Data Science For Astronomers

Image courtesy of www.ska.ac.za

IDIA Inter-University Institute for Data Intensive Astronomy

UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

ARC AFRICAN RESEARCH CLOUD

SKA AFRICA

Data Science for Astronomers AST5004Z

- Lecturers
  - Dr. Bradley Frank, bradley.frank@uct.ac.za
    - Office 535, RW James.
  - Prof. Russ Taylor, russ@ast.uct.ac.za
  - Prof. Patrick Woudt, patrick@ast.uct.ac.za
- Mondays/Wednesdays, 2-4pm.
- 24 Lectures, including a few tutorials.

Data Science for Astronomers
AST5004Z

- Assessment
  - 50% Continuous Assessment.
    - Computational Tutorials.
    - Project.
  - 50% Exam.

Data Science for Astronomers
AST5004Z

- Tutorials: Inter-University Institute for Data Intensive Astronomy, IDIA
- Access:
  - https://idia-pipelines.github.io/access/
  - via Jupyter-Hub or SSH.
  - I leave it to you to figure out how to access the machines :)
- Projects: Astro department will release possible project descriptions soon.

Data Science for Astronomers AST5004Z

- Introduction to astronomy, telescopes and astronomical observations. (2 lectures)
  - The universe in an nutshell for the layman- planets, stars, galaxies, large-scale structure, cosmology.
  - Radio and optical telescopes.
  - International facilities/projects on the ground and in space.
- Astronomical data and data bases. (4 lectures)
  - Astronomical/astrophysical quantities.
  - What does astronomical data look like images, time series (light curves), catalogues.
  - Major astronomical databases.
  - International virtual observatory and online data resources and tools (maybe a small assignment around this).
- Large-scale simulations. (2 lectures)

Data Science for Astronomers AST5004Z

- Software systems. (4 lectures)
  - Python, Jupyter notebooks
  - Processing environments such as casa.
  - Optical processing systems.
- Data to information. (12 lectures)
  - Data processing in optical and radio to images and catalogues.
  - Automation and pipelines.
  - South African development and data challenges (description of MeerKAT LSPs and data issues?)
  - Image processing and information extraction.
  - Source fitting, photometry, noise.
  - Databases, data mining.
  - Visualization and visual analytics.
  - Areas of exploration of machine learning and cognitive computing.
  - Cloud computing and big data.
  - Advanced problems in astronomical data science.

Data Science for Astronomers AST5004Z

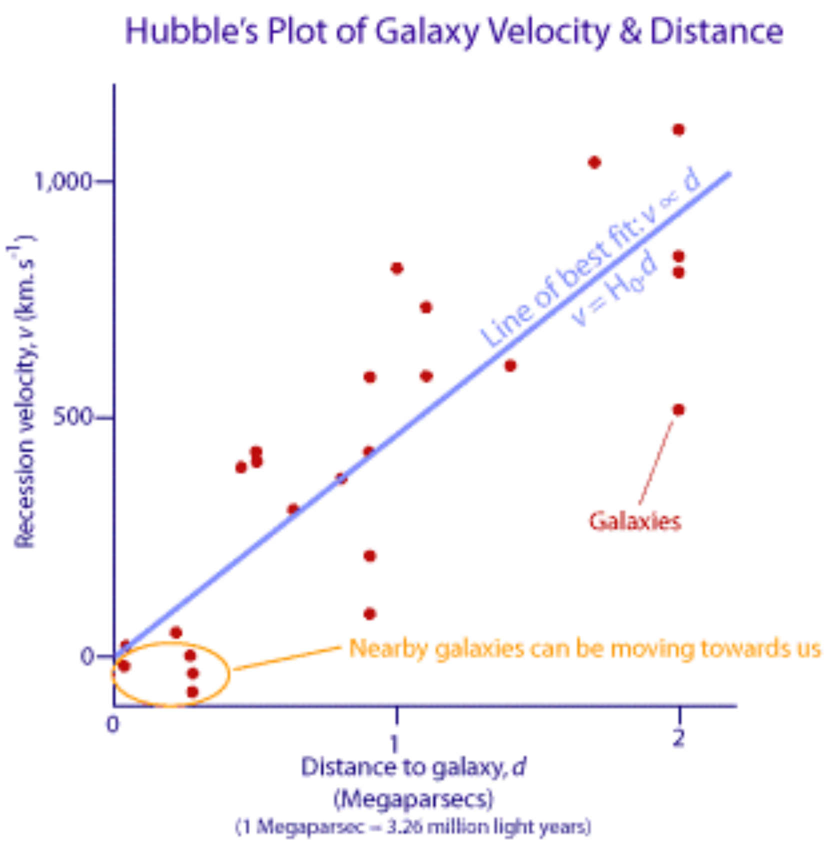
- Aspects of data intensive astrophysics
- Data visualisation and complex databases, advanced statistical tools for astronomical data analysis and computational astrophysics.
- Examples in modern data-intensive astrophysics derived from the global data challenges around MeerKAT, the Square Kilometre Array (SKA).
- Associated projects in radio astronomy, and other large multi-wavelength surveys.
- Use of Bayesian statistics in astronomy.
- Understand the complexity of visualising large data cubes, optimising database operations in the presence of multi-dimensional data, data mining and discovery tools, and the role of large-scale simulations to interpret the significance of astronomical observations.

Astronomy in Southern Africa

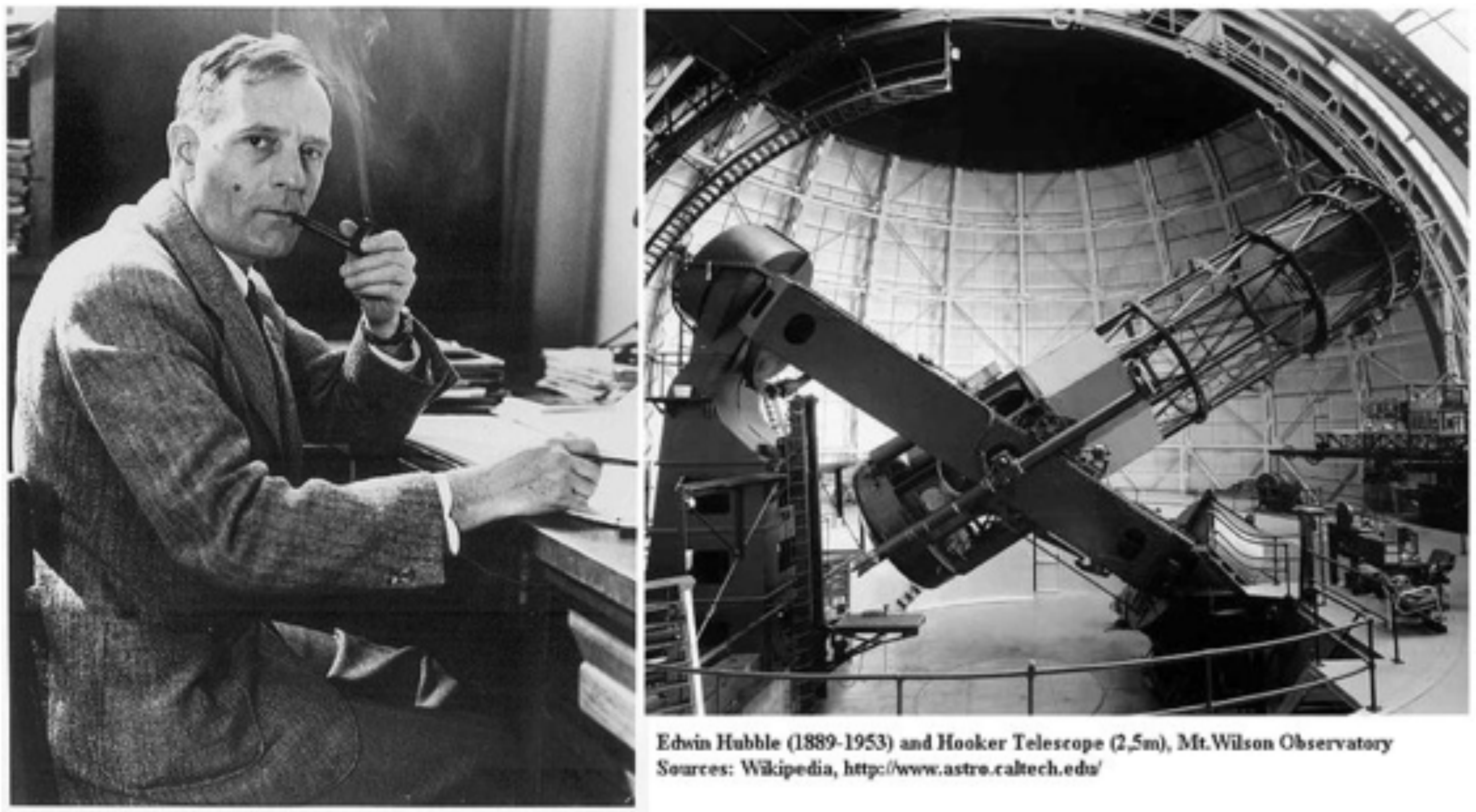
- South African Astronomical Observatory, SAAO.
  - Operations SALT, IRSF, 1.9m, 1.5m.
  - Hosts MeerLICHT.
- South African Radio Astronomical Observatory, SARAO.
  - Hartebeeshoek Radio Astronomical Observatory, HartRAO.
  - KAT-7 and MeerKAT telescopes.
- High Energy Stereoscopic System, HESS.
- South African Space Agency, SANSA.

Why Data Science for Astronomers?

