as well as for the economy and environment on Earth. In the more far away future, humans may be able to take advantage of the million times larger resources of ^{3}He that exist in the atmospheres of the giant planets in the outer parts of the solar system. Humans are capable of colinizing Mars during the next decade, and for those future generations where the challenges of living on Mars are no longer enough, the outer solar system is more exotic. Titan has a nitrogen based atmosphere just like the Earth, but the second most abundant molecule is not oxygen as on Earth but methane as in natural gas. The temperature is minus 180°C, and the atmosphere is so thick and the gravity so low that humans could get around by attaching wings and fly by their own muscle power like a bird. So who knows, maybe one day the colonisation of Titan will not be for economic reasons, as for Mars, but just for the fun of it.

1.10 Is it really *us* to colonize the Galaxy, and is *now* the right time to begin?

According to Fermi's "profety", any intelligent civilisation with technological ability and scientific curiosity will take approximately 5 million years to colonize the entire Galaxy once they start. It looks as if we took the first direct and serious step in this process around 2018, when the first exploratory rockets to bring human colonists to Mars took off to martian orbit, and the first interstellar exploratory satellites were finally under construction. Are we really the first civilisation in the entire Galaxy to have done so? Are we really on our way, or is this araising atempt doomed to fail from the very start?

Figure 1.28. Is that it – our begining colonization of the Galaxy ?