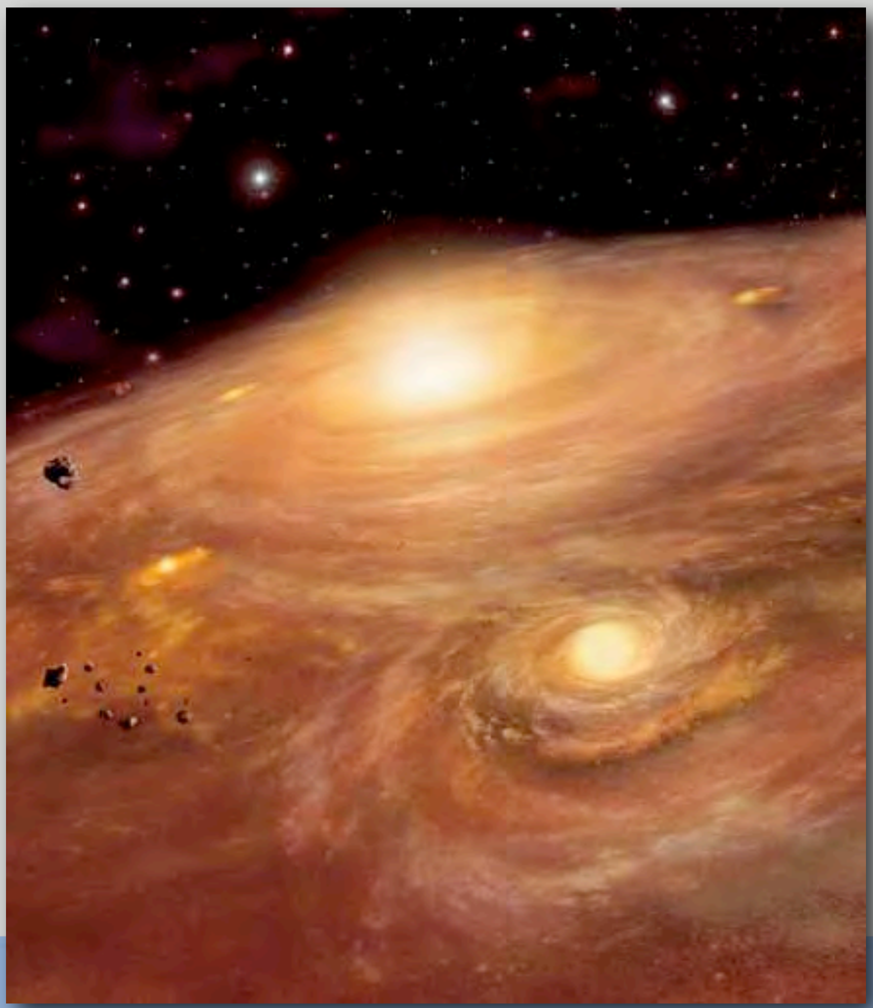
Planet formation and migration

Planetary Sciences Chapter 12

Daphne Stam



Netherlands Institute for Space Research

Netherlands Organisation for Scientific Research

Outline

- Observational constraints
- Star formation
- Early stages of planet formation
- Terrestrial planet formation
- Giant planet formation
- Planetary migration
- Asteroids and comets
- Moons and rings

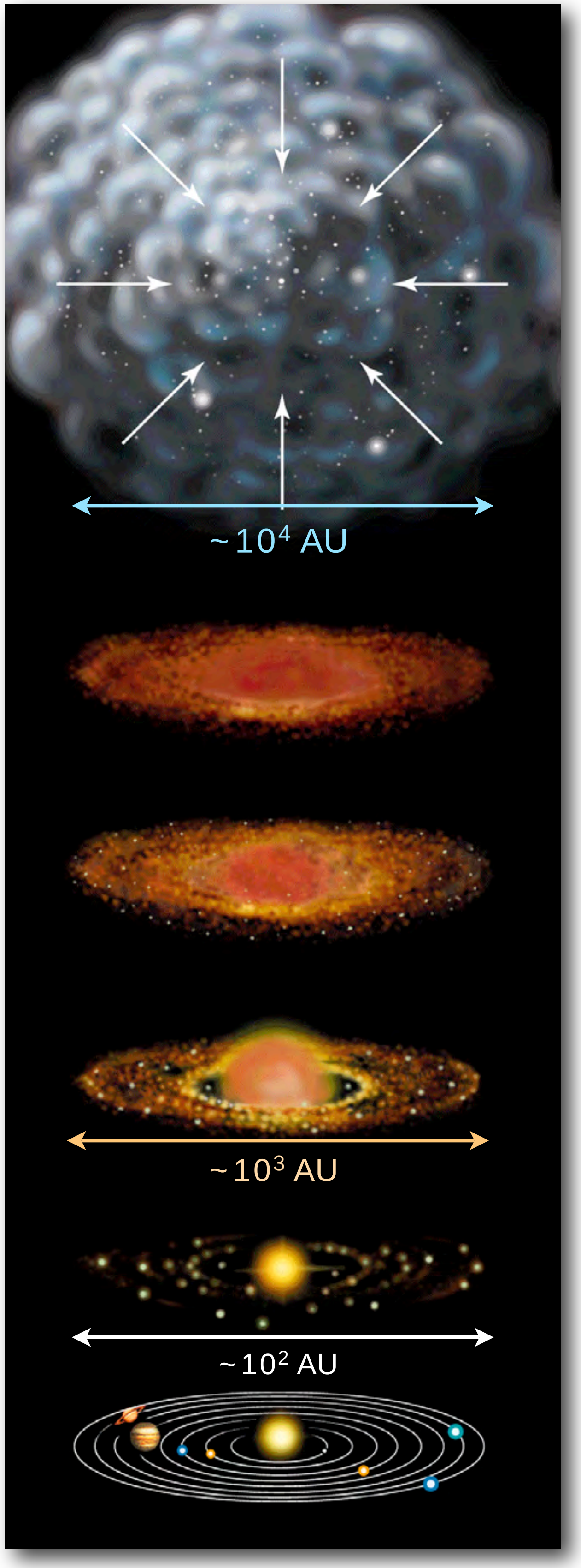
Observational constraints

Any theory on the formation and evolution of planetary systems must explain the following observed properties of the Solar System and exoplanetary systems*:

- Orbits, spacings, and rotation of the planets
- The ages of the planets
- Sizes, densities and compositions of the planets
- Asteroids, comets, moons, rings
- The surface structure of terrestrial planets and moons
- The composition of planetary atmospheres

* Of course very few of these properties are currently observable for exoplanetary systems

The birth of the Solar System in a nutshell



1. Collapse of an interstellar molecular cloud $t=0$
2. Formation of a circumstellar disk
3. Birth of the protostar $t=10^5$ - 10^6 yr
4. Development into main-sequence star $t=10^6$ - 10^7 yr
5. Formation of planetesimals
6. A mature planetary system has been formed $t>10^7$ yr

Interstellar molecular clouds

A giant molecular cloud in Ophiuchus

Credit: Loke Kun Tan/starryscapes.com

Properties of molecular clouds:

- composed mainly of H_2 and some He
- small fraction of other molecules:
  CO, CN, CS, SiO, OH, H_2O, HCN, SO_2,
  H_2S, NH_3, H_2CO, etc ....
- masses vary from ~1 to ~10^5-10^6 M_{Sun}
- temperatures are ~10-30 K
- densities vary from ~10^3 to 10^5 cm^{-3}

Molecular clouds are usually stable against self-gravitational collapse because of internal gas pressure, magnetic fields, turbulent motions, rotation, ...

Collapsing interstellar molecular clouds

The *virial theorem* describes a system in equilibrium:
half the gravitational potential energy of the system equals its kinetic energy:

$$E_G = 2\ E_K$$

Assume that the kinetic energy of our cloud equals its thermal energy:

$$E_K = N\ 3/2\ \ kT$$

N: total number of atoms, k: Boltzmann's constant, T: temperature

The gravitational potential energy of our cloud equals:

$$E_G = 3/5\ \ GM^2/R$$

G: gravitational constant, M: total mass, R: radius of the cloud

Thus, for a cloud in equilibrium (no collapse, no expansion) we have:

$$3/5\ \ GM^2/R = 3\ N\ kT$$

A cloud will expand, if: $3/5\ \ GM^2/R < 3\ N\ kT$

A cloud will collapse, if: $3/5\ \ GM^2/R > 3\ N\ kT$

Collapsing interstellar molecular clouds

Assuming the cloud has a constant density ρ, and m is the mass per atom, we can solve for the critical mass of a cloud, the Jeans mass, M_J:

$$M_J = \left(\frac{5\,kT}{Gm}\right)^{3/2}\left(\frac{3}{4\,\pi\rho}\right)^{1/2}$$

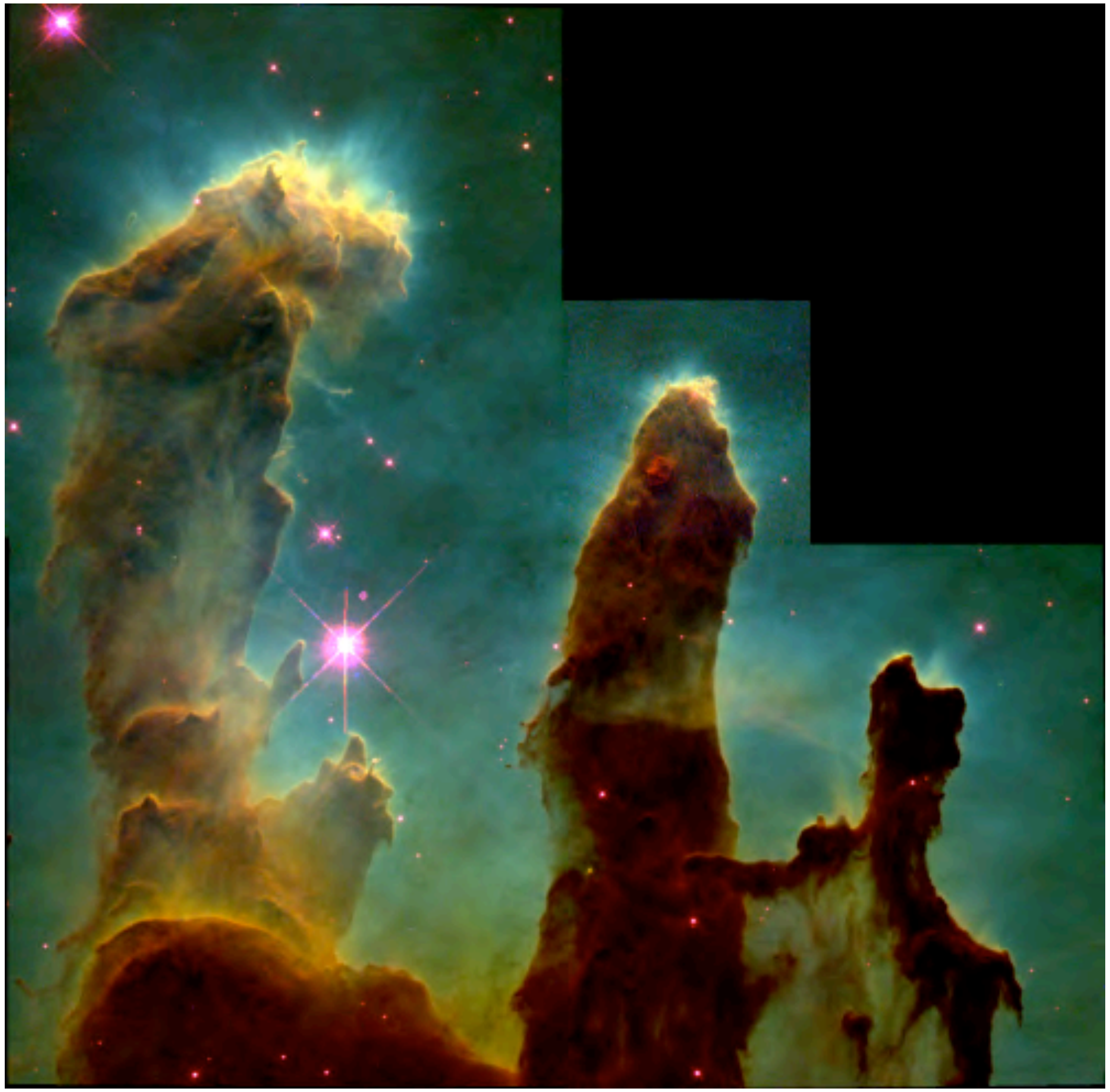
A cloud with $M > M_J$ will collapse if its only means of support is internal thermal pressure.

Possible triggers for collapse:

- galactic pressure changes
- nearby supernova explosions
- expanding hydrogen-regions

The birth of stars:

Stars are born in the densest regions of the collapsed molecular clouds.



Starforming region in the Eagle Nebula

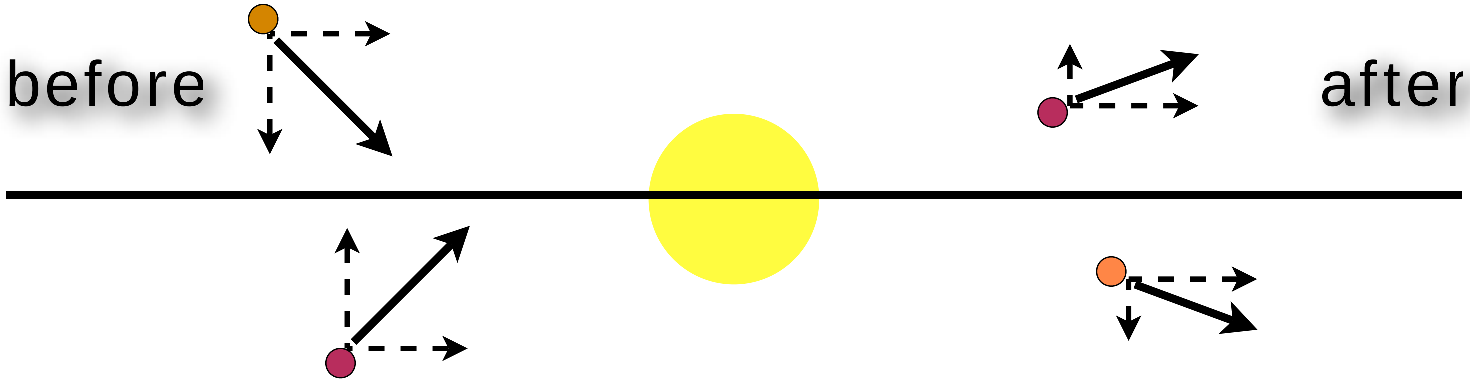
Credit: HST NASA/ESA

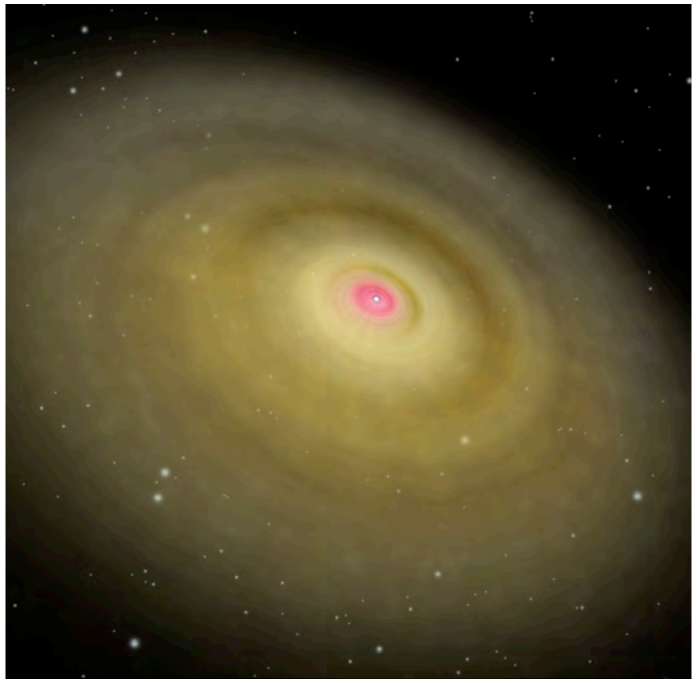
The evolution of young stars:

- the core temperature T increases
- as long as the cloud is sufficiently transparent, excess heat is radiated away, and T remains relatively low
- because of the increasing density, the transparency decreases, and T increases
- once $T > 10^6$ K, deuterium fusion starts, and contraction slows down
- once deuterium runs out, the contraction starts again, T increases
- once $T > 10^7$ K, hydrogen fusion starts, and contraction slows down

The formation of protoplanetary disks

Material from the molecular cloud collapses towards the core, but conservation of angular momentum prevents it from falling onto the protostar. It will orbit the core and settle into a disk around it:



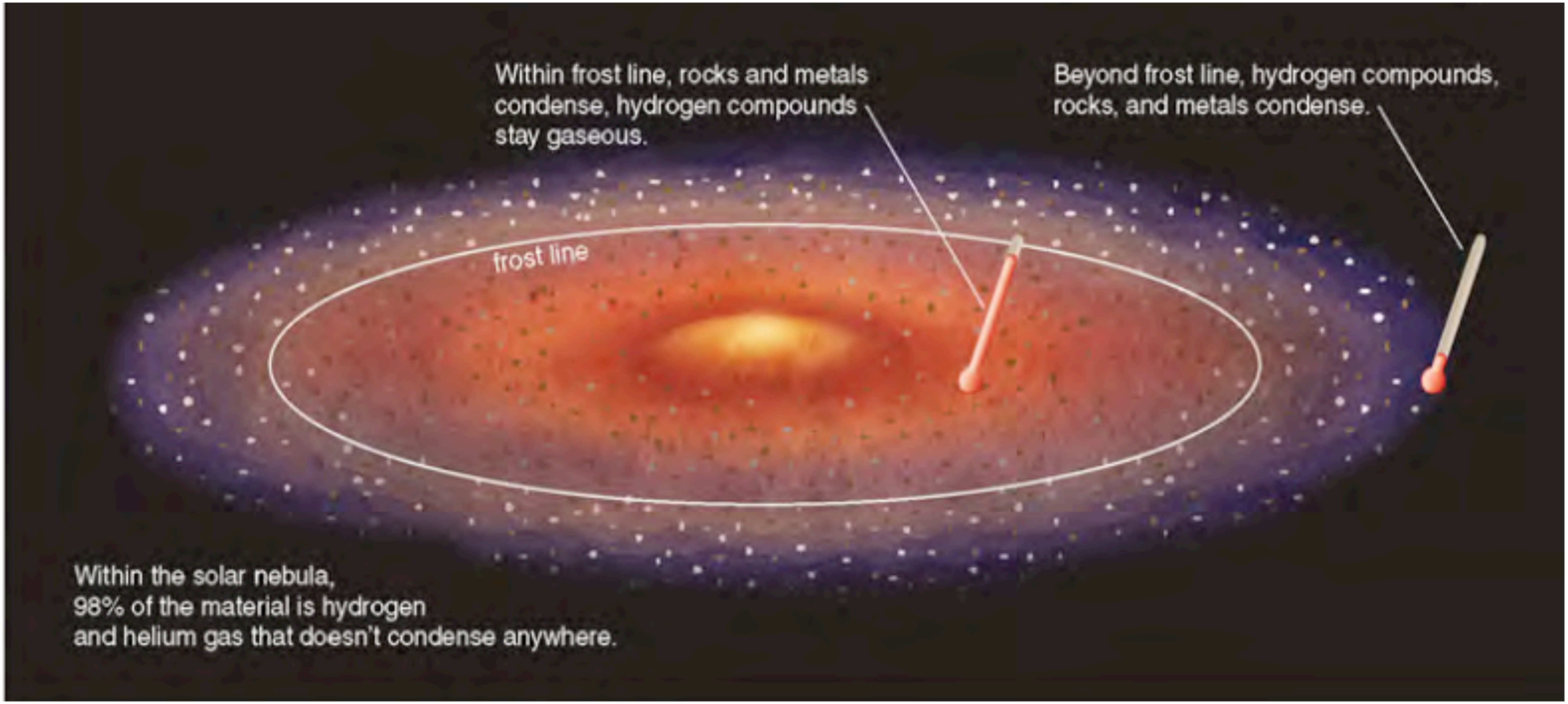


Credit:Subaru Telescope

Excess gravitational potential energy is dissipated as heat in the forming disk. The closer to the core, the higher T, because the more potential energy is released …

Chemistry in a protoplanetary disk

The chemical reactions in the protoplanetary disk depend strongly on the local temperatures, pressures, and mixing processes:

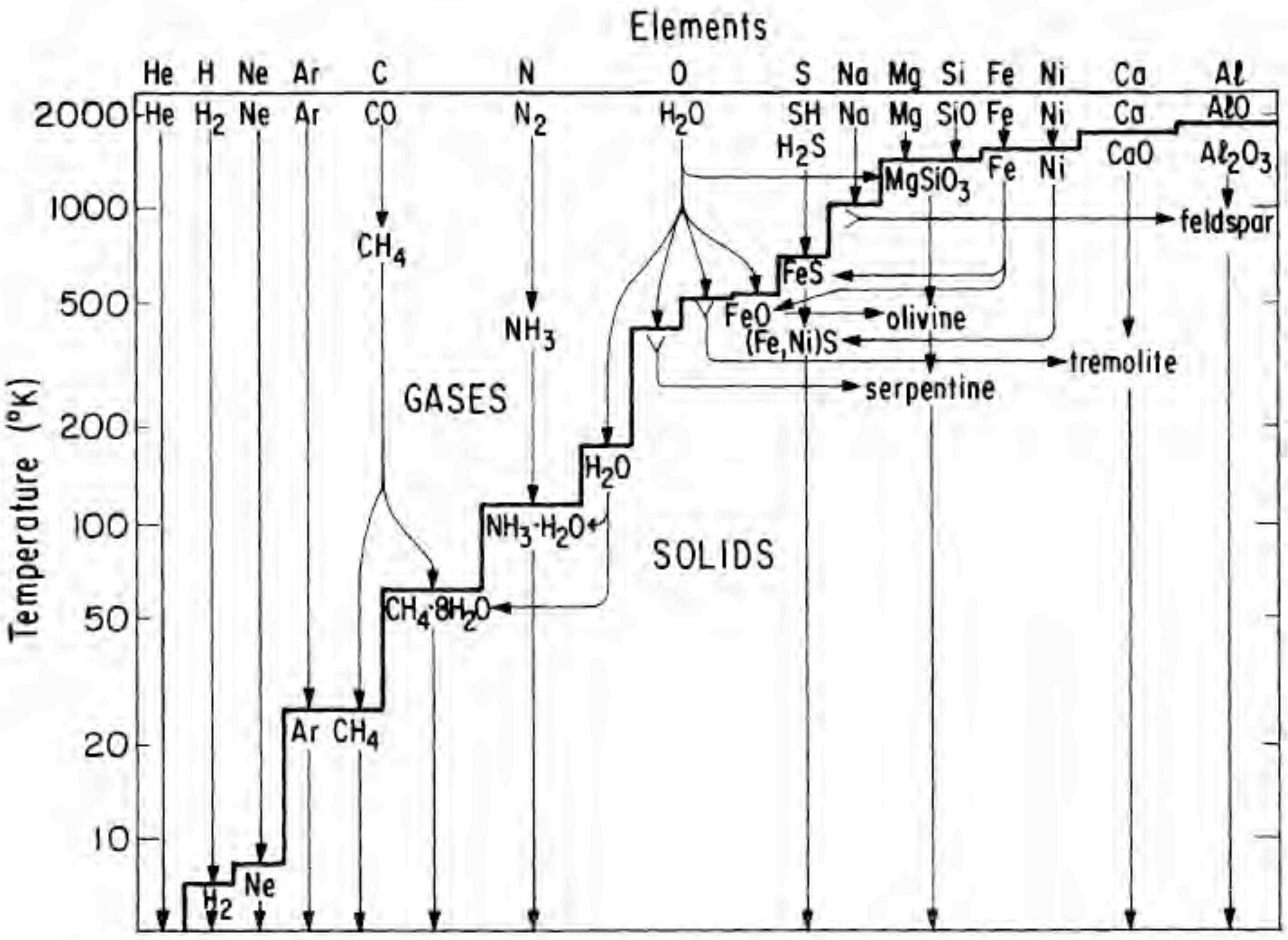


Credit: Addison & Wesley Publishing

The location of the frost/snow/ice line depends on the luminosity of the star, and the density and composition (mixing) of the disk

Condensation in a protoplanetary disk

The major reactions during fully equilibrated cooling of solar material from 2000 to 5 K, at a pressure of ~10^{-3} bar (from Barshay & Lewis, 1976):

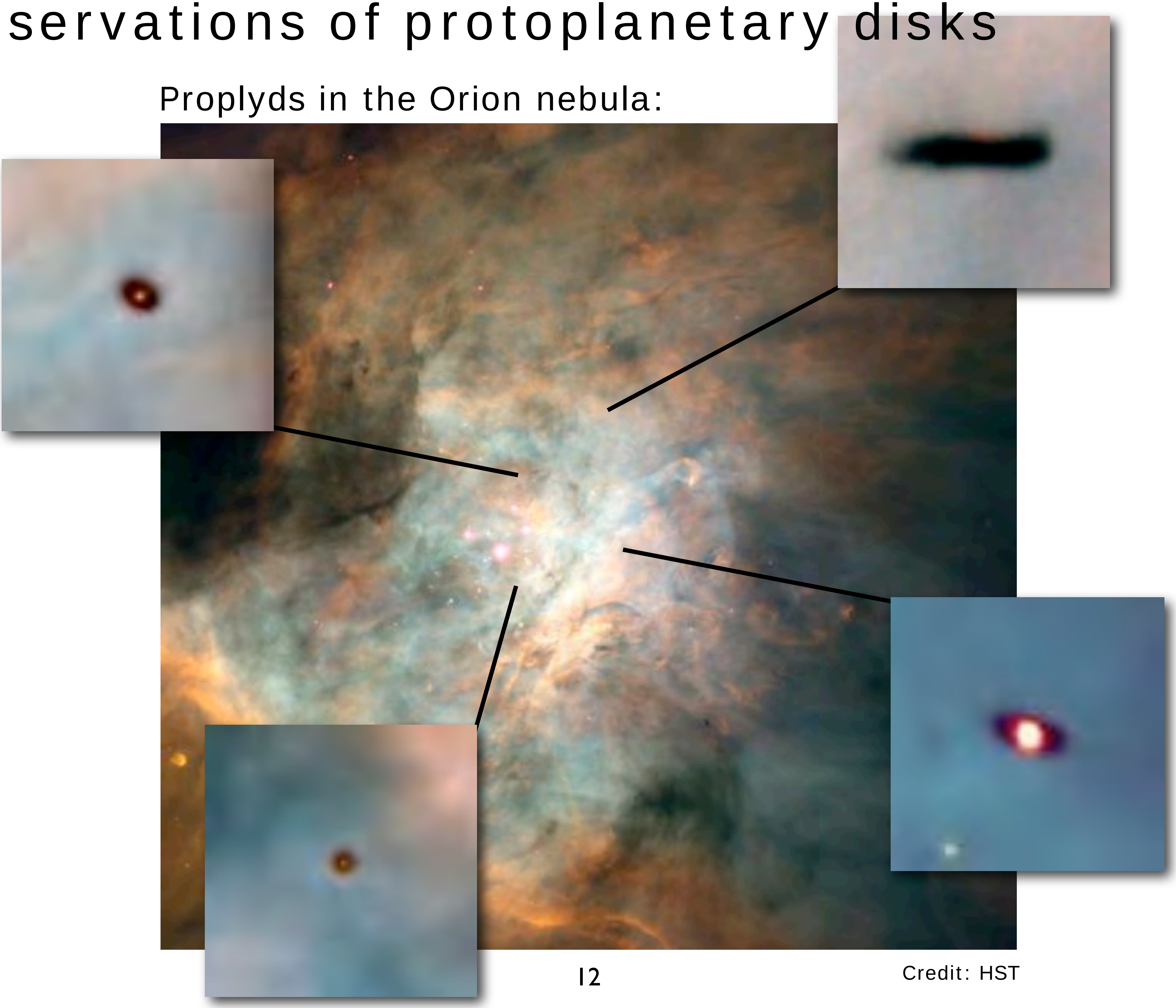


Note 1: at low temperatures (outer disk regions) there is no full chemical equilibrium

Note 2: dynamical mixing processes, transport, and e.g. cosmic ray ionisation are also important

Observations of protoplanetary disks

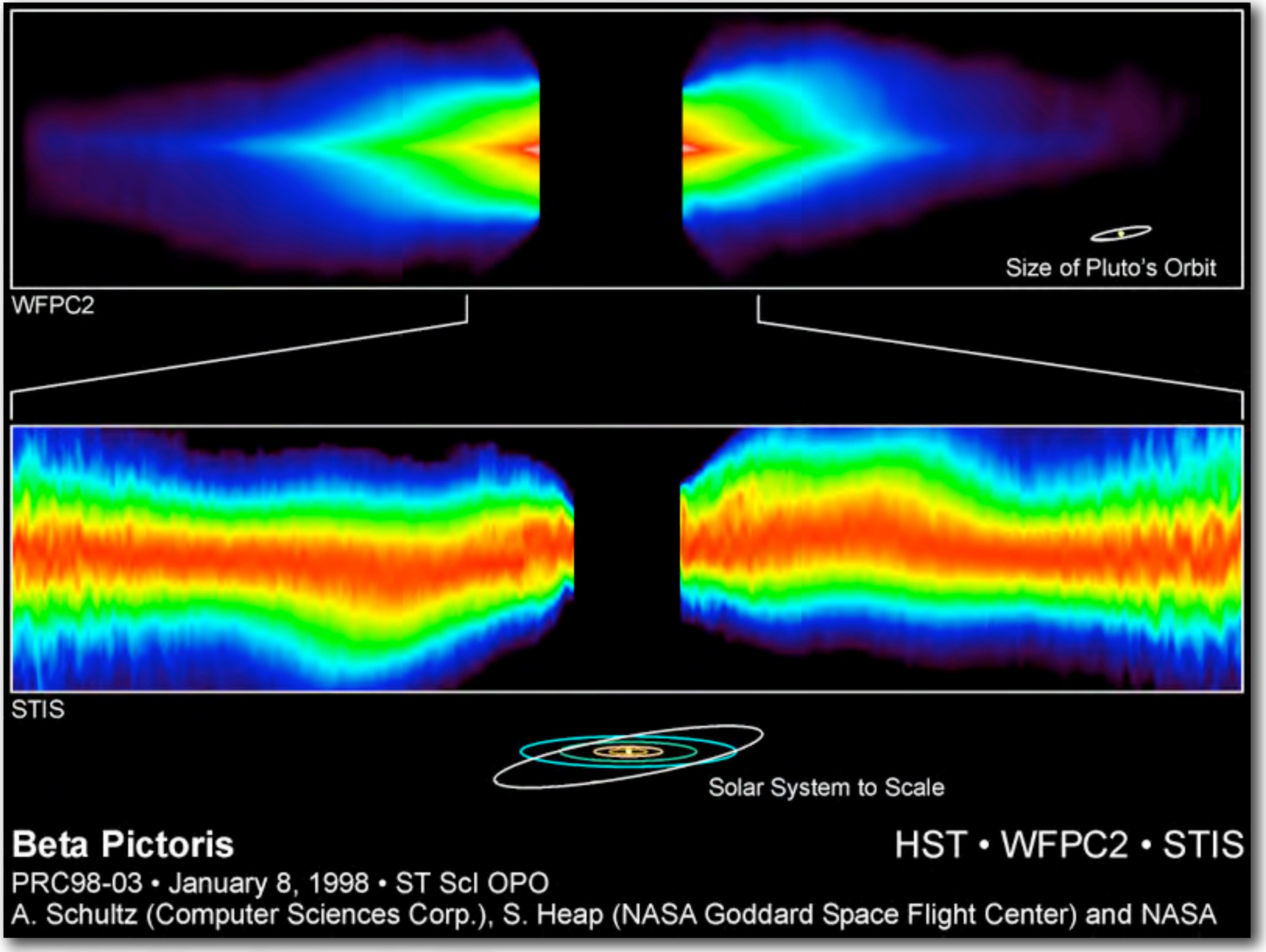
Proplyds in the Orion nebula:



Credit: HST

Observations of protoplanetary disks

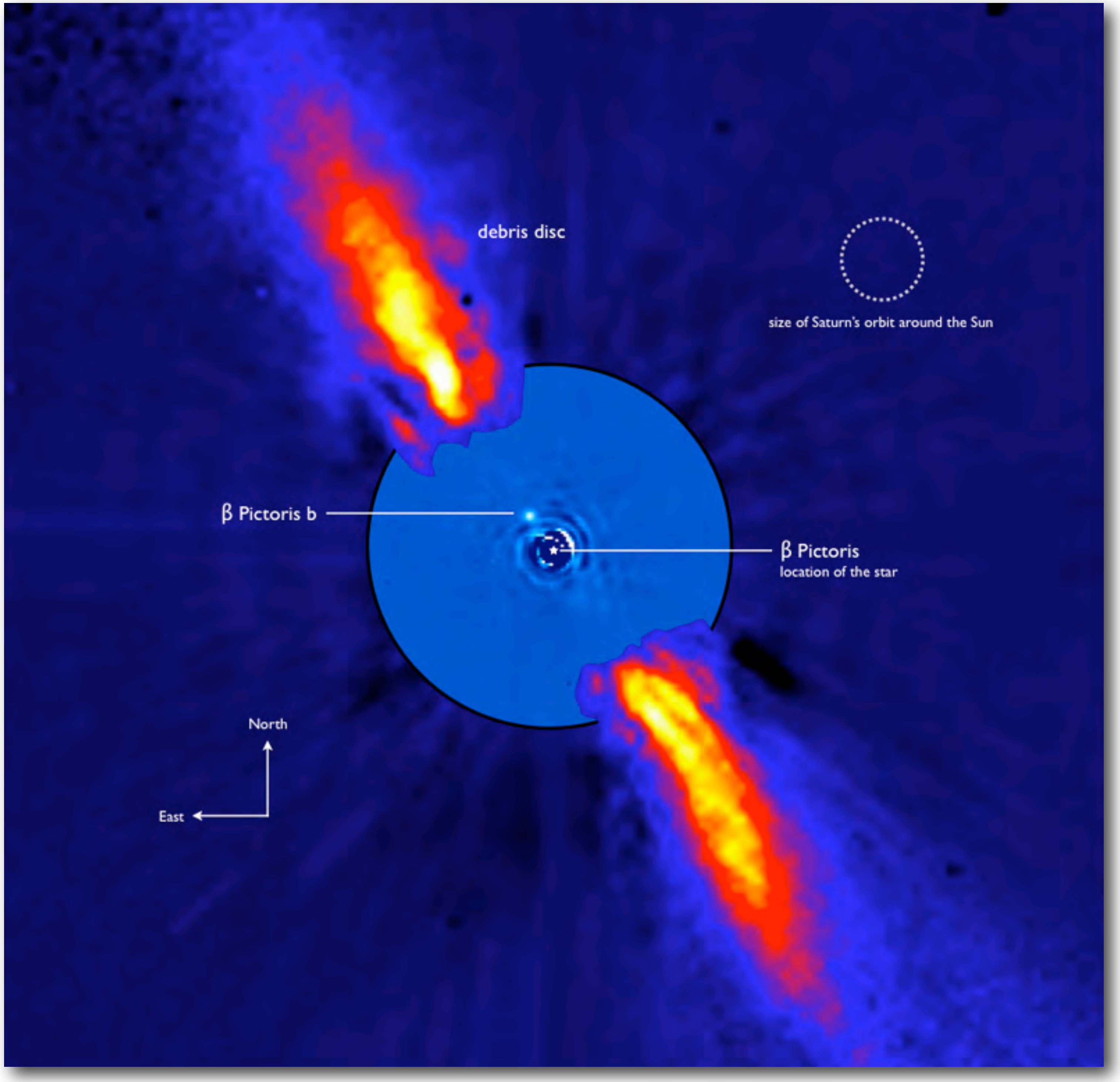
The dust disk around the star β-Pictoris:



This debris disk was first imaged in 1984

Observations of protoplanetary disks

A planet in the dust disk around the star β-Pictoris?

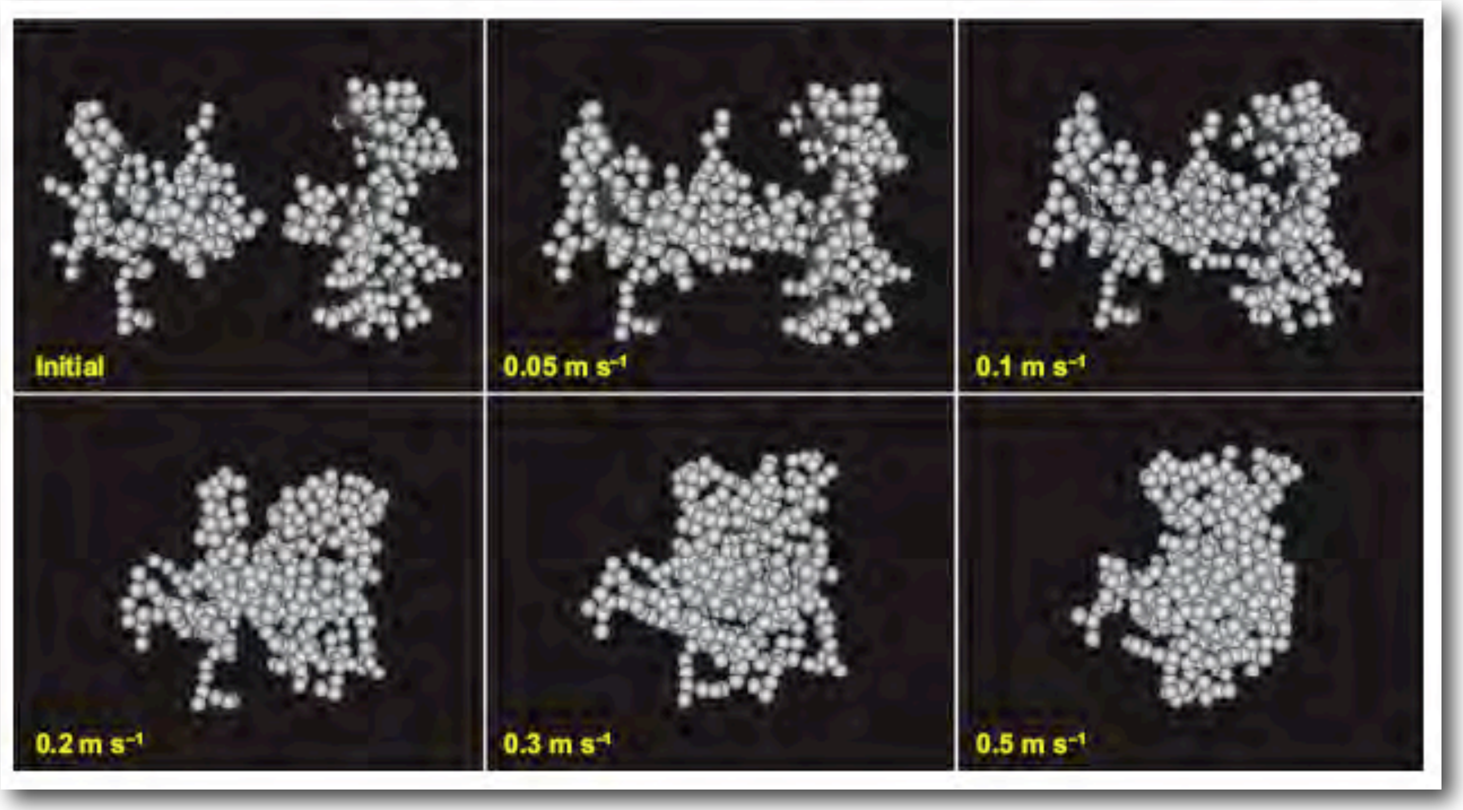


Credit: ESO/Lagrange et al. 2008

Early stages of planet formation

As a disk of gaseous matter cools, various compounds start to condense into microscopic grains that grow through collisions :

- At high temperatures: silicates and iron compounds
- At lower temperatures: water-ice and other ices



Simulations by Paszun & Dominik for collisional growth of dust aggregates assuming different impact velocities.

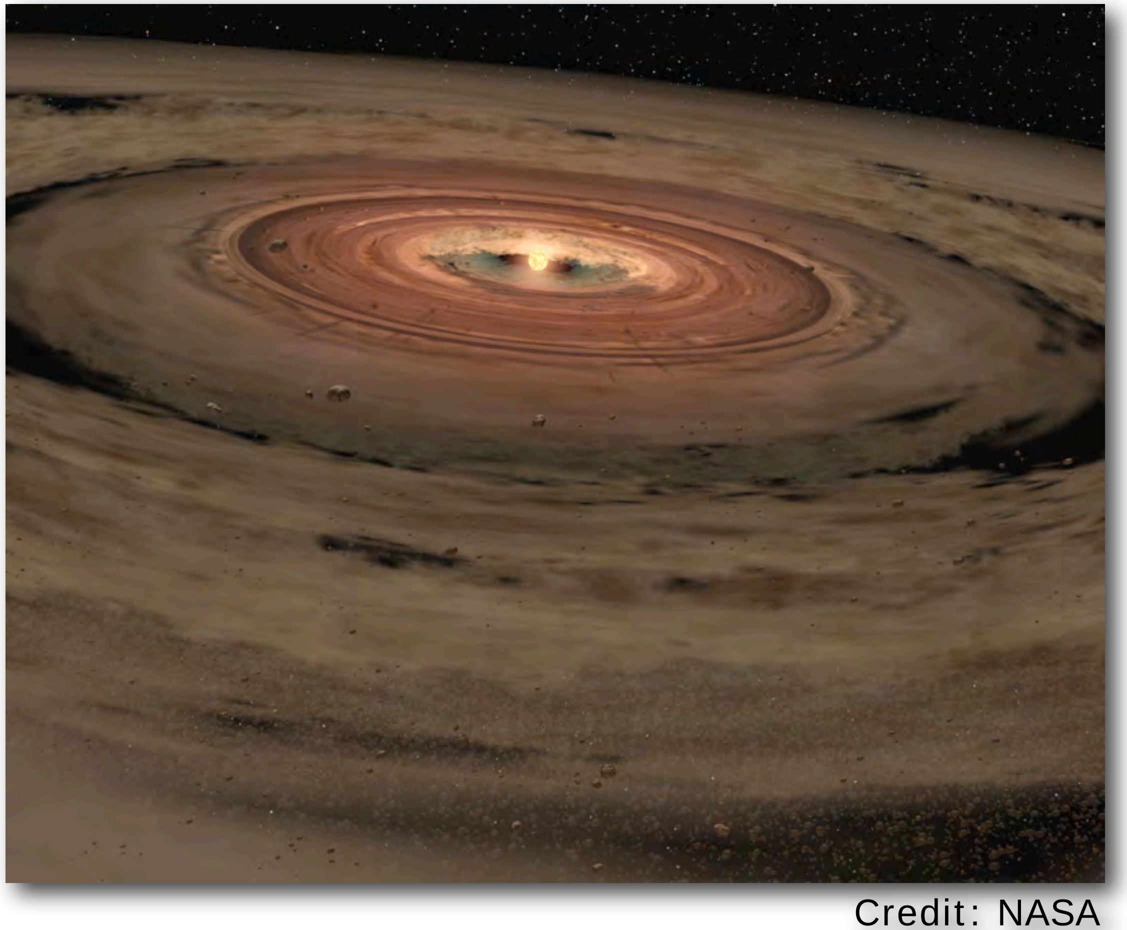
Small (< cm) particles move with the flow of the gas …

Formation of small planetesimals

Intermediate bodies (~ 1 m) are formed through collisions of the fluffy, small particles.

A problem:
The gas rotates slower than the keplerian speed of these bodies ...
Intermediate bodies will suffer from gas drag and spiral towards the star ...*
While large (> km) planetesimals will not suffer from the gas drag ...

The transition from m to km sized bodies should take place rather quickly and is not understood yet!



Credit: NASA

---

* A meter-sized body at 1 AU would spiral into the Sun in ~100 years!

Formation of protoplanets

Kilometer-sized bodies will attract each other gravitationally and grow through collisions.

The largest planetesimals will accrete almost everything that they collide with: *runaway* or *oligarchic accretion*

The isolation mass, M_i, the largest mass (in grams) to which a planetesimal orbiting a 1 M_Sun star can grow by runaway accretion is:

$$M_\mathrm{i} \sim 1.6 \times 10^{25} \, (r^2 \, \sigma)^{3/2}$$

Here, r is the orbital distance and σ is the disk's surface density (in g cm^{-2})

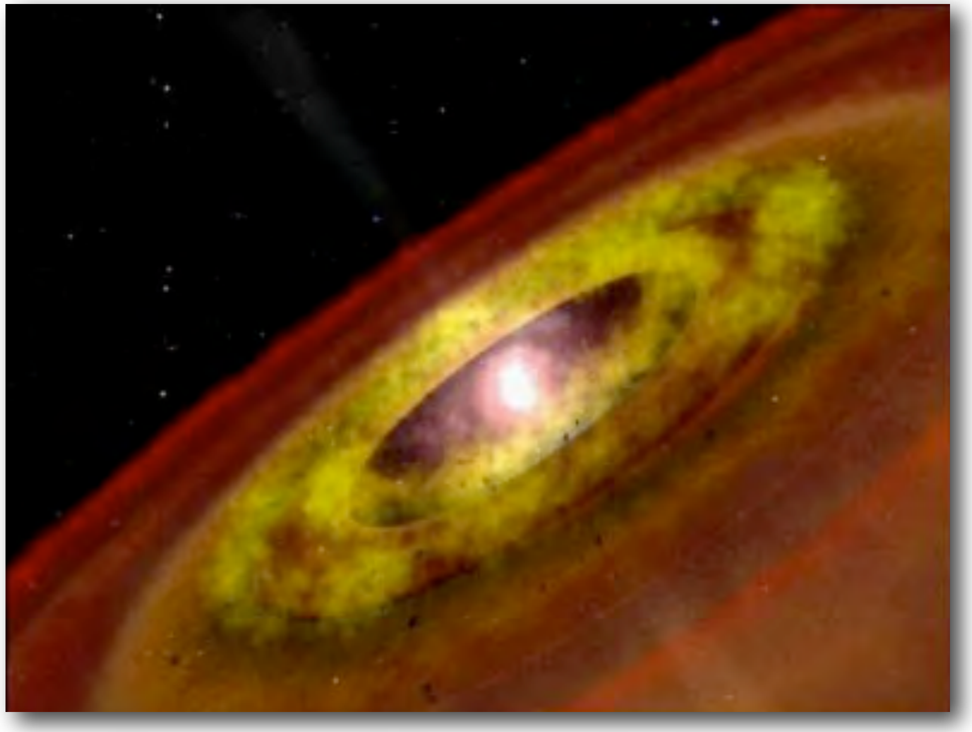
Without 'fresh' material and assuming a disk mass of 0.02 M_Sun,
$M_\mathrm{i} = 0.08 \, M_\mathrm{Earth}$ for the Earth, and ~0.1 $M_\mathrm{Jupiter's\ core}$ for Jupiter's core …

Gas-clearing stage of the disk

Since there is no gas left between the planets, the gas must have been cleared away at some stage during the evolution process.