- Astronomy — a driver for data science.
- Data science before data science.



http://www.atnf.csiro.au/



https://unendlicheweiten.files.wordpress.com/

IDIA Inter-University Institute for Data Intensive Astronomy
UNIVERSITY OF CAPE TOWN IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD
ARC
SKA AFRICA

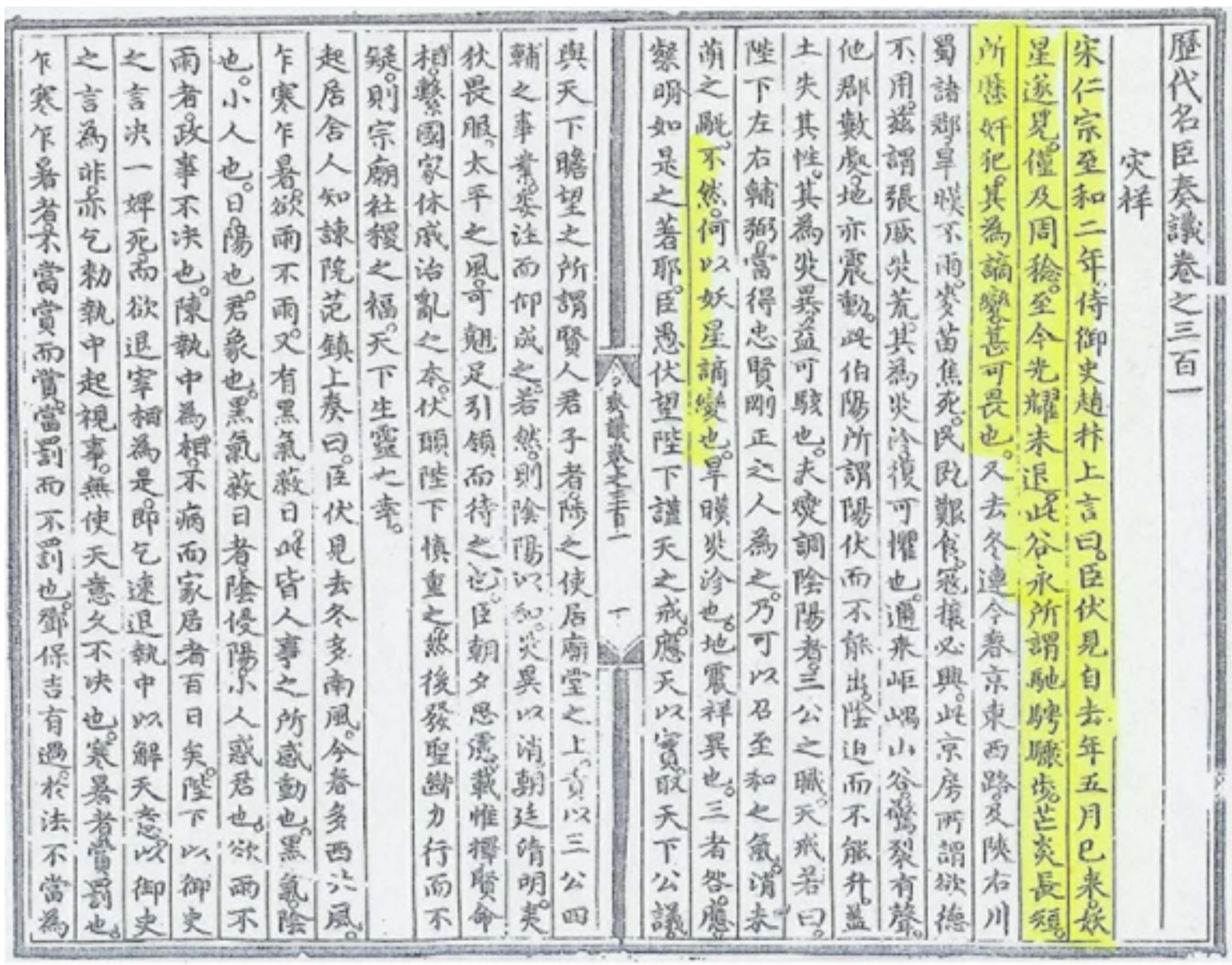
Why Data Science for Astronomers?

- Computers before computers: NASA Human Computers.

http://cdn.history.com

Why Data Science for Astronomers?

- Lots of *historical* data.


歷代名臣奏議卷之三百一

災祥

宋仁宗至和二年侍御史趙抃上言曰臣伏見自去年五月已來妖星遂見僅及周歲至今光耀未退此谷永所謂馳騁驪虒芒炎長短所歷奸犯其為讁變甚可畏也又去冬連今春京東西路及陝右川蜀諸郡旱暵不雨麥苗焦死民既艱食寇攘必興此京房所謂欲德不用茲謂張厥災荒其為災沴復可懼也邇來岠嵎山谷鶯裂有聲他郡數處地亦震動此伯陽所謂陽伏而不能出陰迫而不能升蓋土失其性其為災異益可駭也夫燮調陰陽者三公之職天戒若曰陛下左右輔弼當得忠賢剛正之人為之乃可以召至和之氣消去蕳之䘏不然何以妖星讁變也旱暵災沴也地震祥異也三者咎應纍明如是之著耶臣愚伏望陛下謹天之戒應天以實敢天下公議與天下瞻望之所謂賢人君子者陟之使居廟堂之上責以三公四輔之事夔若注而仰成之若然則陰陽以和災異以消朝廷清明矣狄畏服太平之風可翹足引領而待之也臣朝夕思慮載惟擇賢命相繫國家休戚治亂之本伏願陛下慎重之然後發聖斷力行而不疑則宗廟社稷之福天下生靈之幸

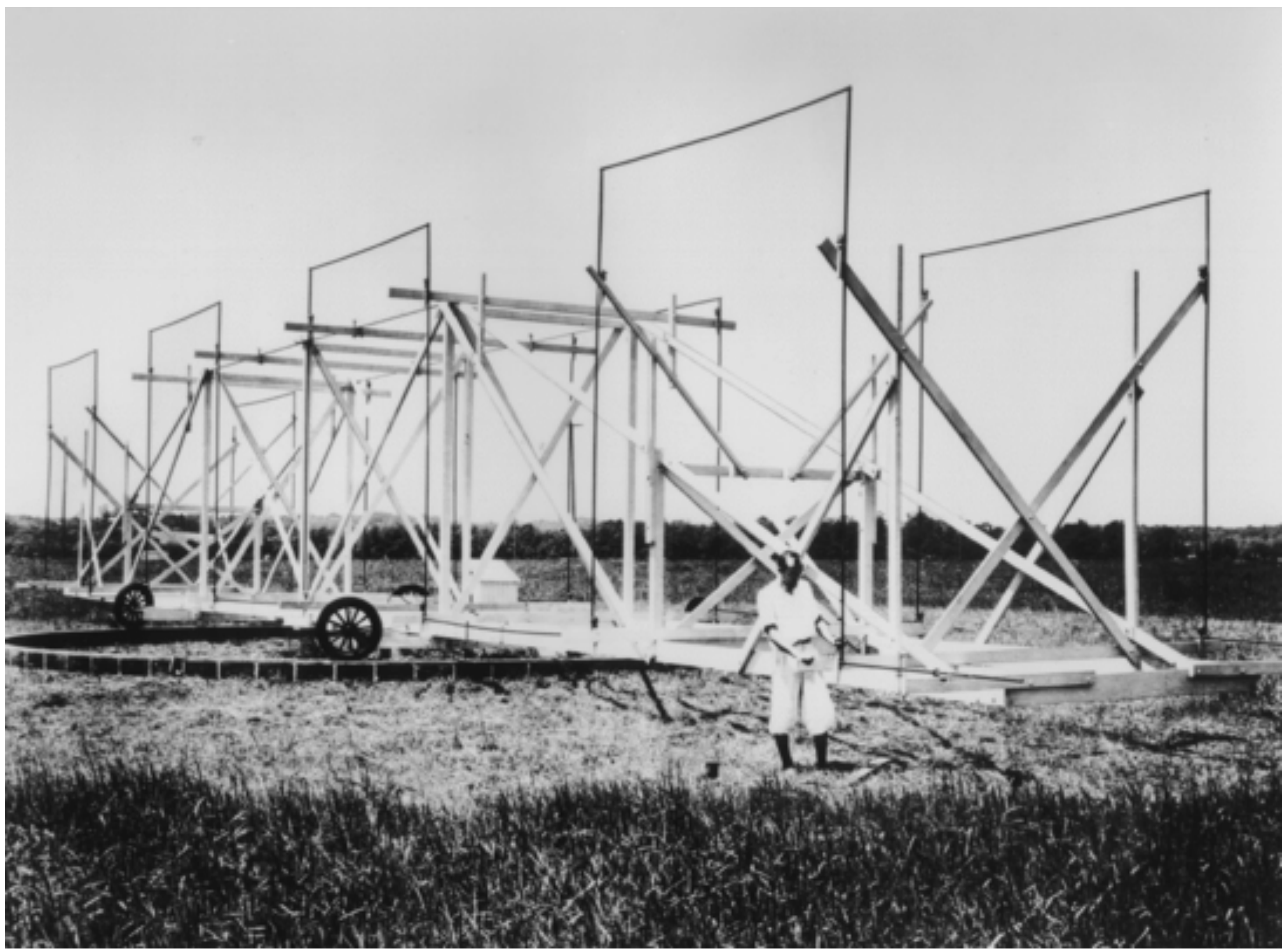
起居舍人知諫院范鎮上奏曰臣伏見去冬多南風今春多西北風乍寒乍暑欲雨不雨又有黑氣蔽日此皆人事之所感動也黑氣陰也小人也日陽也君象也黑氣蔽日者陰侵陽小人惑君也欲雨不雨者政事不決也陳執中為相不病而家居者百日矣陛下以御史之言決一婢死而欲退宰相為是耶乞速退執中以解天意以御史之言為非耶亦乞勑執中起視事無使天意久不決也寒暑者賞罰也乍寒乍暑者不當賞而賞當罰而不罰也鄧保吉有過於法不當為

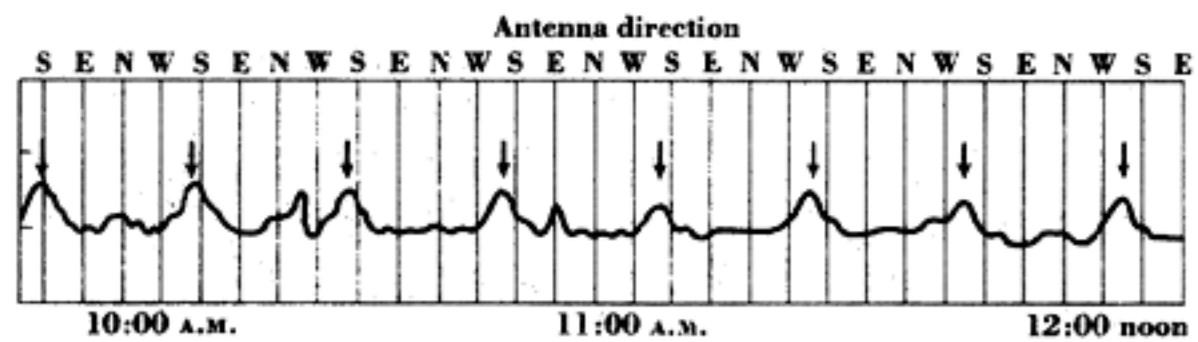
奏議卷之三百一　十

"Simulated image of supernova SN 1054 at the position of modern Crab Nebula, as presumably would have been observed from capital of Song Dynasty at Kaifeng, China during the morning of July 4th, 1054", wikipedia.

Why Data Science for Astronomers?

- Complicated Instrumentation and Algorithms.