Possible explanations:

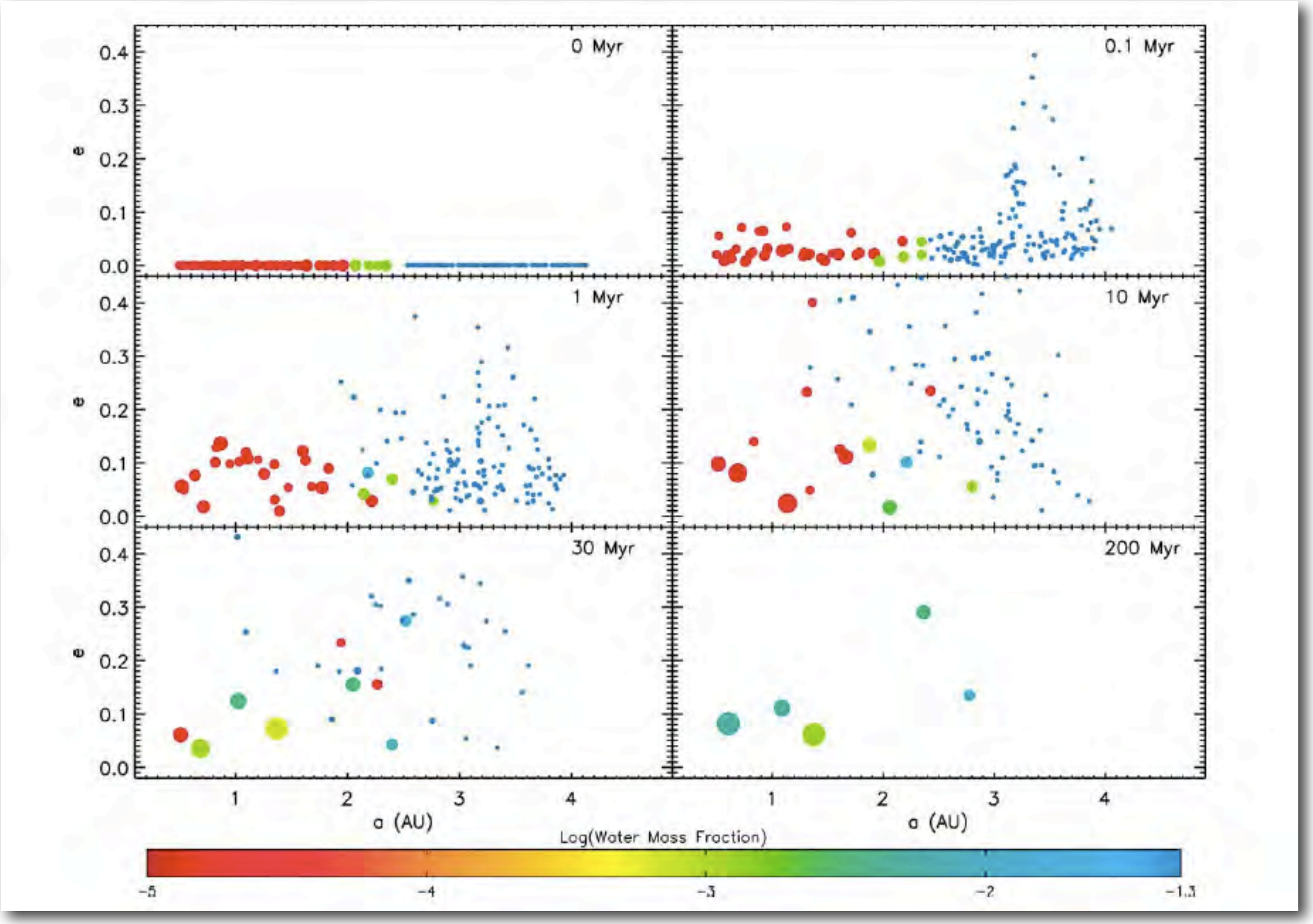
- ~10^6-10^7 yr after its formation, a pre-main sequence star goes through its T Tauri phase with very strong stellar winds that 'blow' the gas out of the system
- the gas 'evaporates' from the disk due to the UV flux of the young star itself or of nearby massive stars



The sun as a T Tauri star. Credit: NASA

Terrestrial planet formation

Numerical simulations of planetesimal accretion (with Jupiter at 5.2 AU):



Credit: Raymond, Quinn, Lunine (2004)

The origin of the Earth's moon

Credit:NASA/Galileo mission

The origin of our moon has long been debated. Some early ideas:

- The Earth and moon formed at the same time and location (as a double-planet system)

  The average density of Earth is 5.5 g/cm^3, that of the moon 3.3 g/cm^3 (Earth has a large iron core, the moon doesn't): these should be similar

- The moon formed somewhere else (where there was less iron) and was captured by the Earth

  The moon has the same oxygen isotope ratio as the Earth, while Mars and meteorites have different ratios. The moon formed thus near the Earth

- The early Earth rotated so rapidly that the moon spun off from the Earth

  This would give the moon its composition, but leaves too much angular momentum in the Earth-moon system

The origin of the Earth’s moon

The Earth’s moon likely formed about 4 - 4.5 billion years ago, through a giant impact of the half-built Earth with a body about half its size:

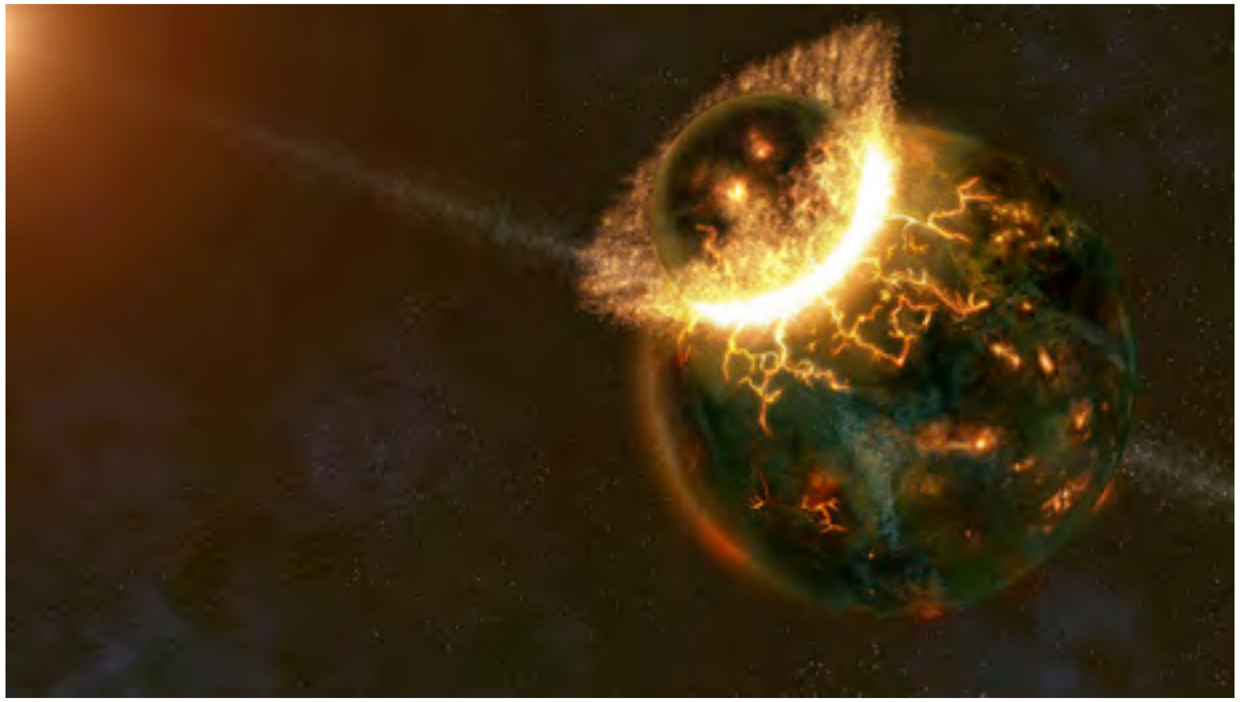


Image credit: National Geographic

After the impact, the Earth had about 2/3 of its current size and the moon formed from the ring of debris. Both the Earth and the moon continued to accumulate impact material after this event.

This chance event would explain why Venus and Mars lack large moons!

The origin of the Earth’s moon

A numerical simulation of the possible events during the giant impact:

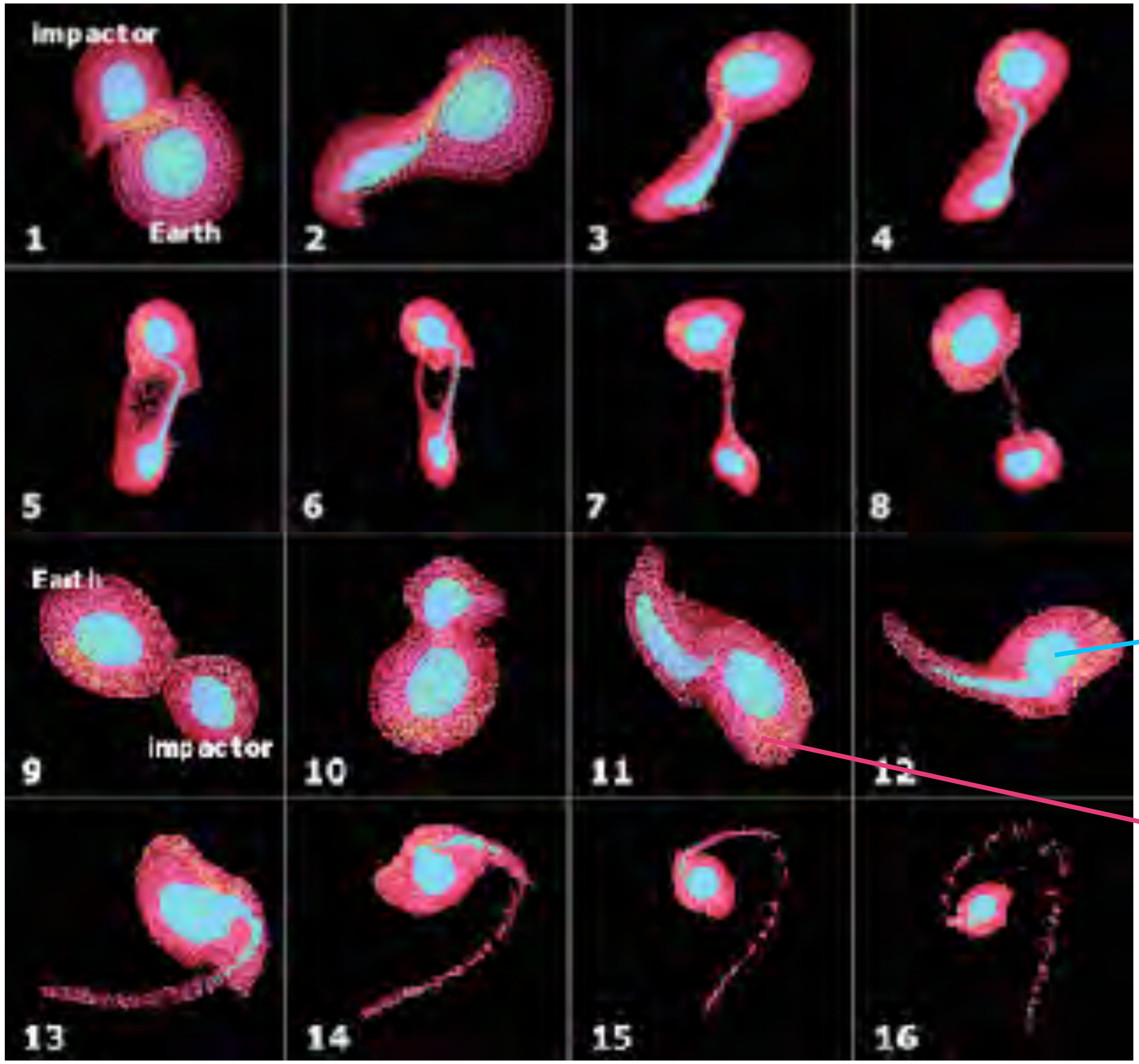
1. The impactor hits the Earth off-centered
2. The impact heats and deforms both bodies

3-8. The impactor ‘bounces off’

9. The impactor hits the Earth again

10-12. Its metallic core merges with the Earth’s

13-16. Rocky material is left orbiting the Earth



Credit: A.G.W. Cameron, Harvard College Observatory

Giant planet formation

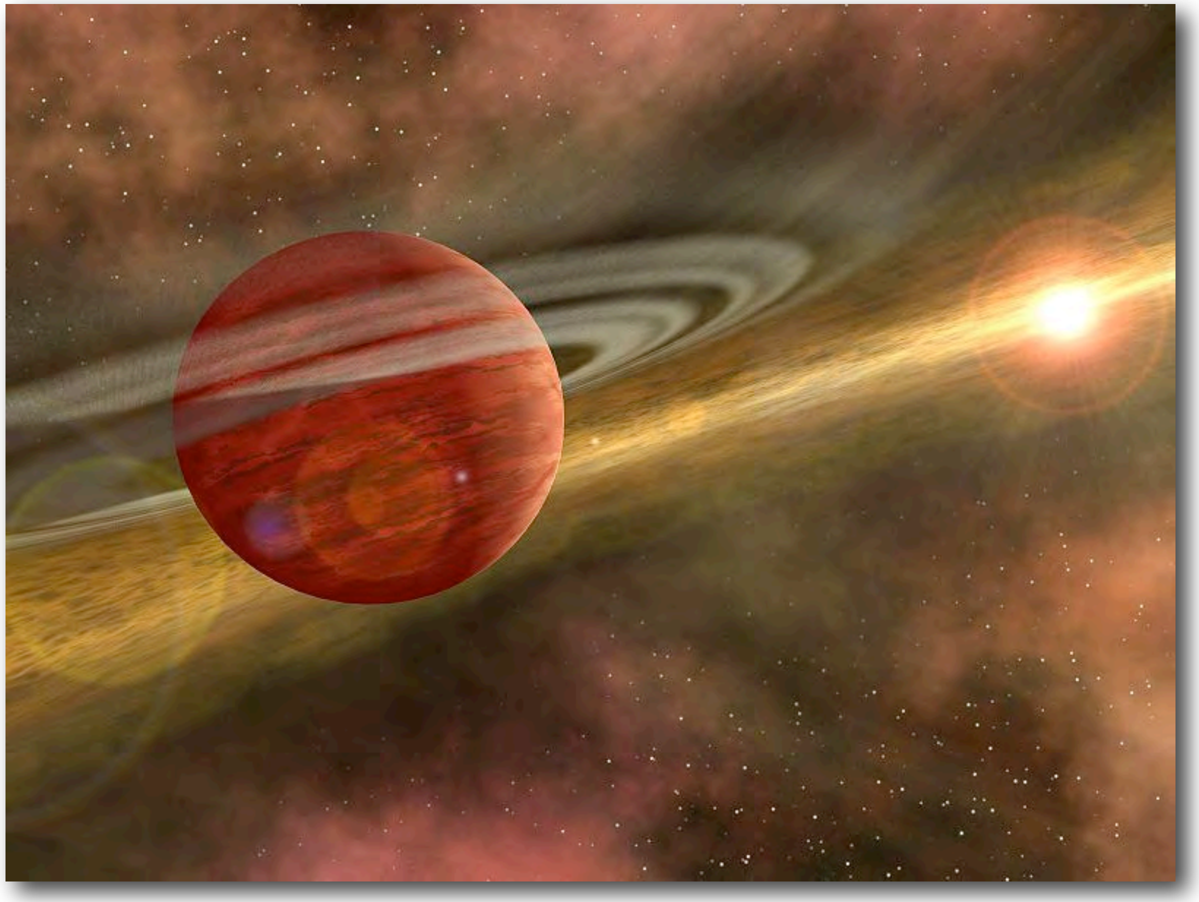
Heavy elements constitute less than 2% of the mass of a solar composition mixture.

The giant planets are enhanced in heavy elements by:

5x Jupiter
15x Saturn
300x Uranus and Neptune

Two main methods are proposed for giant planet formation:

- gas-instability
- core-instability



Gas-instability hypothesis

Gravitational instabilities in the protoplanetary disk form clumps of material, the protoplanets, with the solid parts settling in their cores.

“Top-down planet formation”

Advantages of the gas-instability hypothesis:

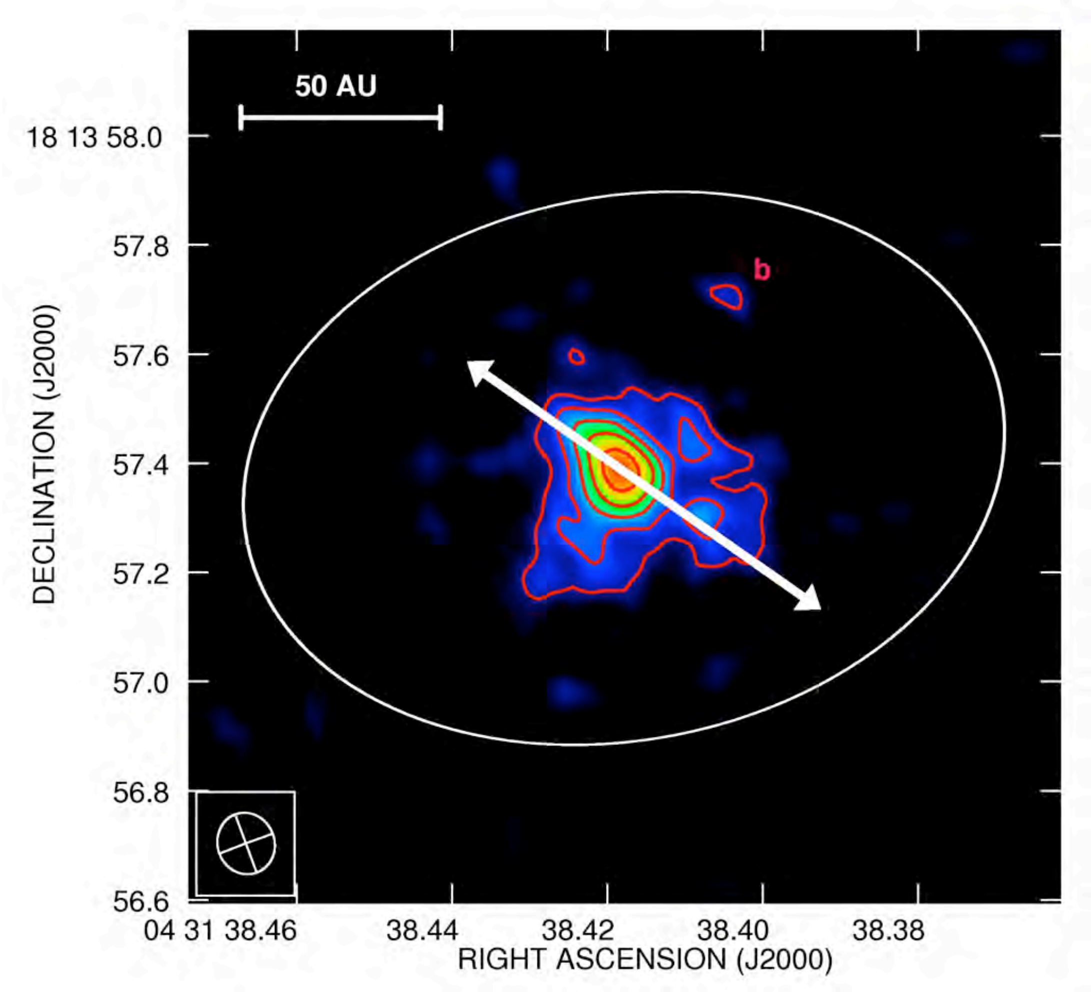
- It explains the similarities between stars, brown dwarfs, and gas giants
- It is fast enough to explain the gas envelopes of the giant planets

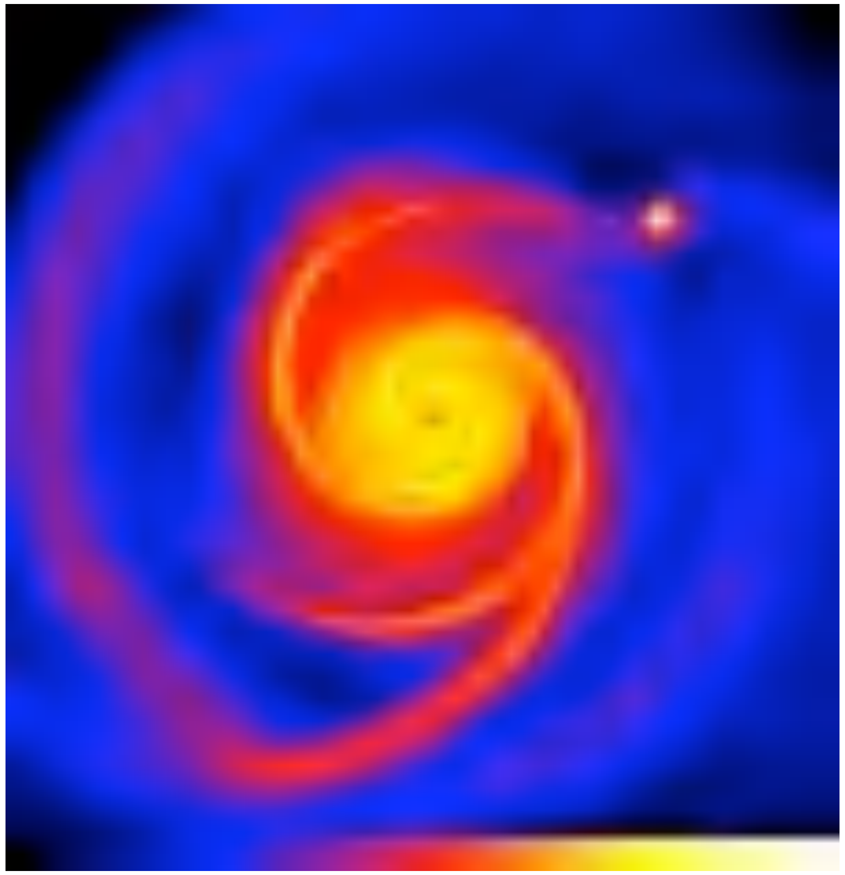
Disadvantages of the gas-instability hypothesis:

- No explanation for the heavy element enrichment of the gas giants
- The instabilities require a very heavy protoplanetary disk (~ 1 M_{Sun})
- The settling of the heavy elements is not straightfoward
- No explanation for asteroids, moons, and comets

Gas-instability hypothesis

Tentative (radio)detection of a protoplanet around the young ($<10^5$ yr) star HL Tau that could have formed out of gas-instabilities:



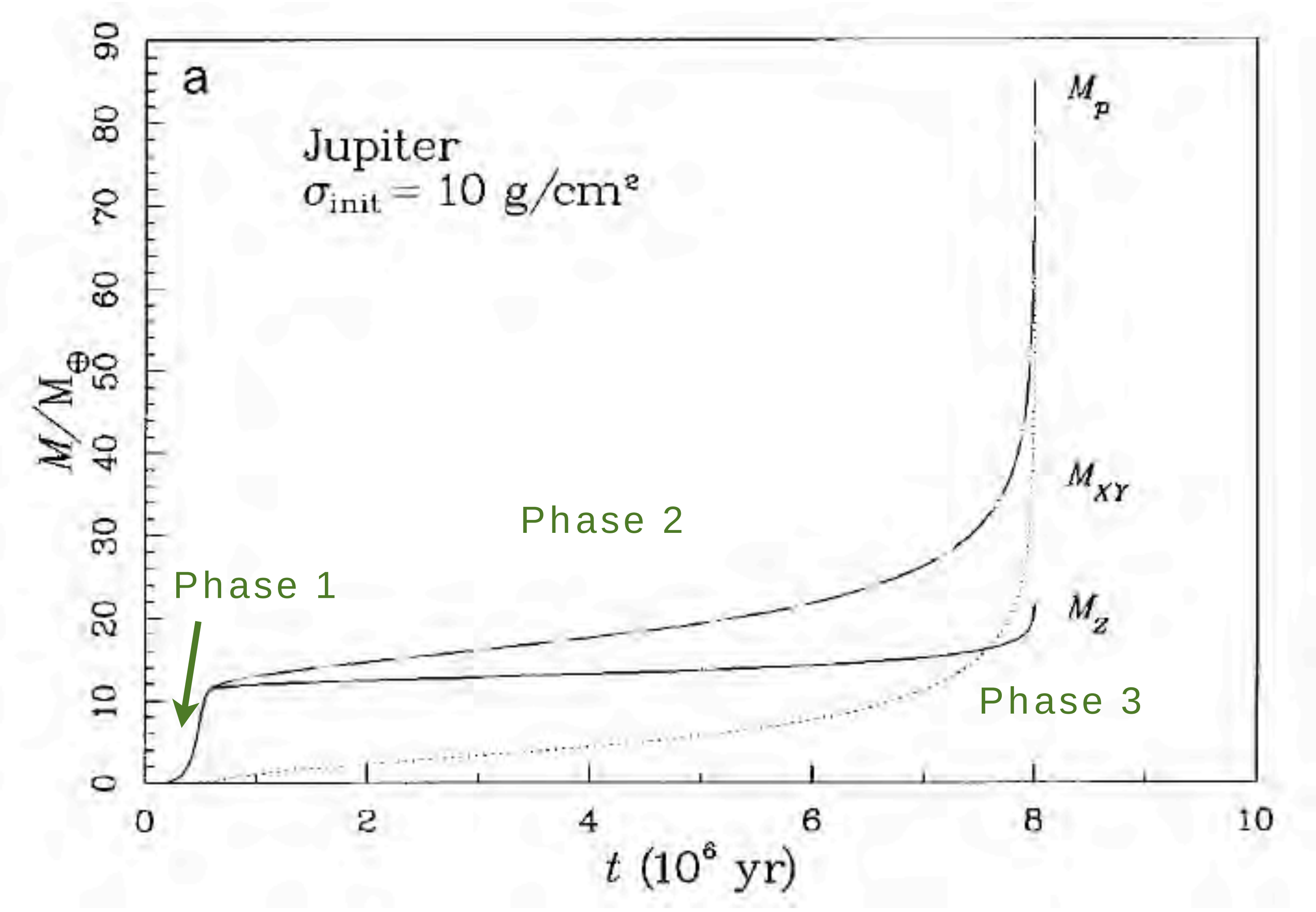


A numerical simulation of the clump’s formation

Another star, XZ Tau, may have passed HL Tau about 16,000 yr ago, triggering the instabilities (Greaves, Richards, Rice, and Muxlow, 2008)

Core-instability hypothesis

The core of a giant planet forms by accretion of planetesimals, and its huge atmosphere by run-away accretion of stellar nebulae gas:



From Pollack et al. (1996): M_p is the total planetary mass, M_Z the mass of the core, and M_{XY} the mass of the atmosphere. The density of the protoplanetary disk is 10 g/cm^3.

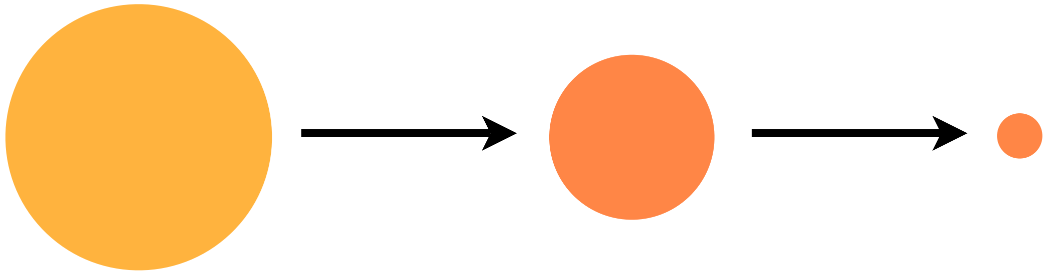
Phase 1 (~ 5.10^5 yr):
The planet accumulates solids by rapid run-away accretion, until it has severely depleted its feeding zone

Phase 2 (~ 7.10^6 yr):
The planet accumulates solids and volatiles at similar rates

Phase 3:
The planet accumulates volatiles by rapid run-away accretion

Core-instability hypothesis

At the end of the run-away gas accumulation phase, the planet cools down, the pressure decreases, and the planet starts to contract. This contraction releases gravitational energy, which is emitted as radiation:



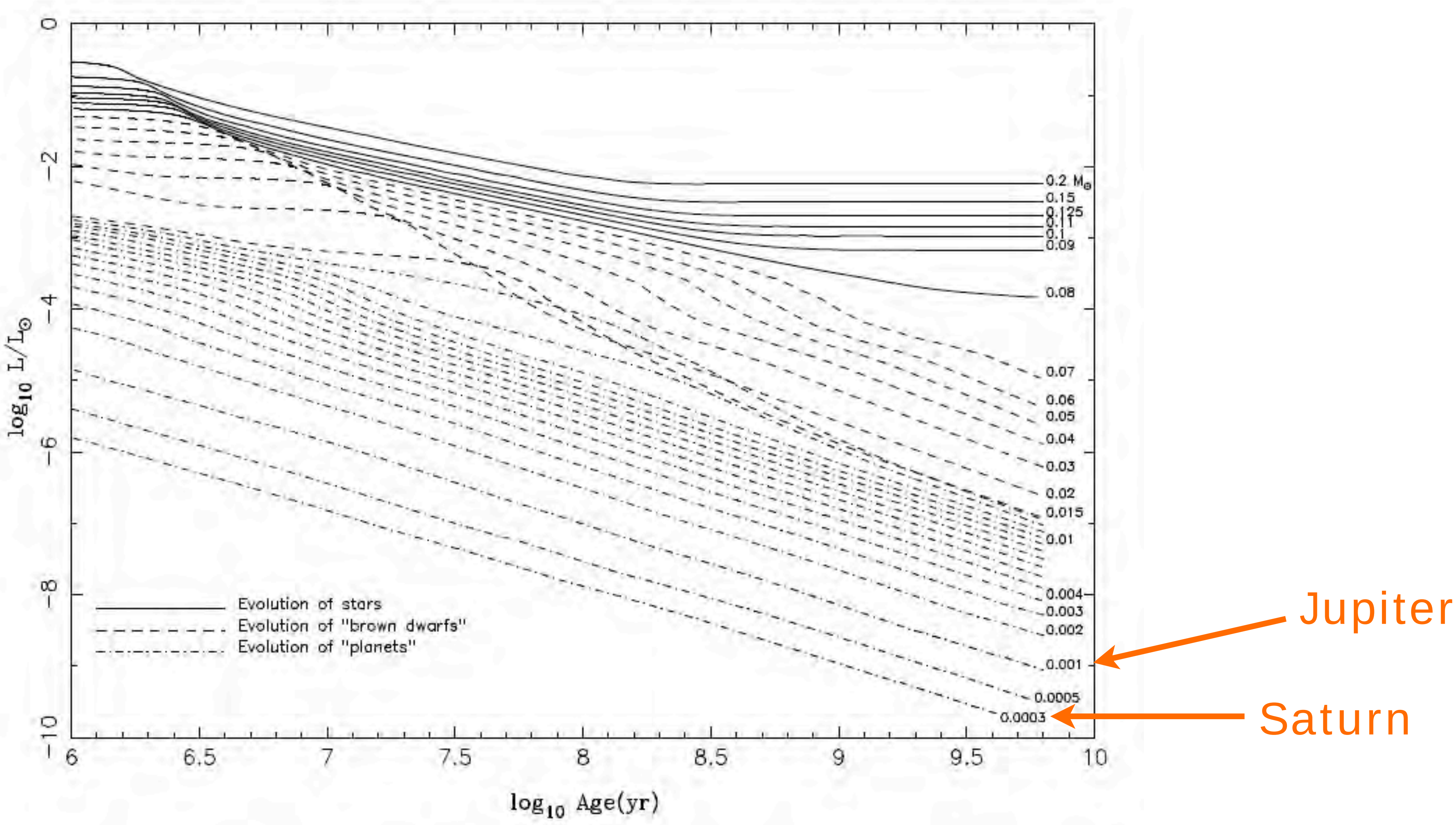
The ratio of the gravitational energy E_G to the planet's luminosity L gives the initial 'Kelvin-Helmholtz' time-scale t_{KH} of the contraction:

$$t_{KH} = \frac{E_G}{L} \sim \frac{GM^2}{RL}$$

With G the gravitational constant, M the planet's mass, and R its initial radius.