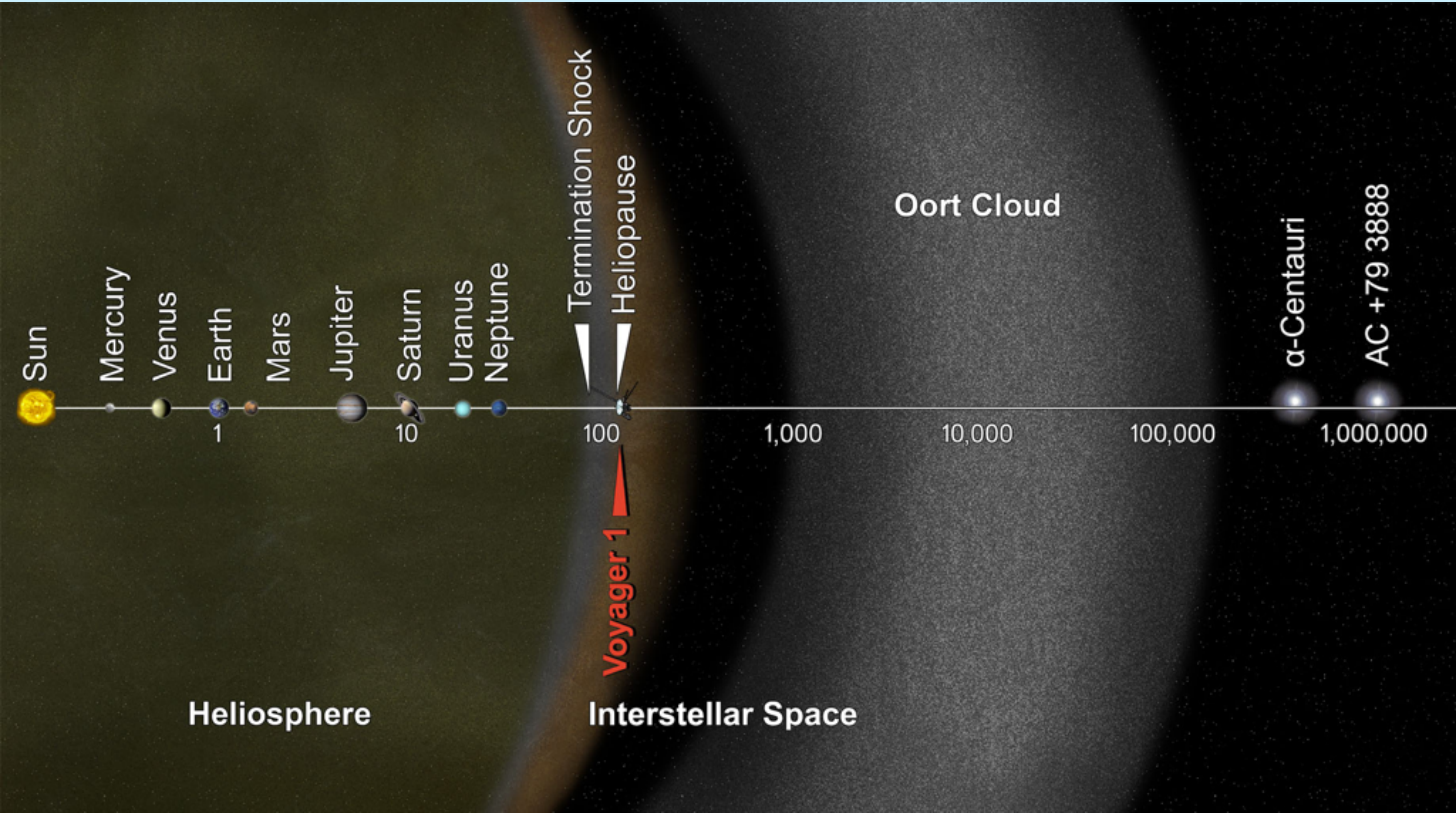


Astronomy

- Astronomy — study of celestial phenomena.
  - Planets, stars, the Milky Way, Galaxies.
  - Study a wide variety of physical processes, which we cannot reproduce on Earth!

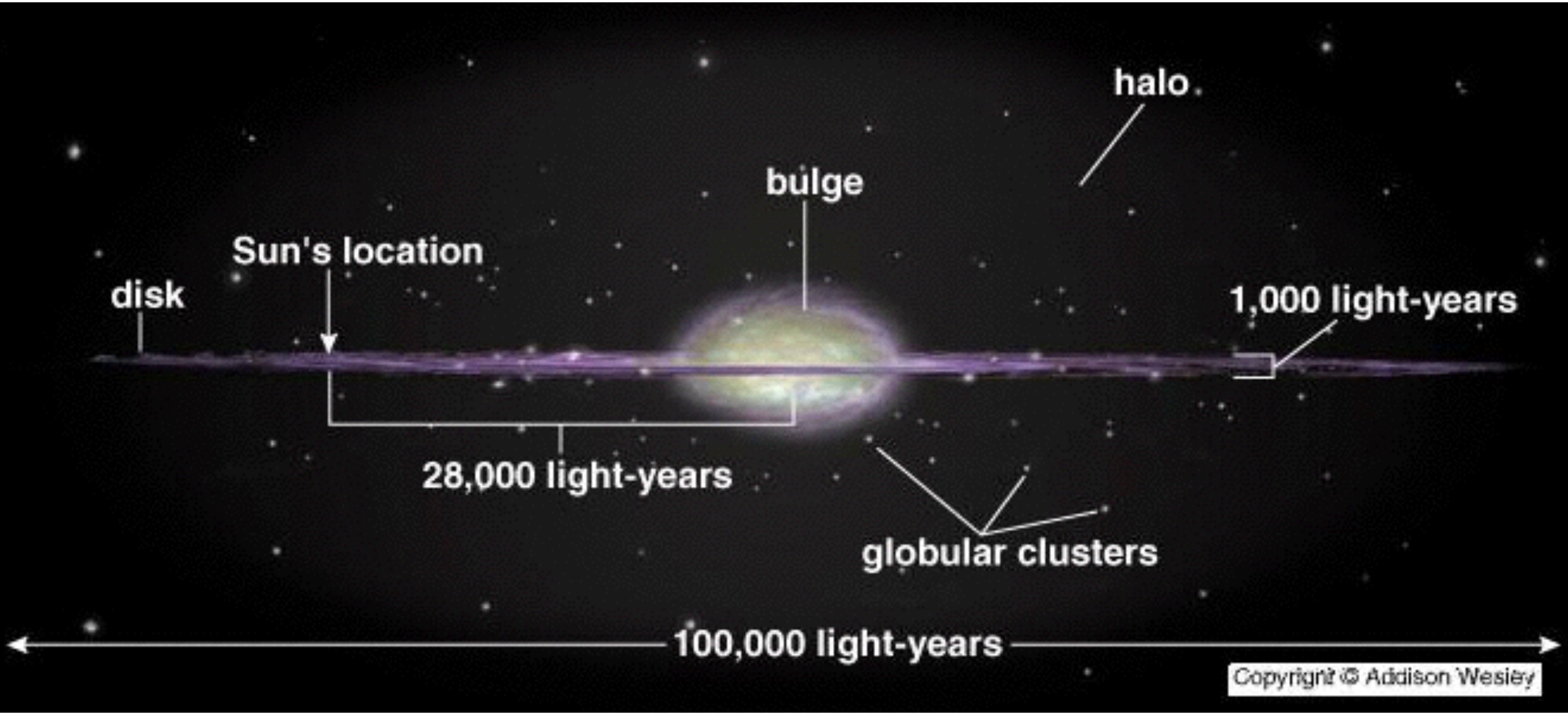
A short note on distances...

- 1 light-year = 9.5×10^{12} km.
- 1 Astronomical Unit (AU) = 1.5×10^{8} km, 8 light-minutes.
- 1 Parsec = 3.1×10^{13} km, 206265 AUs, 3.3 light-years.
- 1 kpc = 3261 light-years.
- 1 Mpc = 3.3×10^{6} light-years.
- Redshift: z=1 is equivalent to a luminosity distance of 6701Mpc.
    - Q: What is the difference between luminosity distance and angular distance?