Core-instability hypothesis

After ~10^4 years for Jupiter and Saturn, and ~10^5 years for Uranus and Neptune, the contraction slowed down, and the temperature and luminosity decreased:



From Burrows et al. (1997): The evolution of the luminosity of stars, brown dwarfs and planets measured in Solar luminosities versus time after formation.

Core-instability hypothesis

“Bottom-up planet formation”

Advantages of the core-instability hypothesis:

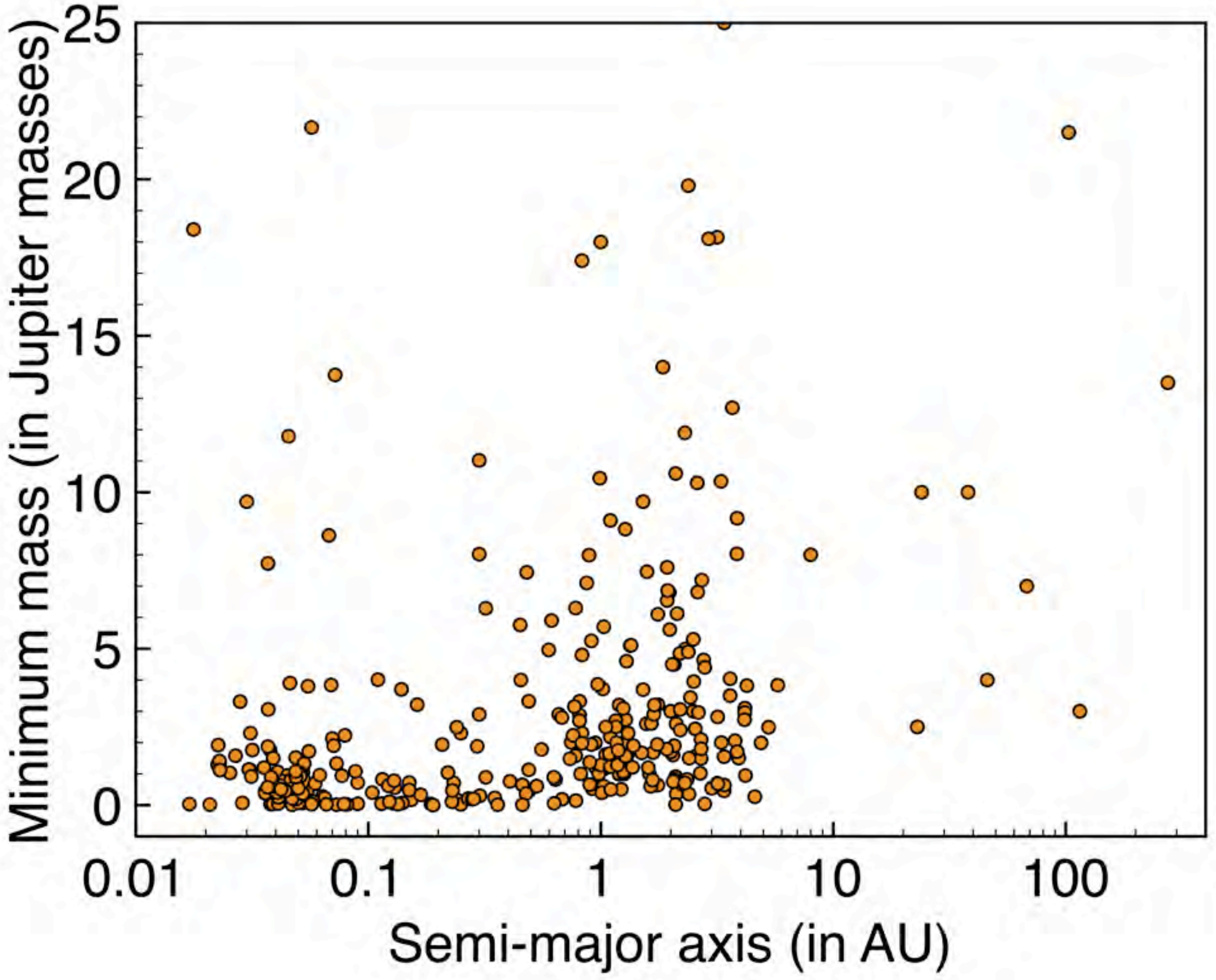
- Seems the best mechanism for terrestrial planet formation
- Leads to large core masses
- Can explain non-solar compositions (gas giants and ice giants)
- Would lead to circular orbits due to random accretion of planetesimals

Disadvantages of the core-instability hypothesis:

- Might require unrealistically large core masses
- The total time required for forming the planets might be longer than the lifetime of the disk
- Gap formation in the disk might cut-off the atmospheric accumulation

Planet migration

Many exoplanets are found to orbit their star in orbits much tighter than common in the Solar System*:



Planets and planetesimals are thought to form in the outer regions of a circumstellar disk, and then to migrate inward (sometimes outward) because of gravitational interaction with the gas and dust in the disk.

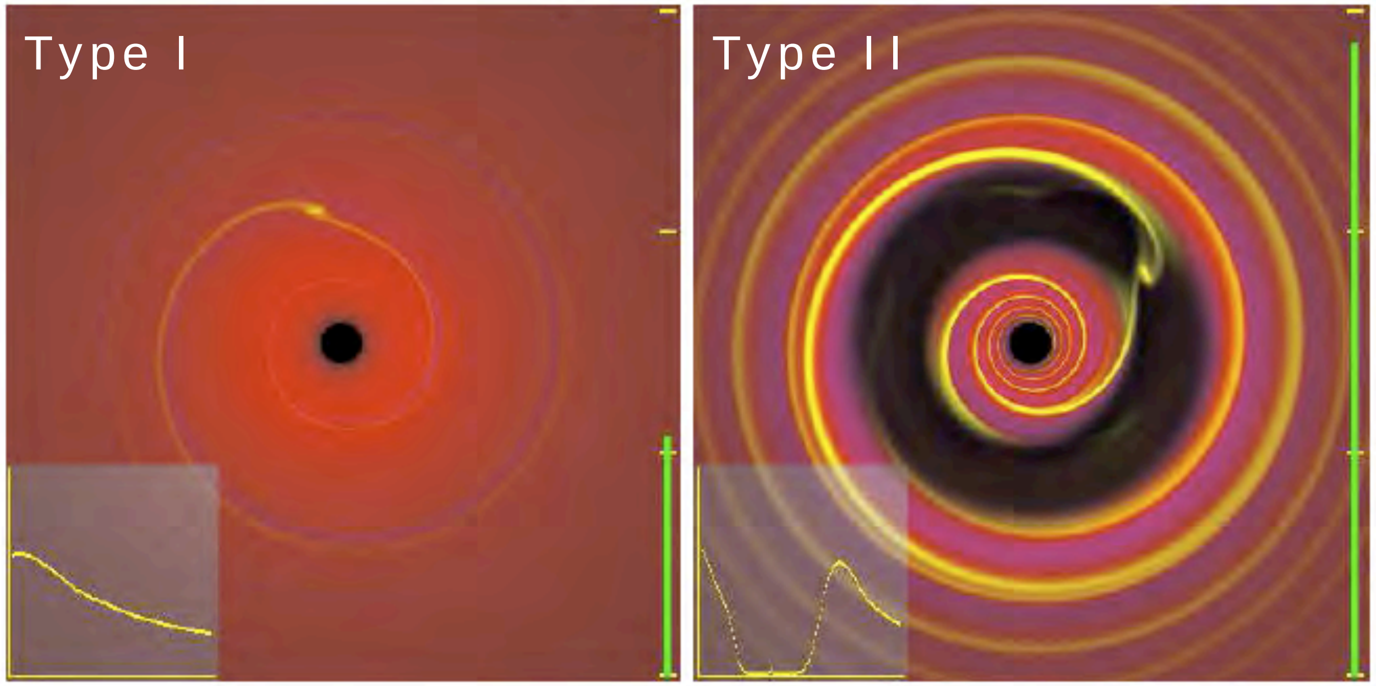
*This can very well be an observational bias! In time, more planetary systems like our own can be discovered

Planet migration

Orbital migration of a solid body involves a change of angular momentum to either gas or other solid bodies in the system, most likely due to gravitational interactions.*

Besides gravitational scattering, we distinguish two types of migration:

Type I: the planetary mass is smaller than the disk
Type II: the planetary mass is large enough to open up a gap

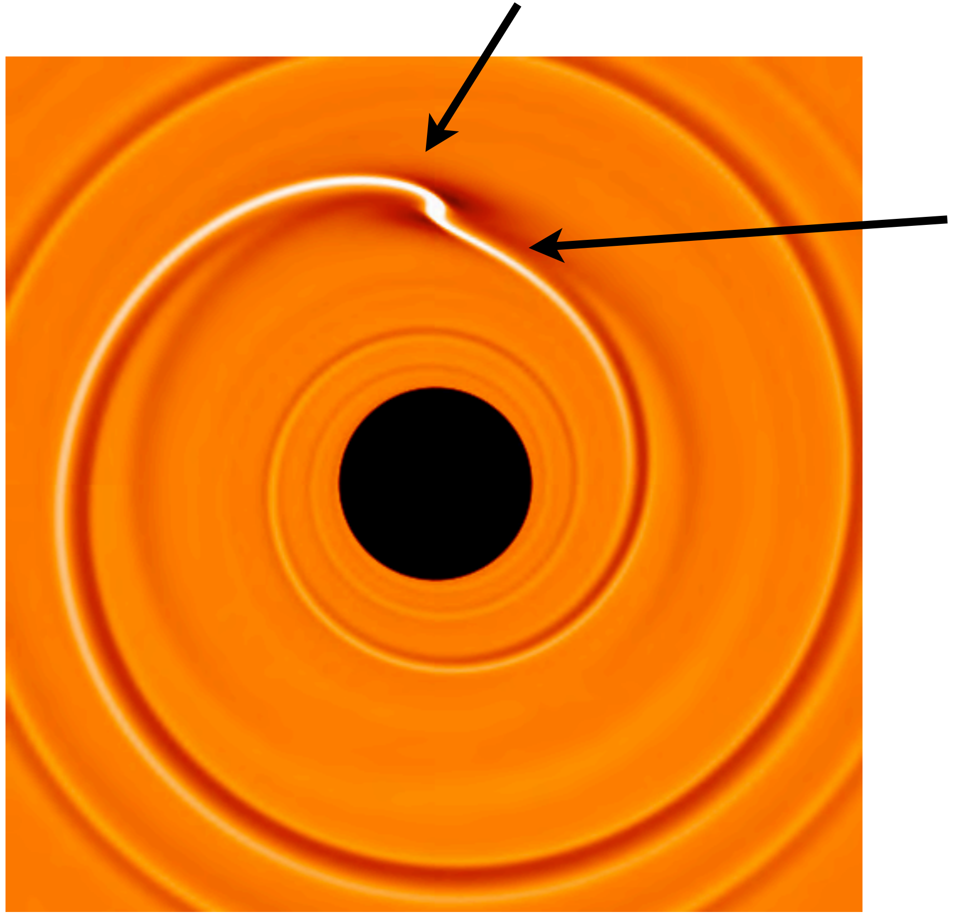


Armitage & Rice (2005)

* Aerodynamic drag, which is important for the orbital evolution of meter-sized planetesimals, is negligible for planetary masses

Type-I migration

The outer disk region orbits slower than the planet, and the gravitational attraction of the spiral waves tries to slow down the planet, and thus move it inwards.



The inner disk region orbits faster than the planet, and the gravitational attraction of the spiral waves tries to speed up the planet, and thus move it outwards.

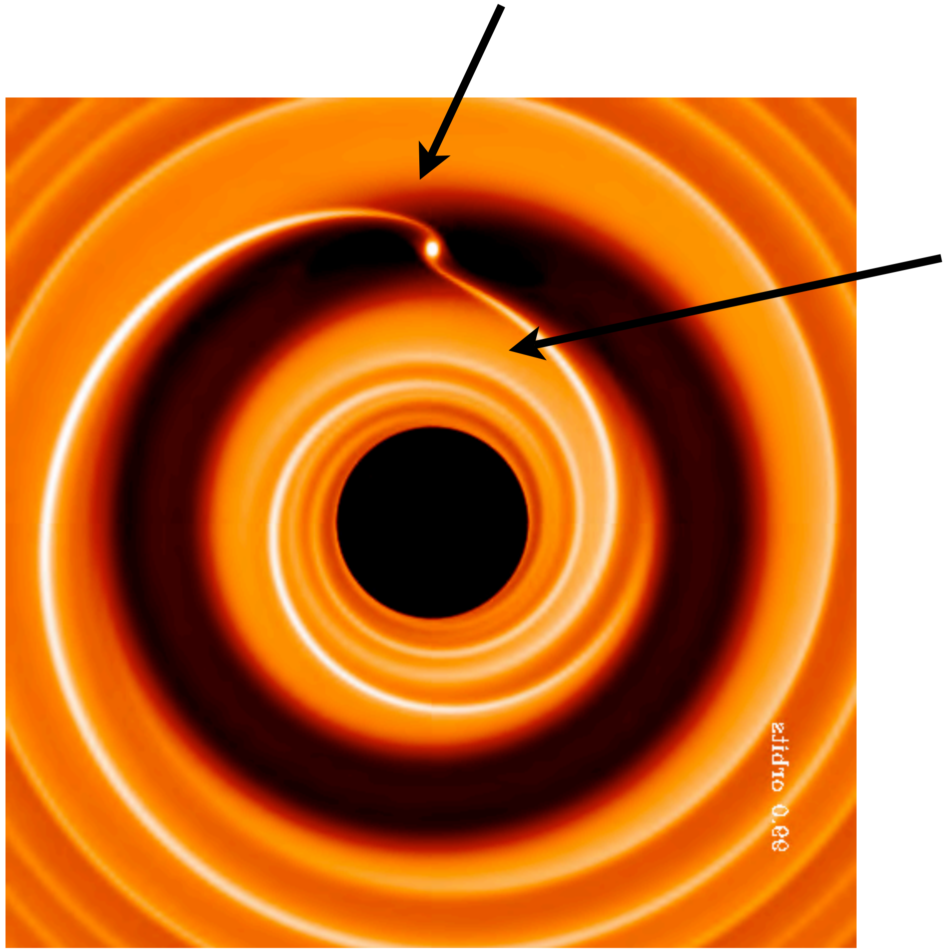
F. Masset (www.maths.qmul.ac.uk/~masset/moviesmpegs.html)

The inward force is generally larger than the outward force … the planet thus moves towards its star … fast!