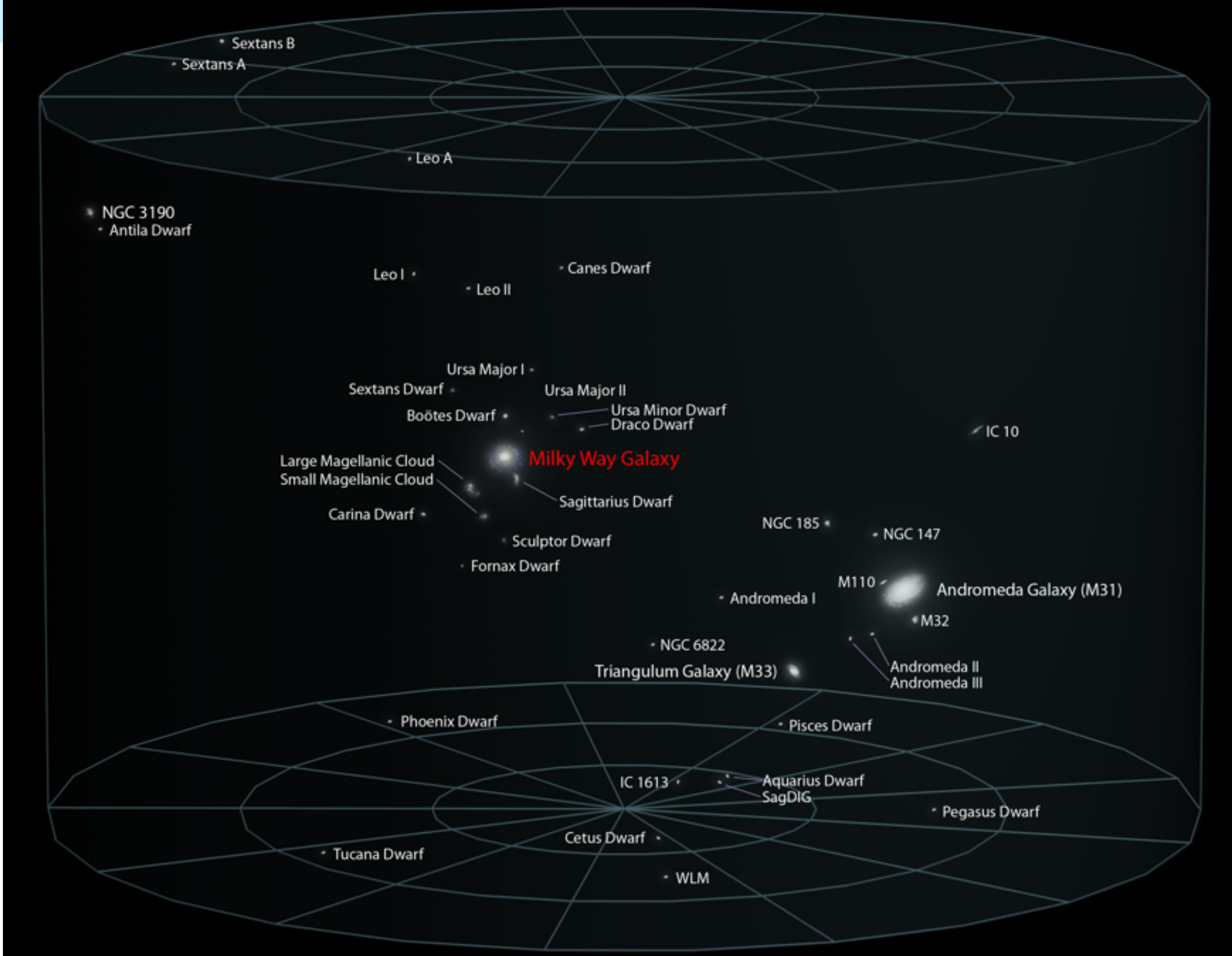


https://www.jpl.nasa.gov

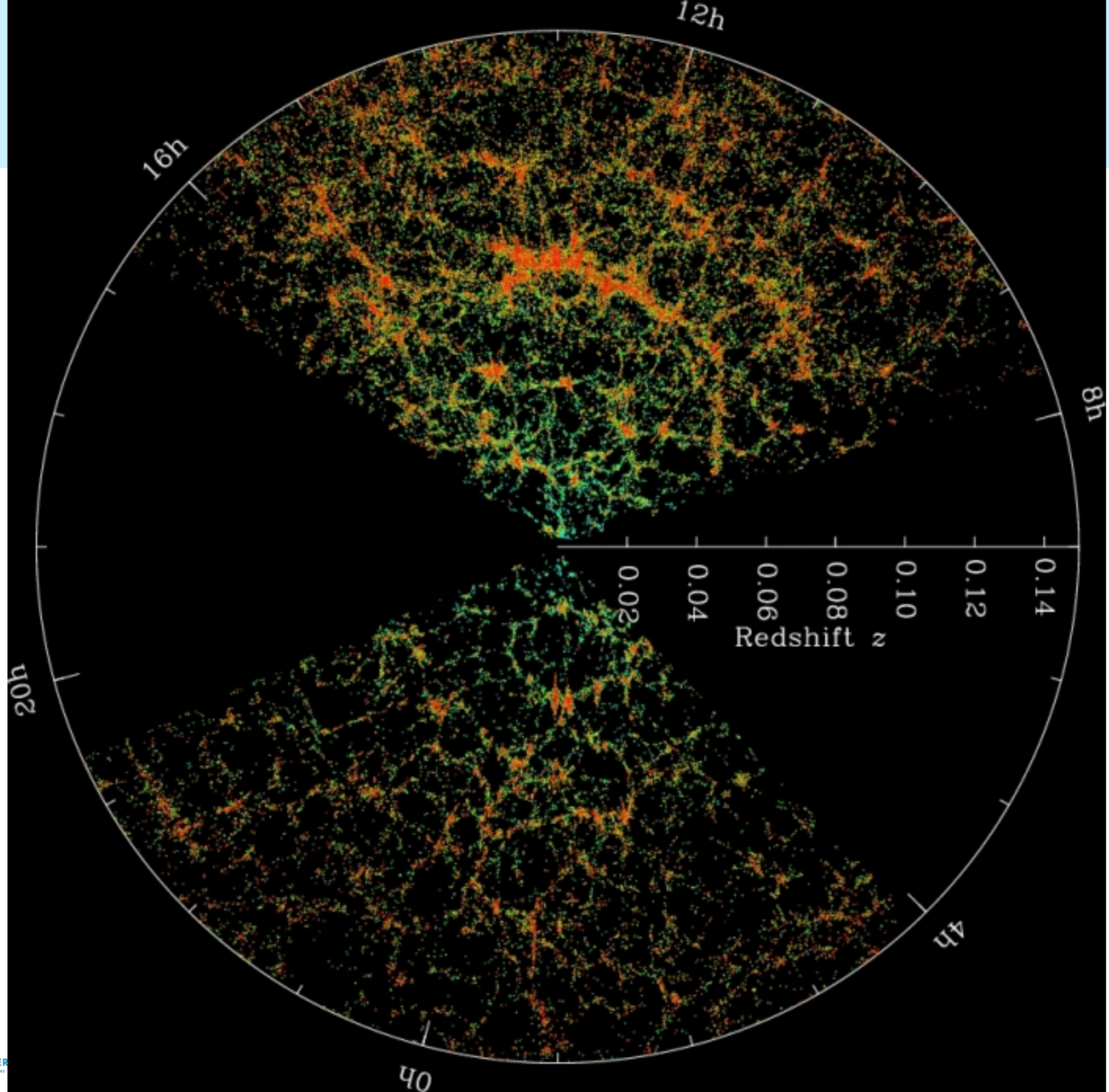
The Milky Way

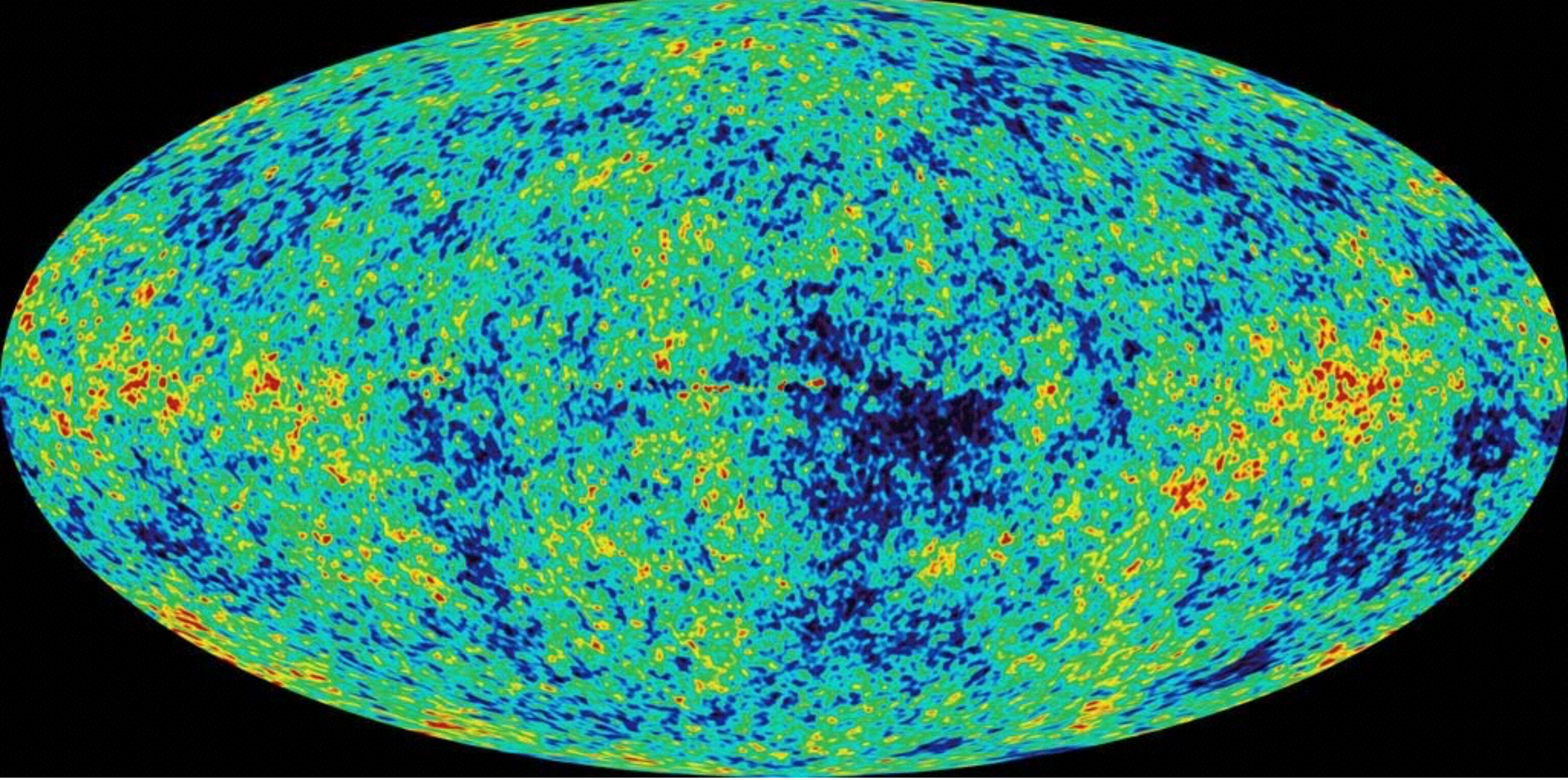


Local Galactic Group

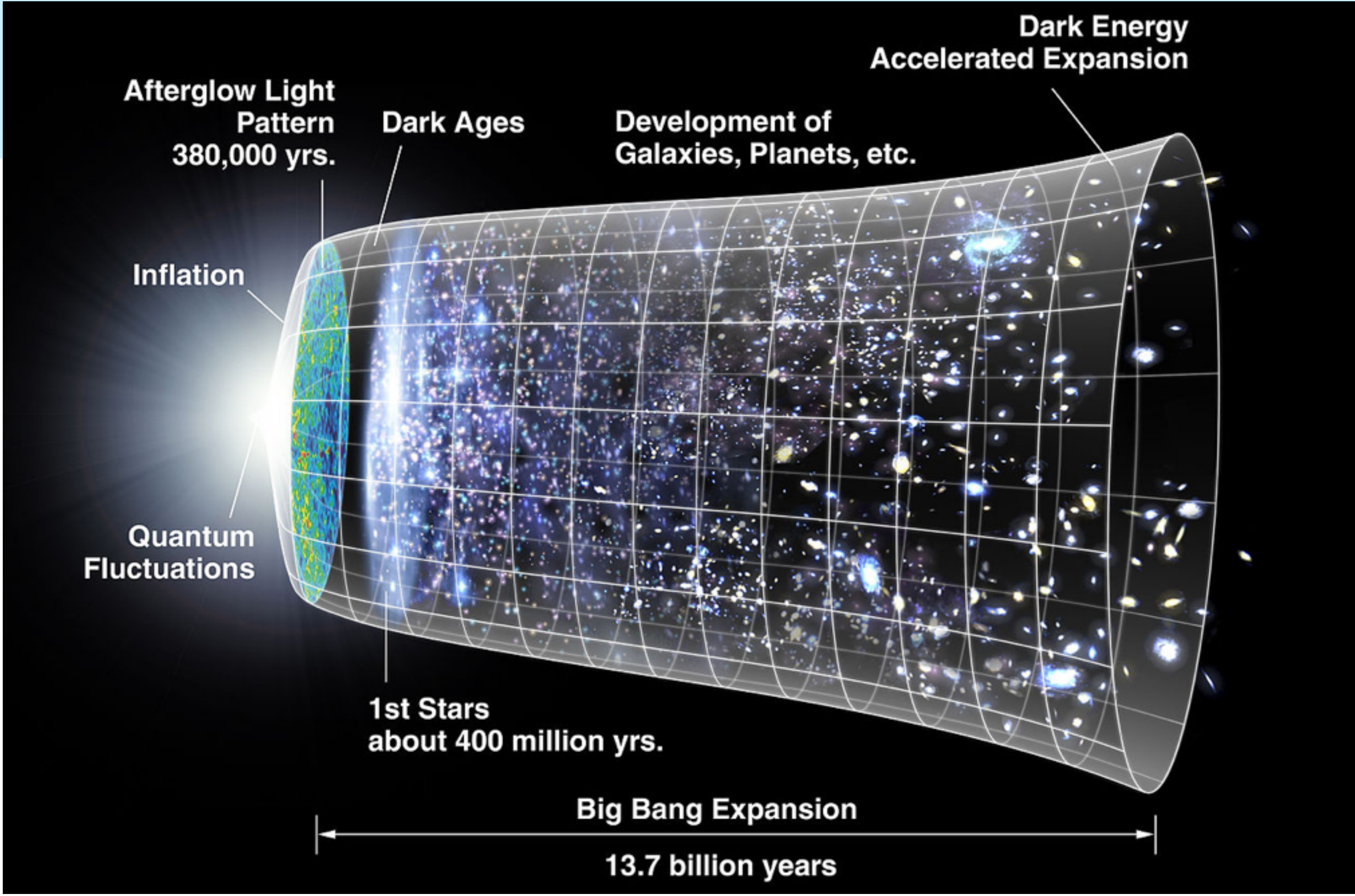



HD 112734
NGC 4881
HD 112886
NGC 4895
NGC 4907
NGC 4860
IC 4045
NGC 4865
HD 112887
IC 4012
IC 4040
IC 4026
IC 4021
NGC 4908
NGC 4883
IC 4011
IC 4051
IC 4041
IC 3955
NGC 4884/9
NGC 4873
IC 3998
NGC 4864
IC 4033
NGC 4874
NGC 4898
NGC 4906
NGC 4876
NGC 4869
NGC 4921
IC 3973
IC 3976
IC 3960
IC 3949
NGC 4919
NGC 4911
IC 3959


12h
16h
8h
20h
4h
0h
0.02
0.04
0.06
0.08
0.10
0.12
0.14
Redshift z



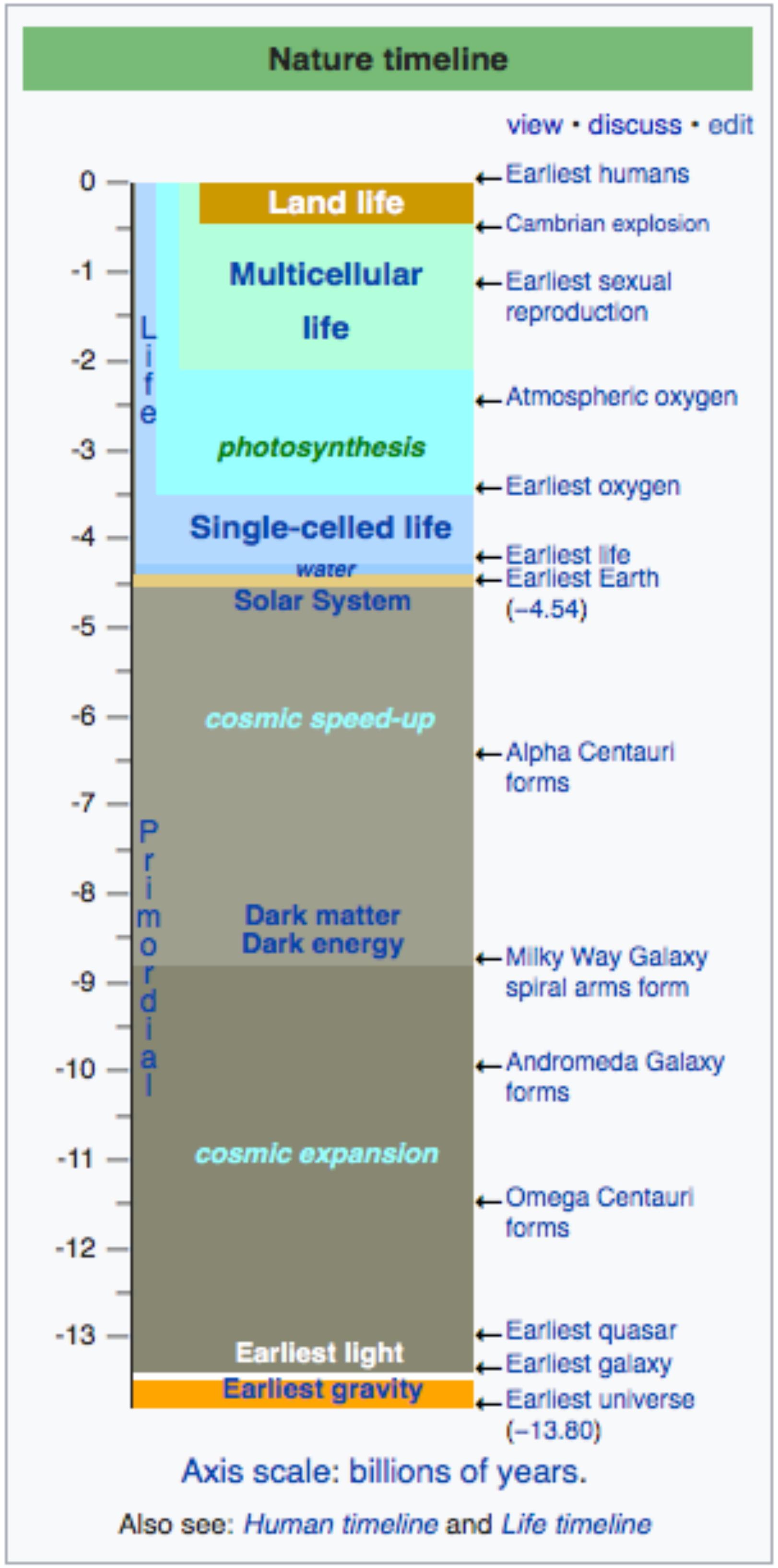
Cosmic Microwave Background Radiation https://www.lsst.org



NASA/WMAP Science Team

Nature's Timeline

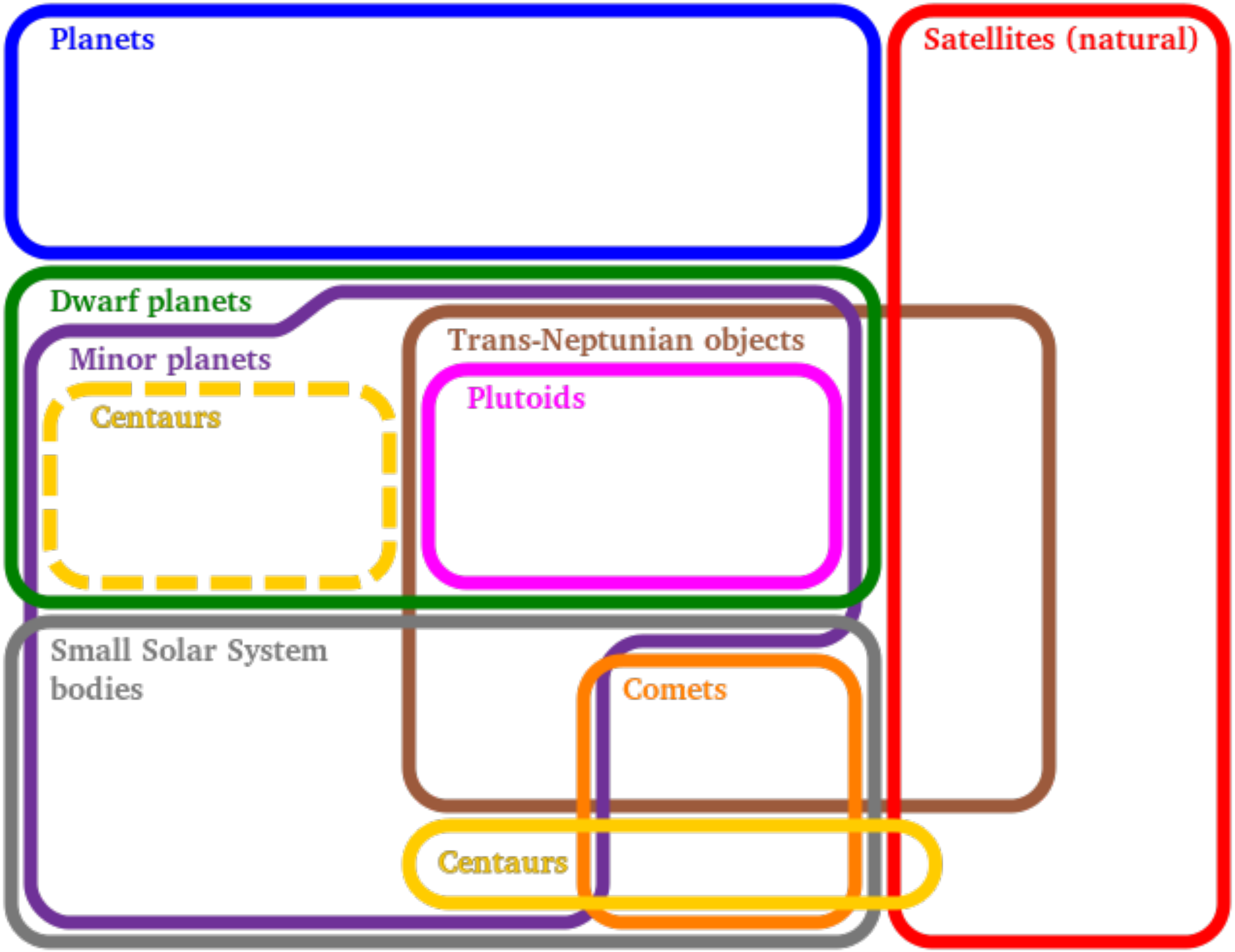
- Wikipedia.