The time scale for a planet with mass M_p is ~ $10^5\ (M_\mathrm{p}/M_\oplus)^{-1}$ yr

Type-II migration

The outer disk region tries to accrete onto the star: if the disk is heavy enough it will push the planet towards the star



The inner disk region accretes onto the star and will deplete if the planet is not spiralling towards the star

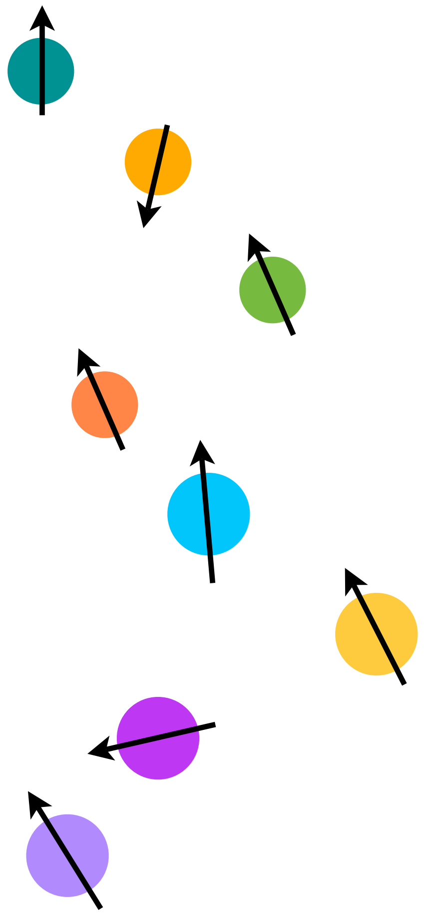
F. Masset (www.maths.qmul.ac.uk/~masset/moviesmpegs.html)

The time scale for type II migration is much longer than for type I …

Planetary rotation

The sidereal rotation periods and obliquity angles (measured with respect to their orbit) of the Solar System planets:

Planet	rotation period	obliquity angle
Mercury	58.646 days	0.1°
Venus	243.019 days	177.3°
Earth	23.93 h	23.45°
Mars	24.62 h	25.19°
Jupiter	9.92 h	3.12°
Saturn	10.66 h	26.73°
Uranus	17.24 h	97.86°
Neptune	16.11 h	29.56°



From C.F. Yoder, Astrometric and geodetic properties of Earth and the Solar System. In *Global Earth Physics, A handbook of Physical Constants*, AGU Reference Shelf 1, American Geophysical Union [1995]

For the gaseous planets, the rotation periods indicate the rotation periods of their magnetic fields

The origin of planetary rotation

A planet's direction of rotation and its obliquity angle are determined during the accumulation stage, probably by a few large impacts. Why the prograde*** rotations?

- the hydrodynamic accretion of the atmospheres of the gas giants would lead to prograde rotation
- accretion from a uniform, dynamically cold disk of planetesimals would give a rocky planet a slow retrograde rotation
- rapid prograde rotation of a rocky planet can be achieved when a planet opens up a gap in the disk and accretes material from the edges of its feeding zone
- a slow decay of planetesimal orbits towards the protoplanet due to drag, would give a prograde rotation
- the observed rotation might still be a chance occurrence …

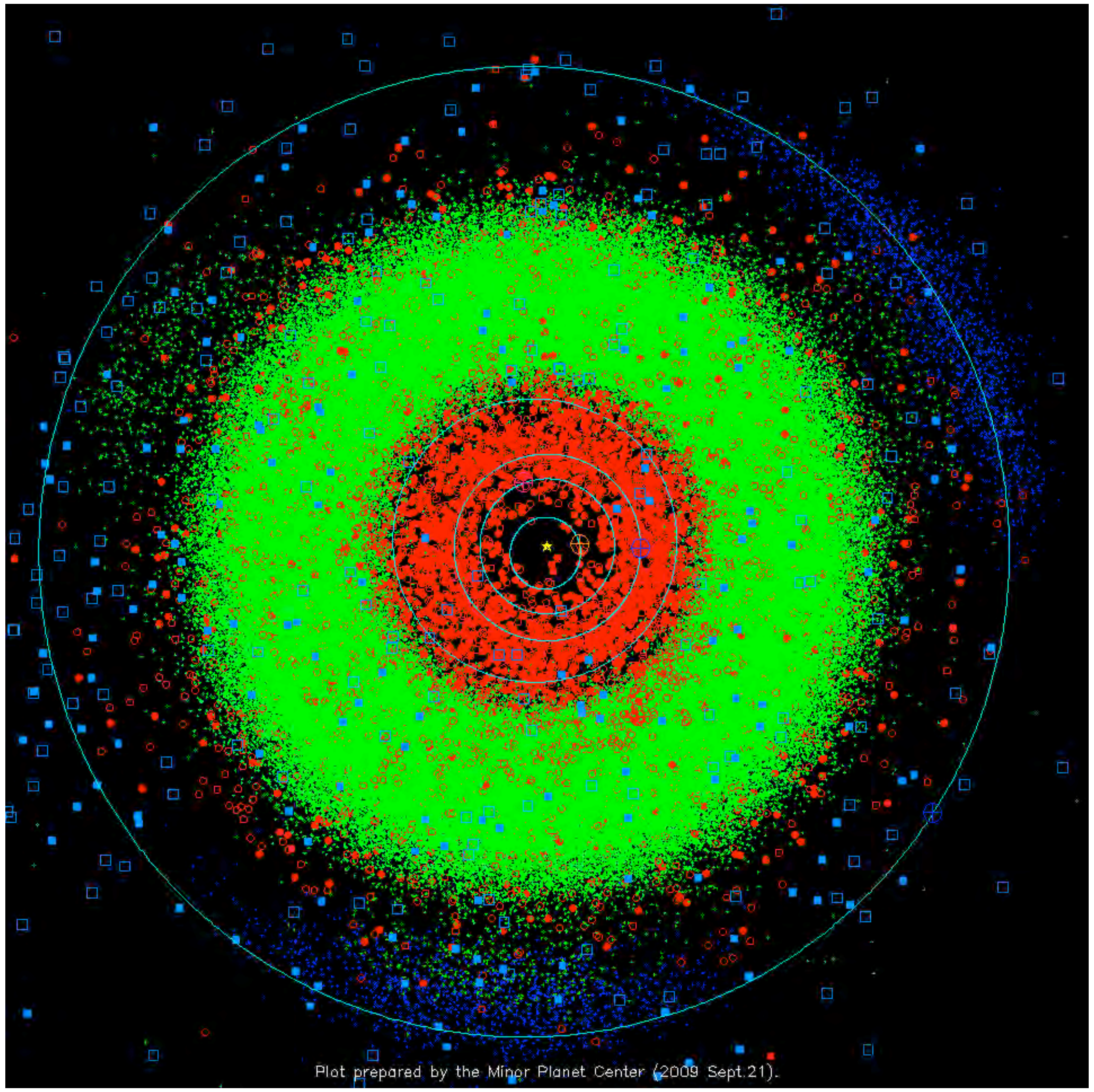
* The rotation of Mercury is strongly influenced by gravitational interaction with the Sun (it is in a 3:2 spin-orbit resonance)

* The retrograde rotation of Venus can be explained by interaction with its thick atmosphere

* The obliquity of Uranus is usually attributed to a giant collision (GC) with an Earth-sized planetesimal at the end of the accretion

Origin and evolution of asteroids

Asteroids or minor planets can be viewed as remnant planetesimals although some have been melted and there has been substantial collisional evolution among bodies in the main asteroid belt.



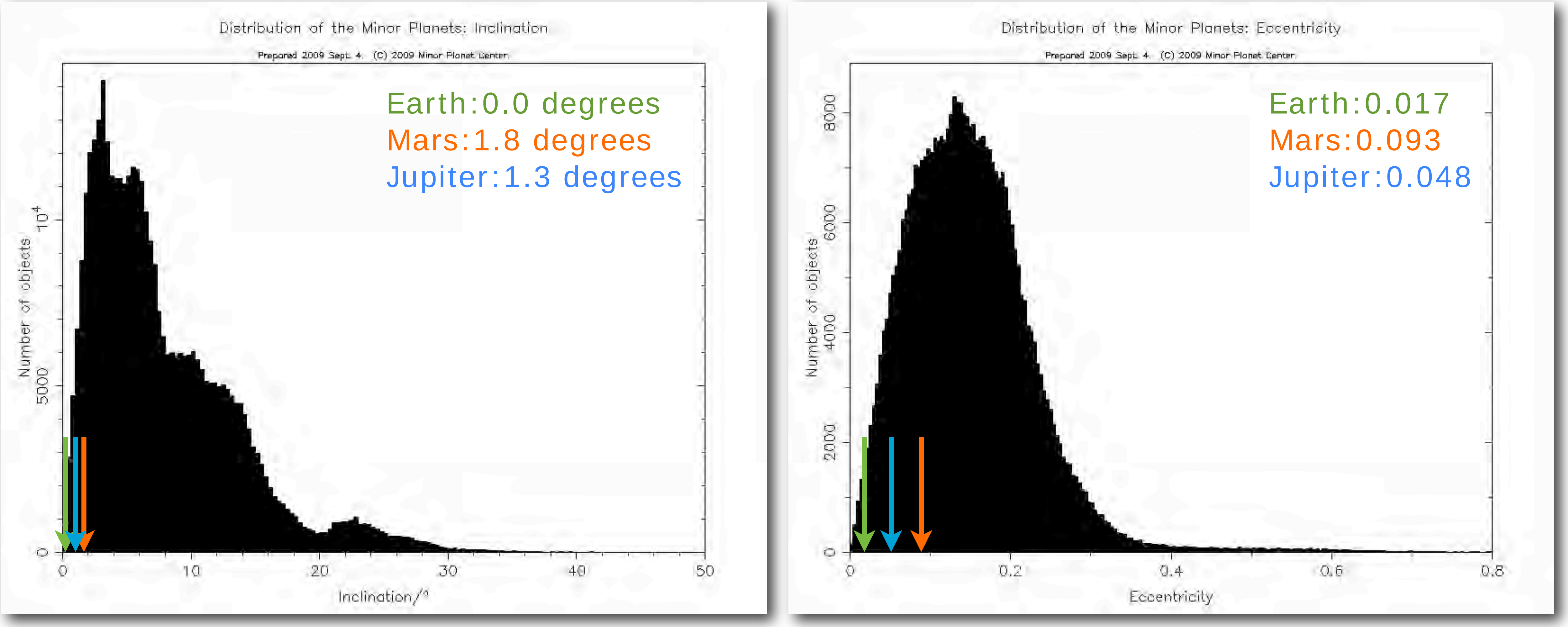
Current asteroid locations projected onto the ecliptic plane. The orbits of Jupiter and the inner planets are shown. The blue squares are comets. Credit: IAU Minor Planet Center

- There are more than 10,000 known asteroids
- They show a large diversity in composition
- The largest asteroid, Ceres, has radius R ~470 km*
- Number N of asteroids varies with R as $N_{R+dR} \sim R^{-3.5}$
- Total mass in asteroid belt ~ 5.10^{-4} $M_\oplus$
- Too little mass to form a full-sized planet
- Less mass than expected at these orbital distances

* A rocky body with a typical density of 3.5 g/cm^3 is approximately spherical if R > 350 km

Origin and evolution of asteroids

Gravitational disturbances of Jupiter are most likely responsible for the mass depletion in the asteroid belt. This would also explain the observed orbital properties of asteroids (large inclination angles and high eccentricities):



The distribution of orbital inclination angles (left) and eccentricities (right) of known asteroids. Also indicated: the inclination angles and eccentricities of the Earth, Mars, and Jupiter. Credit: IAU Minor Planet Center

Origin and evolution of comets

Comets consist primarily of water-ice and dust, but more volatile species (S_2, N_2^+, CO) have also been detected. Most comets must have formed at the outer edges of the protoplanetary disk, > 20 AU, where $T < 20 - 30$ K.

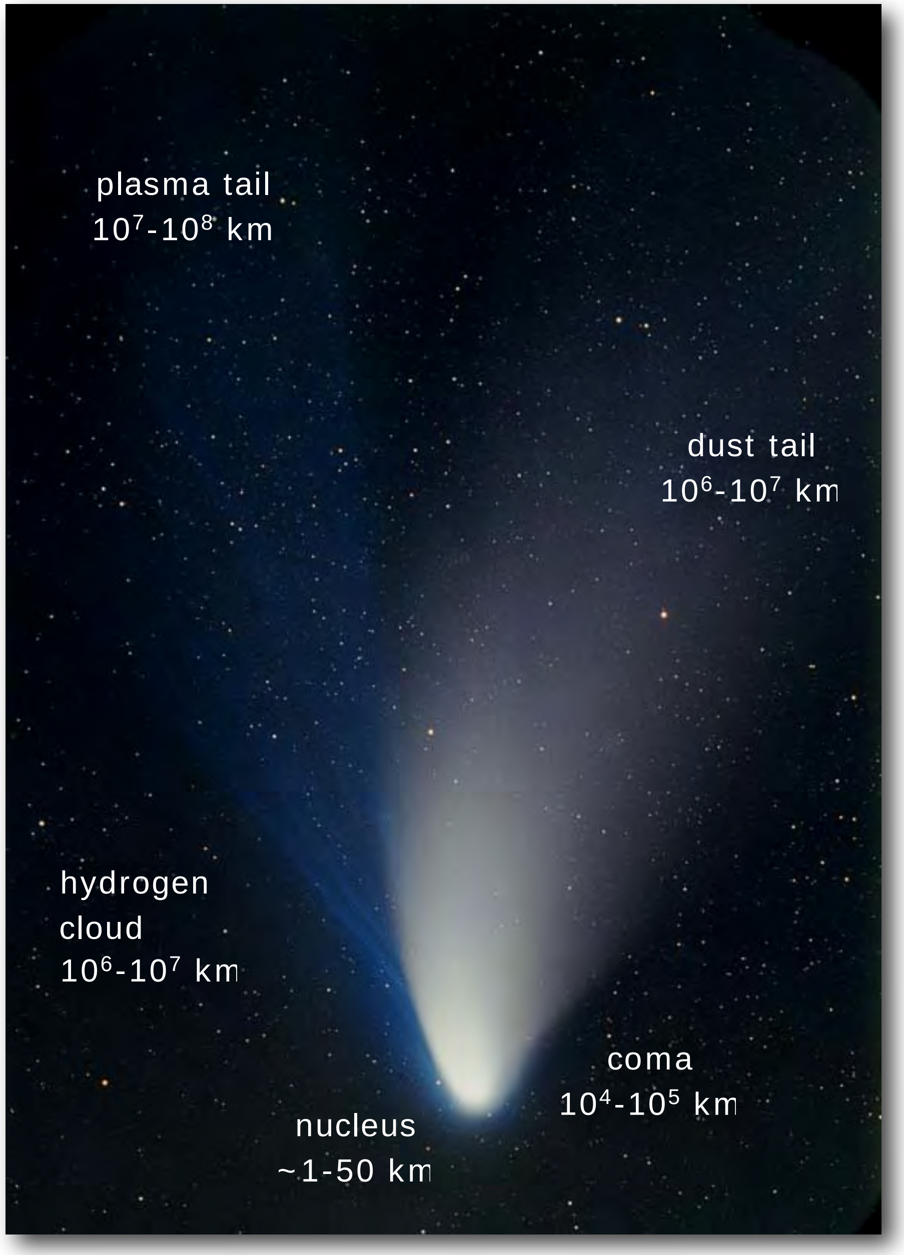
The main comet reservoirs in the Solar System:

The Kuiper Belt*

- 40-50 AU
- Short Period (SP) (< 200 yr) comets
- $10^9 - 10^{10}$ objects (total mass $< 1\ M_\oplus$)

The Oort Cloud

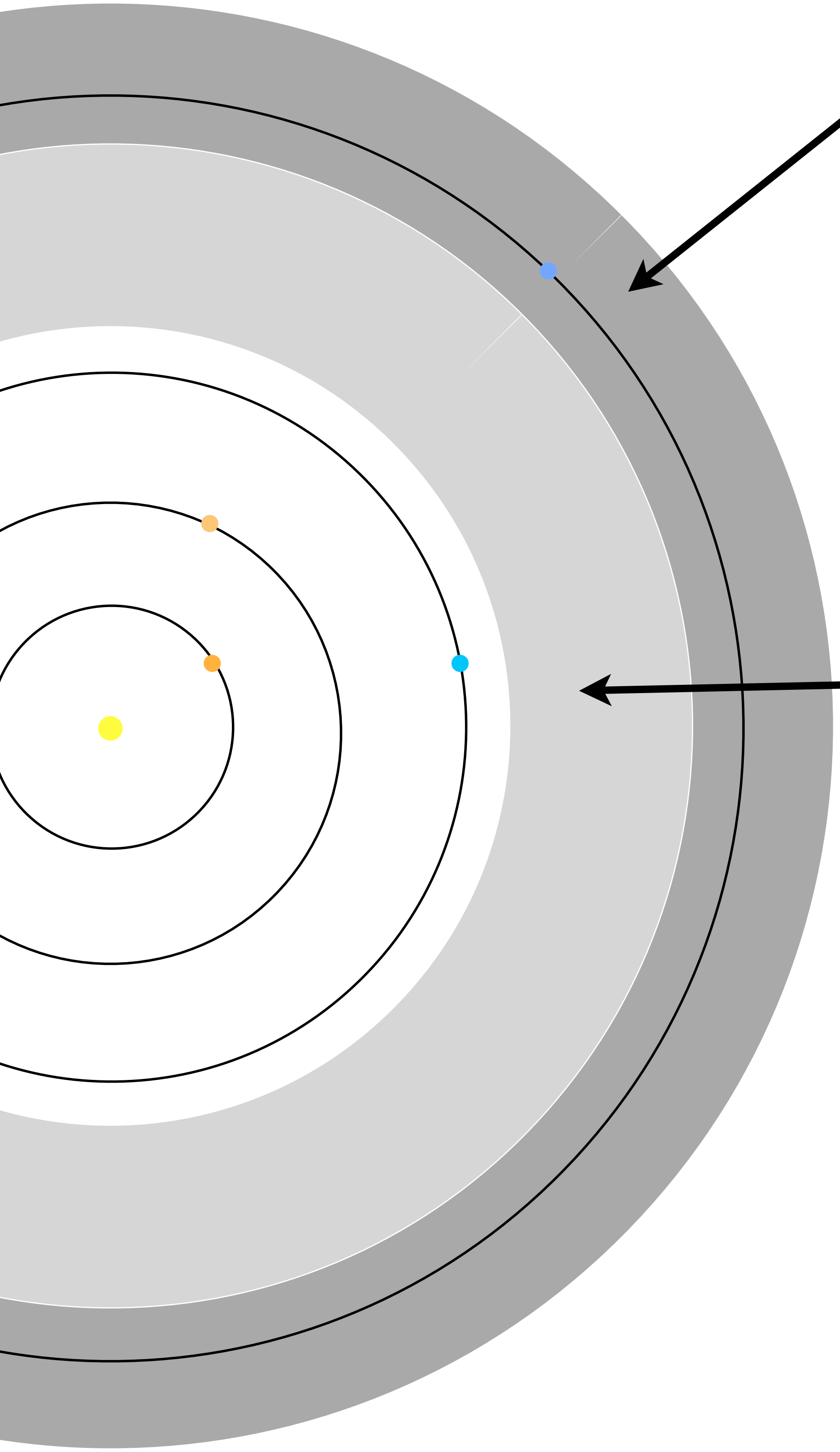
- $> 10^4$ AU
- Long Period (LP) (> 200 yr) comets
- $\sim 10^{12}$ objects



Comet Hale-Bopp (1997)
Credit: Loke Kun Tan/starryscapes.com

* Sometimes called the Edgeworth-Kuiper Belt

Origin and evolution of comets



origin of Kuiper Belt comets

- current gravitational perturbations by the outer planets and local collisions

origin of Oort cloud comets

- gravitationally perturbed inwards by Uranus & Neptune, then outwards mainly by Jupiter
- current gravitational perturbations by passing stars and galactic tides

Planetary satellites

Regular satellites:

- prograde, low eccentricity orbits near the planet’s equatorial plane
- well within the planet’s Hill sphere*

Formed in-situ from circumplanetary disk

Irregular satellites:

- prograte or retrograde, high eccentricity, high inclination orbits
- outside the regular satellites system
- usually quite small (non-spherical)

Formed elsewhere, were captured by the planet

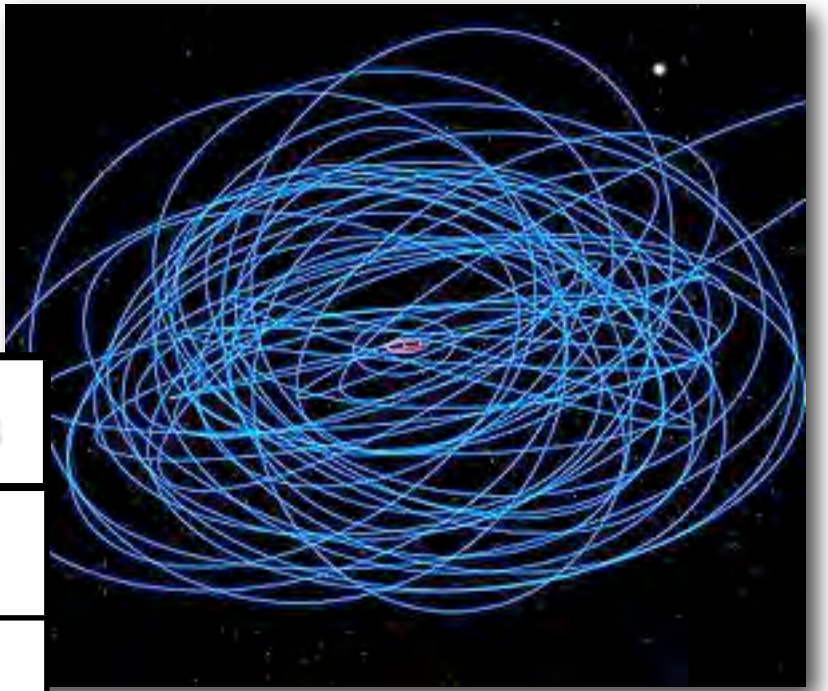
*The Hill sphere is the area (with radius R_H) in which the gravity of a secondary body (m_2) exceeds that of a primary body (m_1) that is located at distance a:

$$R_H = \left(\frac{m_2}{3(m_1 + m_2)}\right)^{1/3} a$$

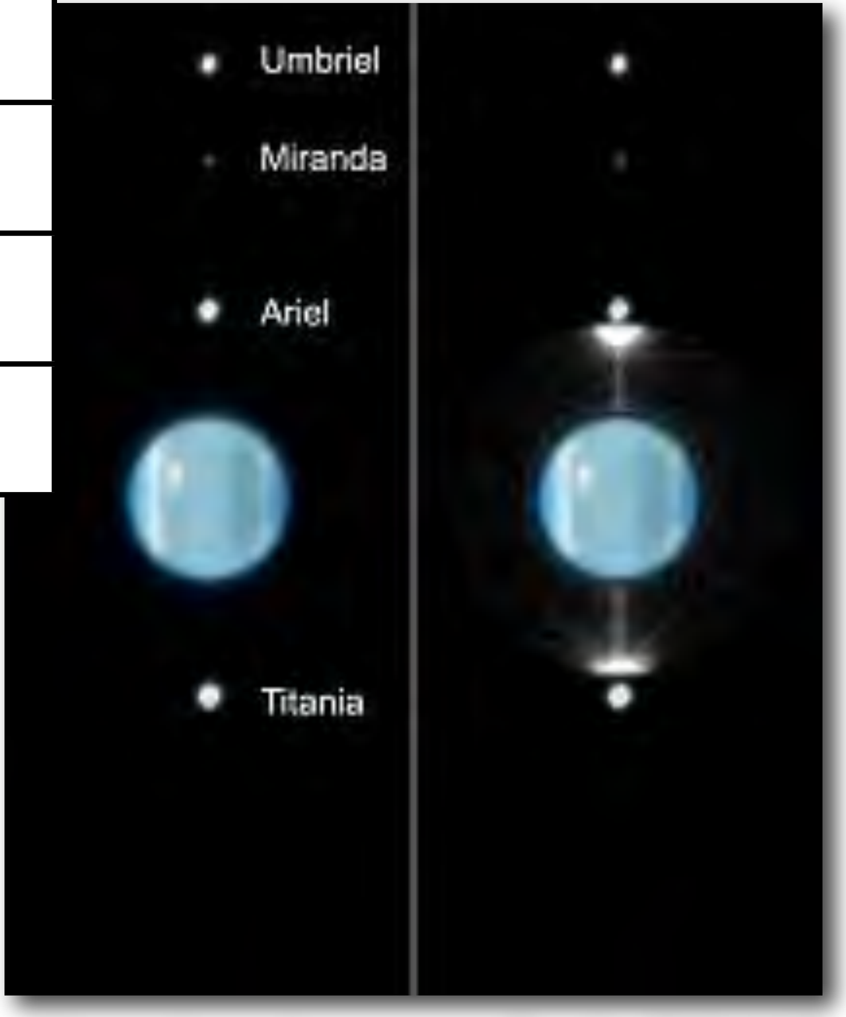
Planetary satellites

Satellites of the Solar System planets:

Planet	**# regular satellites**	**# irregular satellites**
Earth	1 (Moon)	-
Mars	-	2 (Phobos & Deimos)
Jupiter	8 (the Amalthean and Galilean moons)	55
Saturn	23 (largest: Titan)	38
Uranus	18 (largest: Titania)	9
Neptune	6	7 (largest: Triton)



The moons of Saturn



Uranus' regular moons. Credit: ESO/VLT

An example of an irregular satellite:

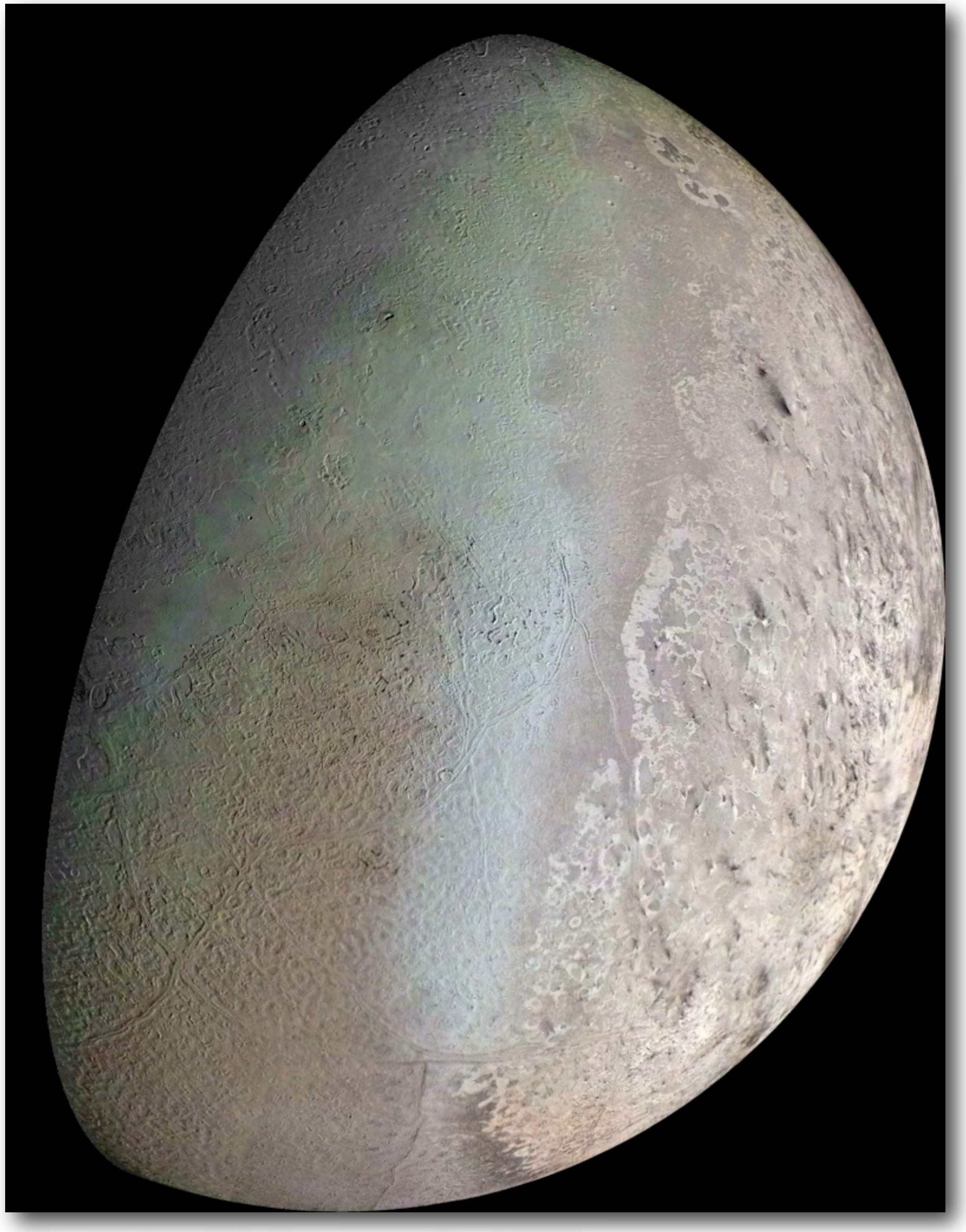
A false-colour image of Saturn’s moon Hyperion obtained during a fly-by of Cassini at a distance of ~60,000 km:

Credit: NASA/JPL/ Space Science Institute (Sept. 26, 2005)

Year of discovery	1848
Mass	$1.8 \cdot 10^{19}$ kg ($3 \cdot 10^{-6}$ $M_\oplus$)
Dimensions	360x280x225 km
Mean density	1.4 g/cm^3
Mean orbital radius	1,481,000 km
Orbital period	21.3 days
Orbital direction	prograde
Orbital eccentricity	0.1042
Orbital inclination angle	0.43°
Rotational period	chaotic
Geometric albedo	0.3
Visual magnitude	14.19

Another irregular satellite:

A false-colour image of Neptune’s largest moon Triton:



Credit: NASA/Voyager 2 (1989)

Year of discovery	1846
Mass	$2.1\cdot10^{22}$ kg (0.004 $M_\oplus$)
Diameter	2700 km
Mean density	2.06 g/cm^3
Mean orbital radius	354,760 km
Orbital period	5.877 days
Orbital direction	retrograde
Orbital eccentricity	0.000016
Orbital inclination angle	157°
Rotational period	synchronous (5 d,21 h)
Geometric albedo	0.76
Visual magnitude	13.47