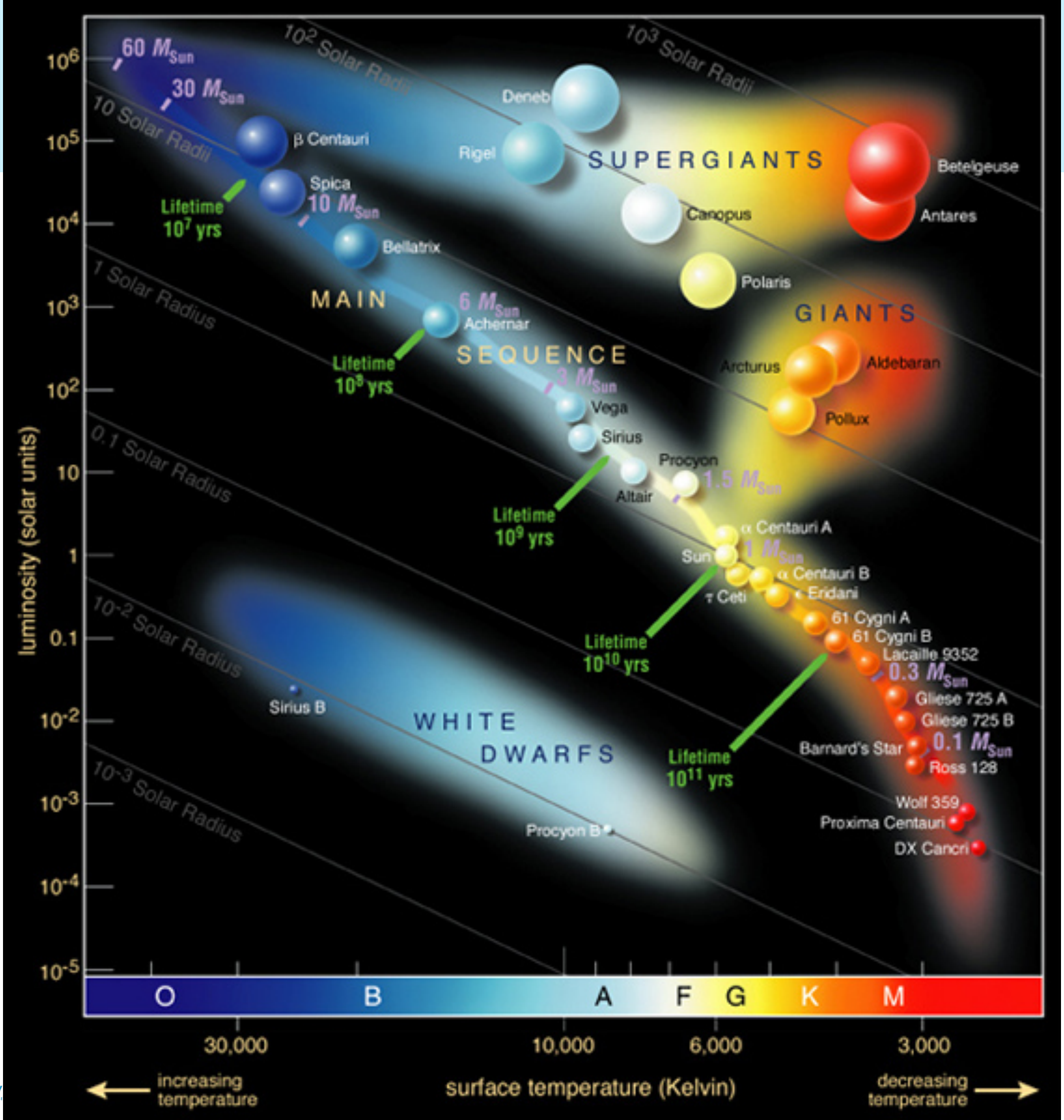


Planets

- The IAU...resolves that planets and other bodies in the Solar System be defined into three distinct categories in the following way:
    1. A planet [1] is a celestial body that (a) is in orbit around the Sun, (b) has sufficient mass for its self-gravity to overcome rigid body forces so that it assumes a hydrostatic equilibrium (nearly round) shape, and (c) has cleared the neighbourhood around its orbit.
    2. A "dwarf planet" is a celestial body that (a) is in orbit around the Sun, (b) has sufficient mass for its self-gravity to overcome rigid body forces so that it assumes a hydrostatic equilibrium (nearly round) shape [2], (c) has not cleared the neighbourhood around its orbit, and (d) is not a satellite.
    3. All other objects [3] orbiting the Sun shall be referred to collectively as "Small Solar System Bodies".
- [1] The eight planets are: Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, and Neptune.
- [2] An IAU process will be established to assign borderline objects into either dwarf planet and other categories.
- [3] These currently include most of the Solar System asteroids, most Trans-Neptunian Objects (TNOs), comets, and other small bodies.

Planets


Planets
Satellites (natural)
Dwarf planets
Minor planets
Trans-Neptunian objects
Centaurs
Plutoids
Small Solar System bodies
Comets
Centaurs



https://cdn.thinglink.me

Main Sequence Stars

Spectral Type:	O	B	A	F	G	K	M
Temperature:	40 000K	20 000K	8500K	6500K	5700K	4500K	3200K
Radius (Sun=1):	10	5	1.7	1.3	1.0	0.8	0.3
Mass (Sun=1):	50	10	2.0	1.5	1.0	0.7	0.2
Luminosity (Sun=1):	100 000	1000	20	4	1.0	0.2	0.01
Lifetime (million yrs):	10	100	1000	3000	10 000	50 000	200 000
Abundance:	0.00001%	0.1%	0.7%	2%	3.5%	8%	80%

Giant Stars

Low mass stars near the end of their lives.

Spectral Type:	Mainly G, K or M
Temperature:	3000 to 10 000K
Radius (Sun=1):	10 to 50
Mass (Sun=1):	1 to 5
Luminosity (Sun=1):	50 to 1000
Lifetime (million yrs):	1000
Abundance:	0.4%

White Dwarfs

Dying remnant of an imploded star.

Spectral Type:	D
Temperature:	Under 80 000K
Radius (Sun=1):	Under 0.01
Mass (Sun=1):	Under 1.4
Luminosity (Sun=1):	Under 0.01
Lifetime (million yrs):	–
Abundance:	5%

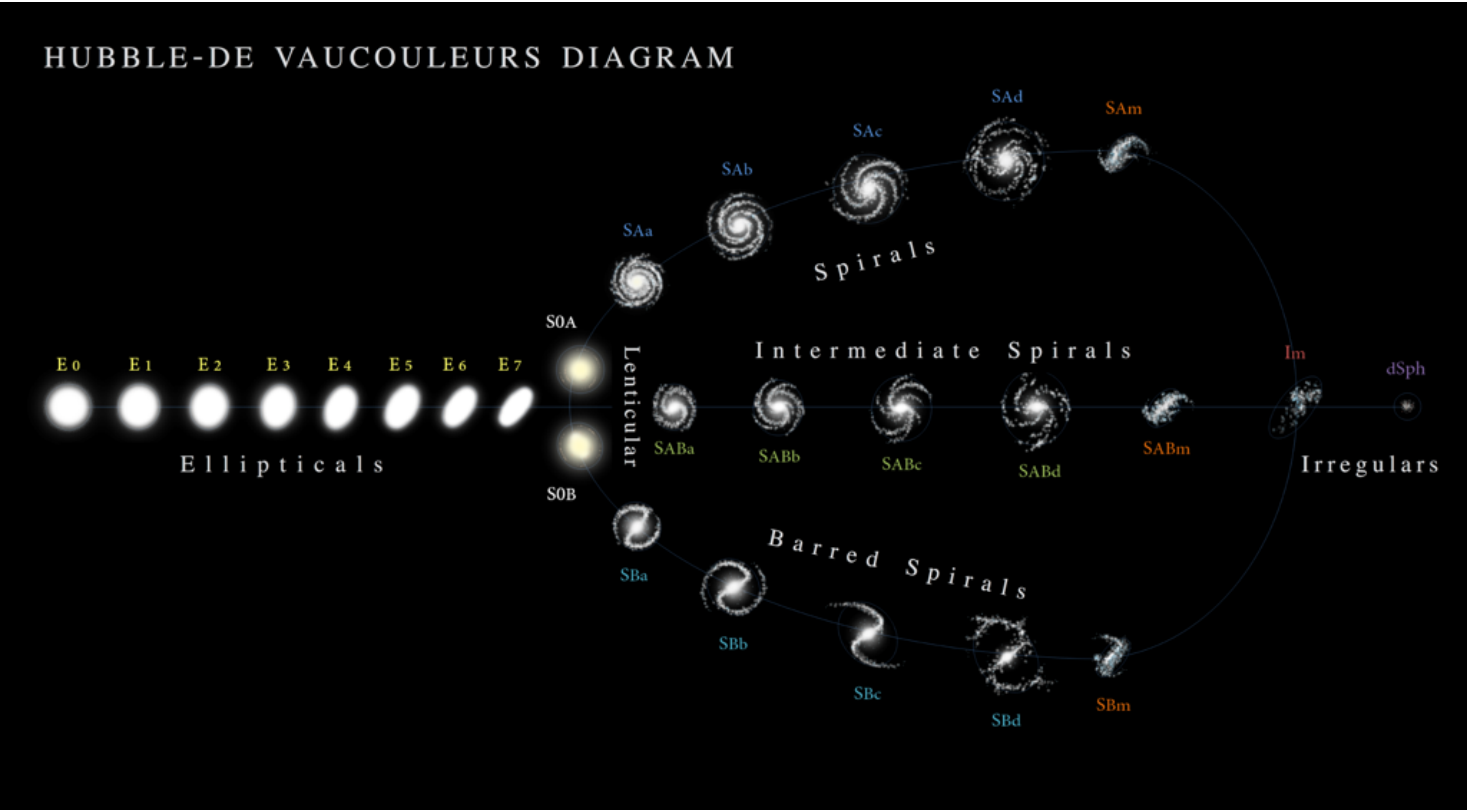
Supergiant Stars

High mass stars near the end of their lives.

Spectral Type:	O, B, A, F, G, K or M
Temperature:	4000 to 40 000K
Radius (Sun=1):	30 to 500
Mass (Sun=1):	10 to 70
Luminosity (Sun=1):	30 000 to 1000 000
Lifetime (million yrs):	10
Abundance:	0.0001%

r powell

Galaxies

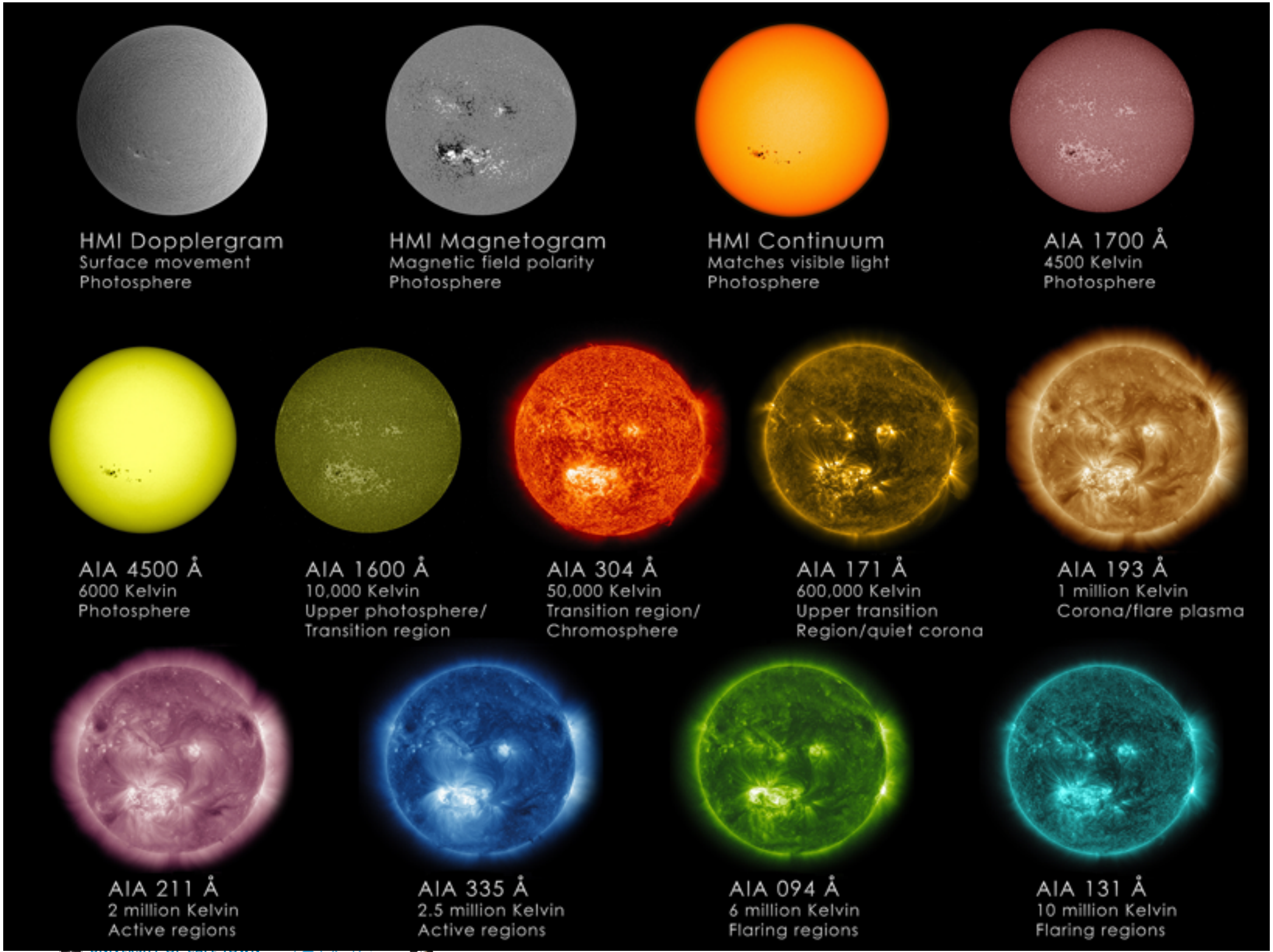


HUBBLE-DE VAUCOULEURS DIAGRAM
E0
E1
E2
E3
E4
E5
E6
E7
Ellipticals
S0A
S0B
Lenticular
SAa
SAb
SAc
SAd
SAm
Spirals
Intermediate Spirals
SABa
SABb
SABc
SABd
SABm
Barred Spirals
SBa
SBb
SBc
SBd
SBm
Im
dSph
Irregulars

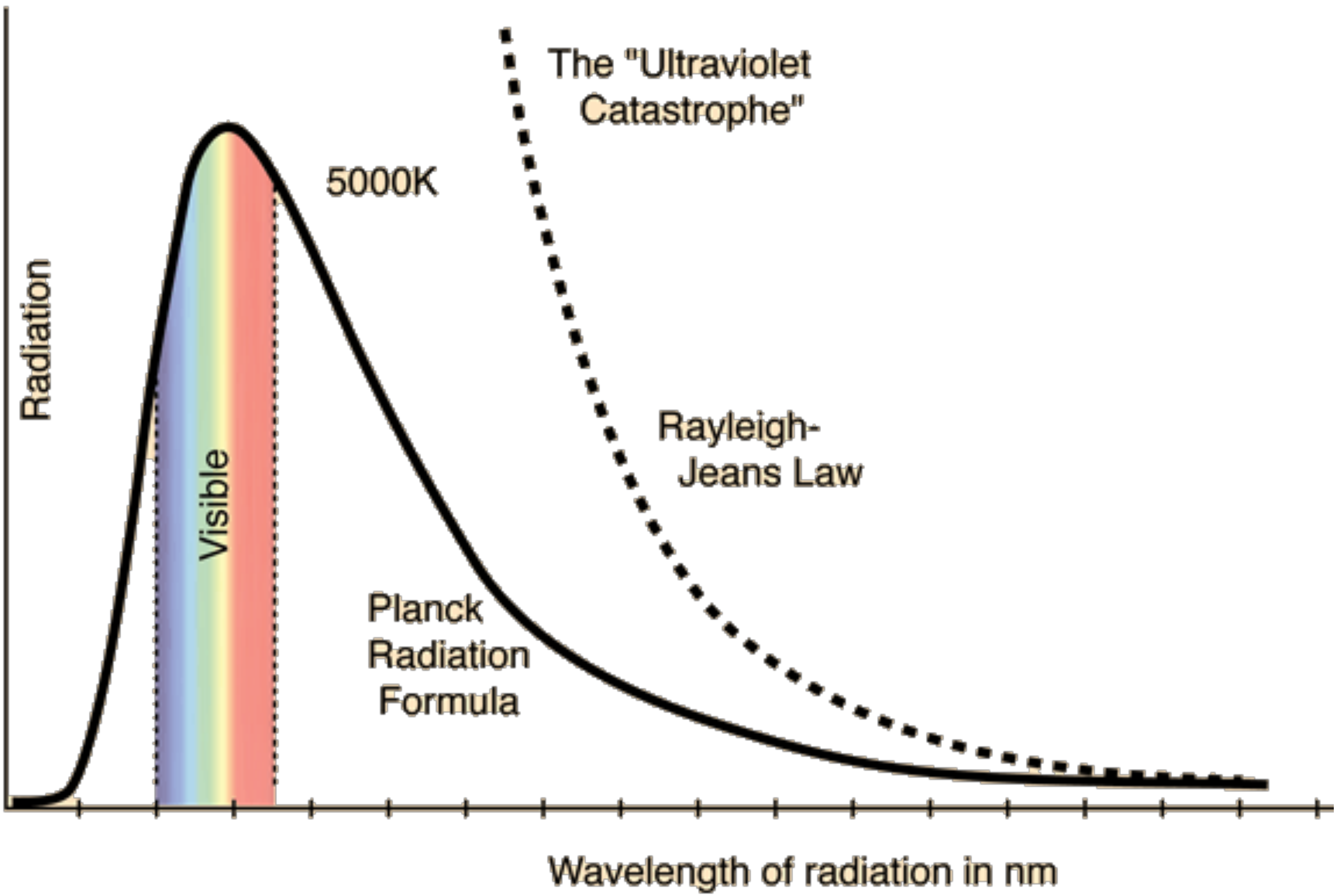
Astronomical Objects and Phenomena

- Blackbody Radiation, e.g., stars/gas/plasma.
- Plasma, e.g., free-free emission.
- Synchrotron emission, e.g., bremstrahlung.
- Spectral lines (quantum effects).
  - Neutral Hydrogen — spin-flip transition.
  - Molecules — vibrational and rotational modes.
  - Masers — Microwave Amplification by Stimulated Emission.

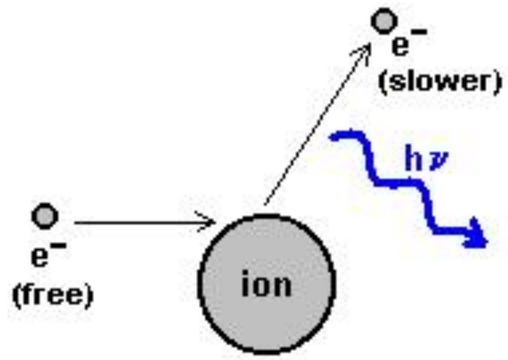
Blackbody Emission

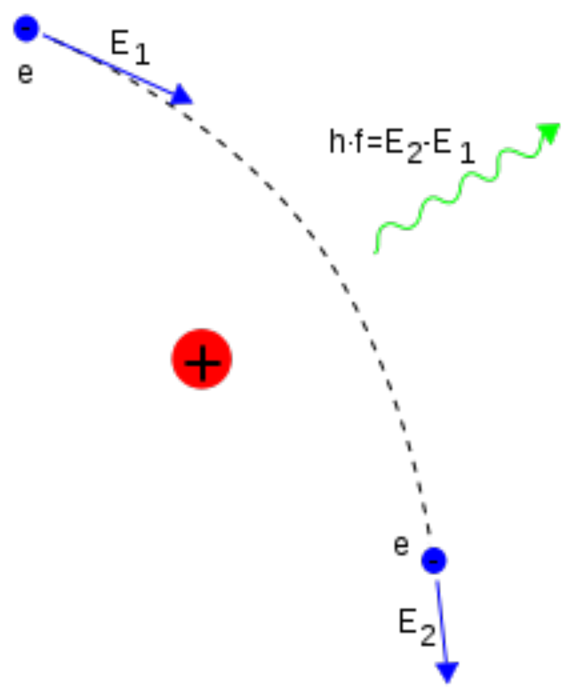


Blackbody Emission



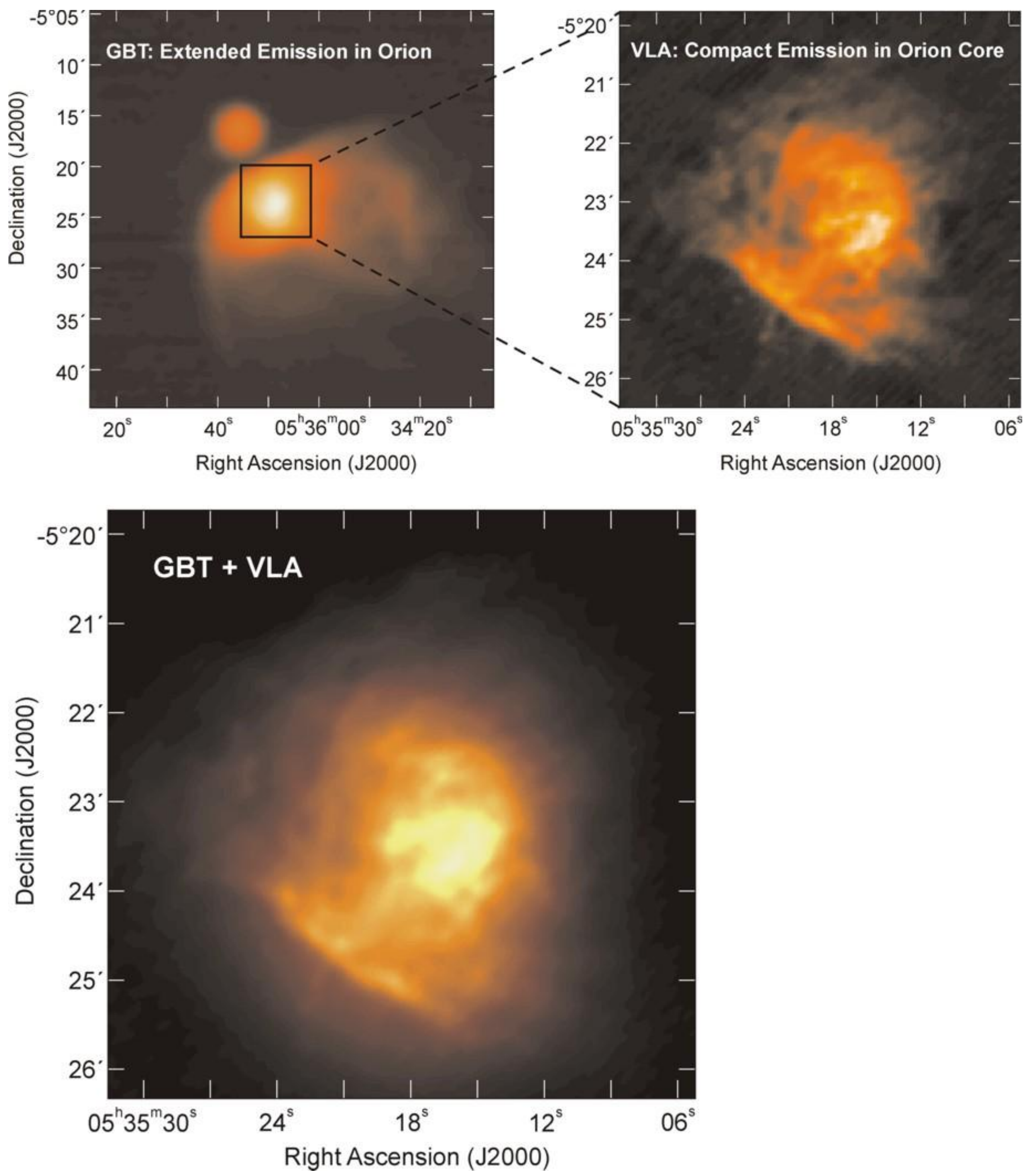
Free-free emission (radio)





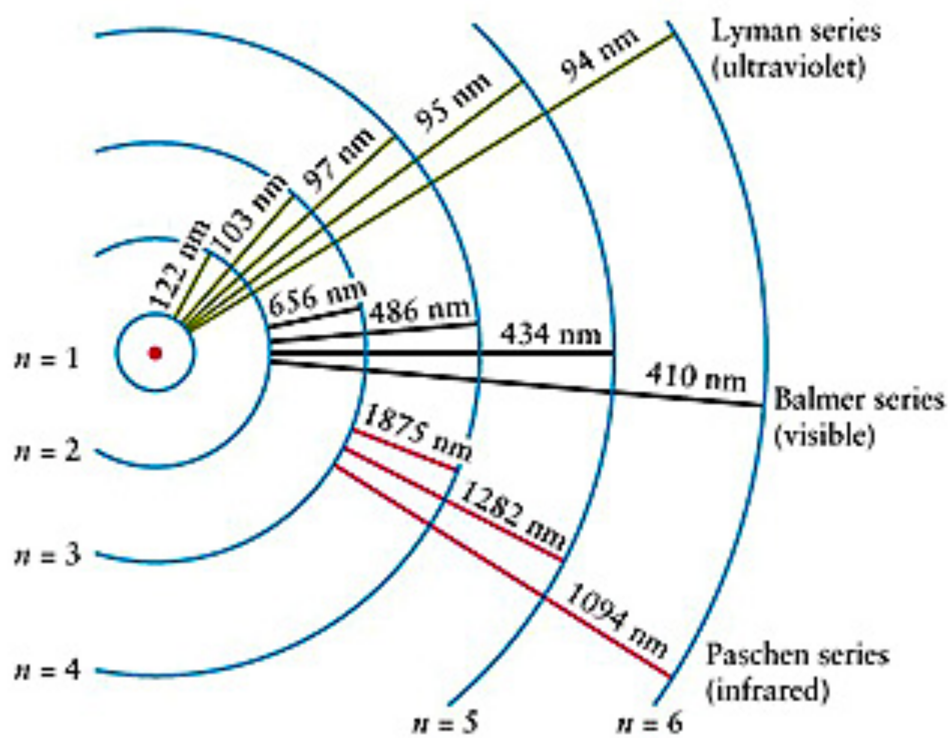
- Interstellar Plasmas
- Electrons and ions dissociated
- High kinetic temperatures
- Encounter results in change in momentum
- (Similar principle to synchrotron)
- Emission

Free-free emission (radio)

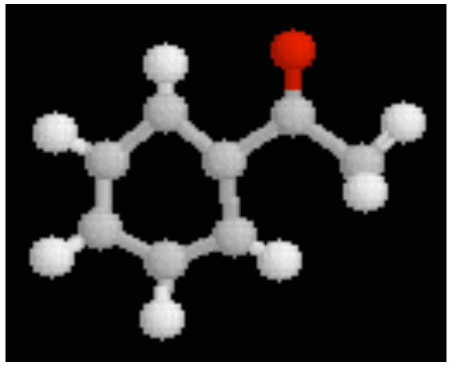


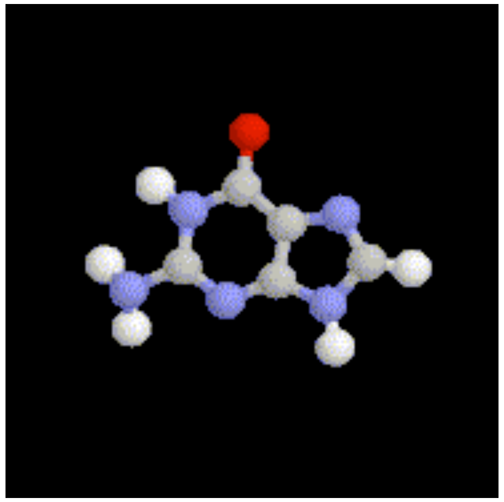
Spectral Lines

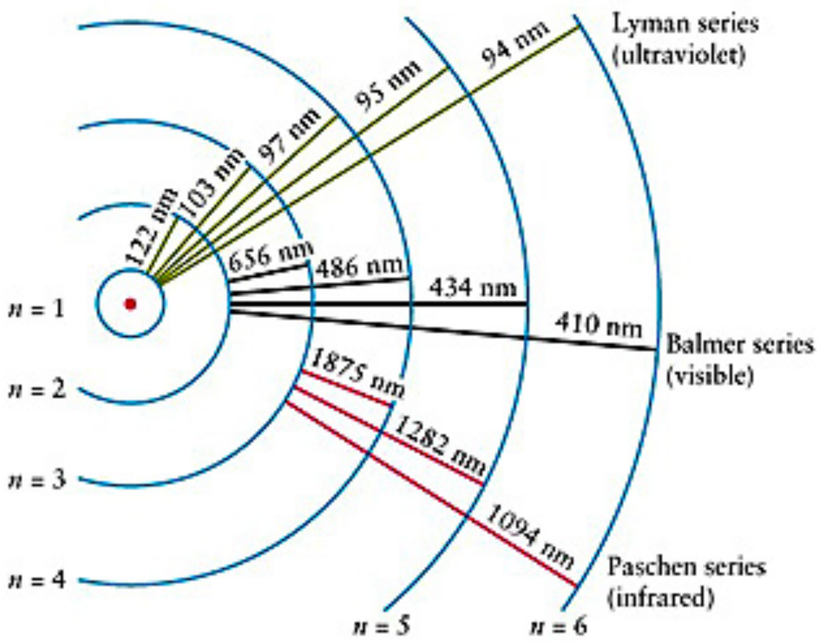
- Quantum processes, excitation/de-excitation, leads to well-defined characteristic absorption/emission of photons.
- Can be calculated; observations of spectral lines provide a very precise analysis of the population/conditions of astrophysical gases.

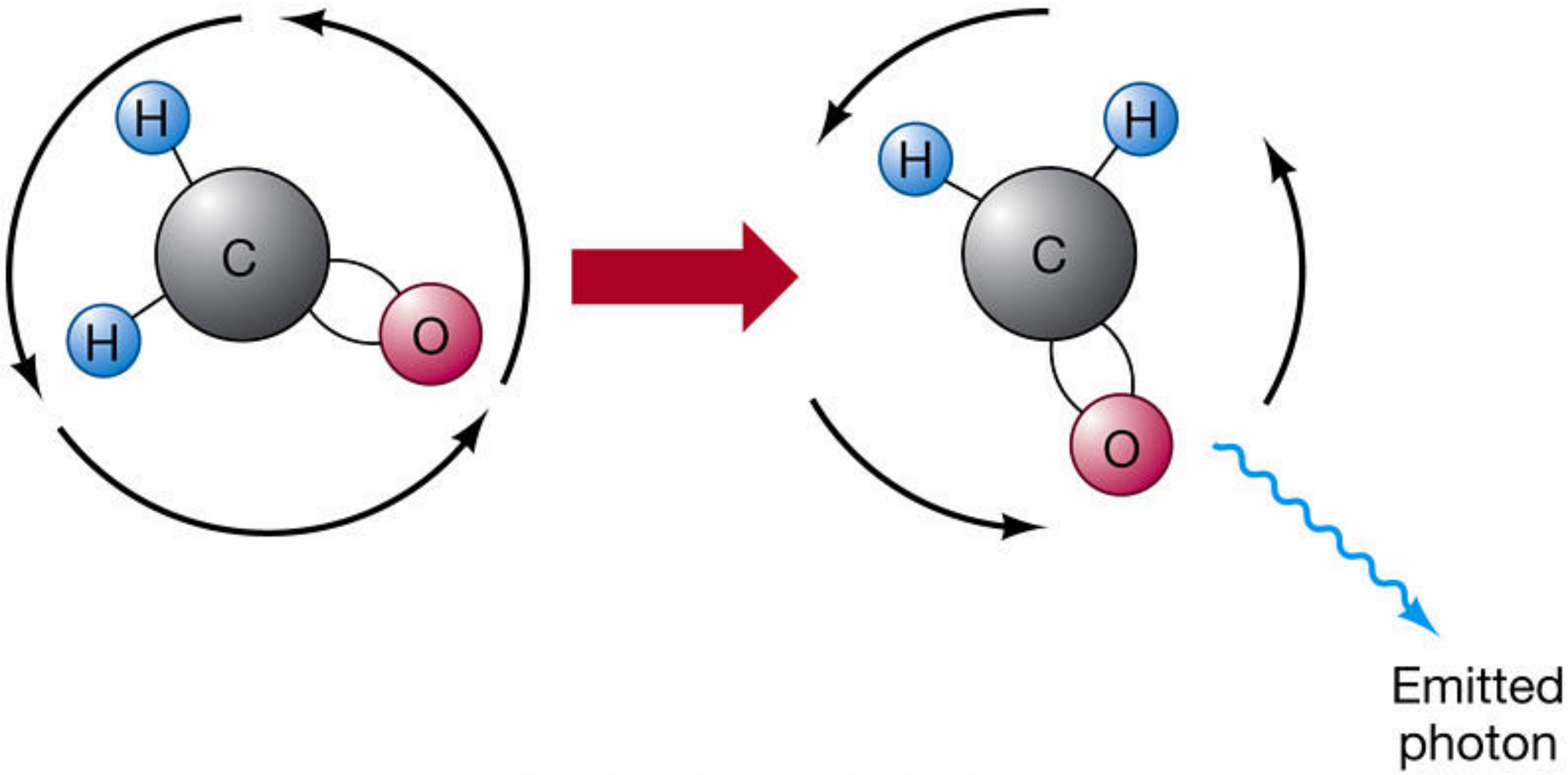


Spectral Lines



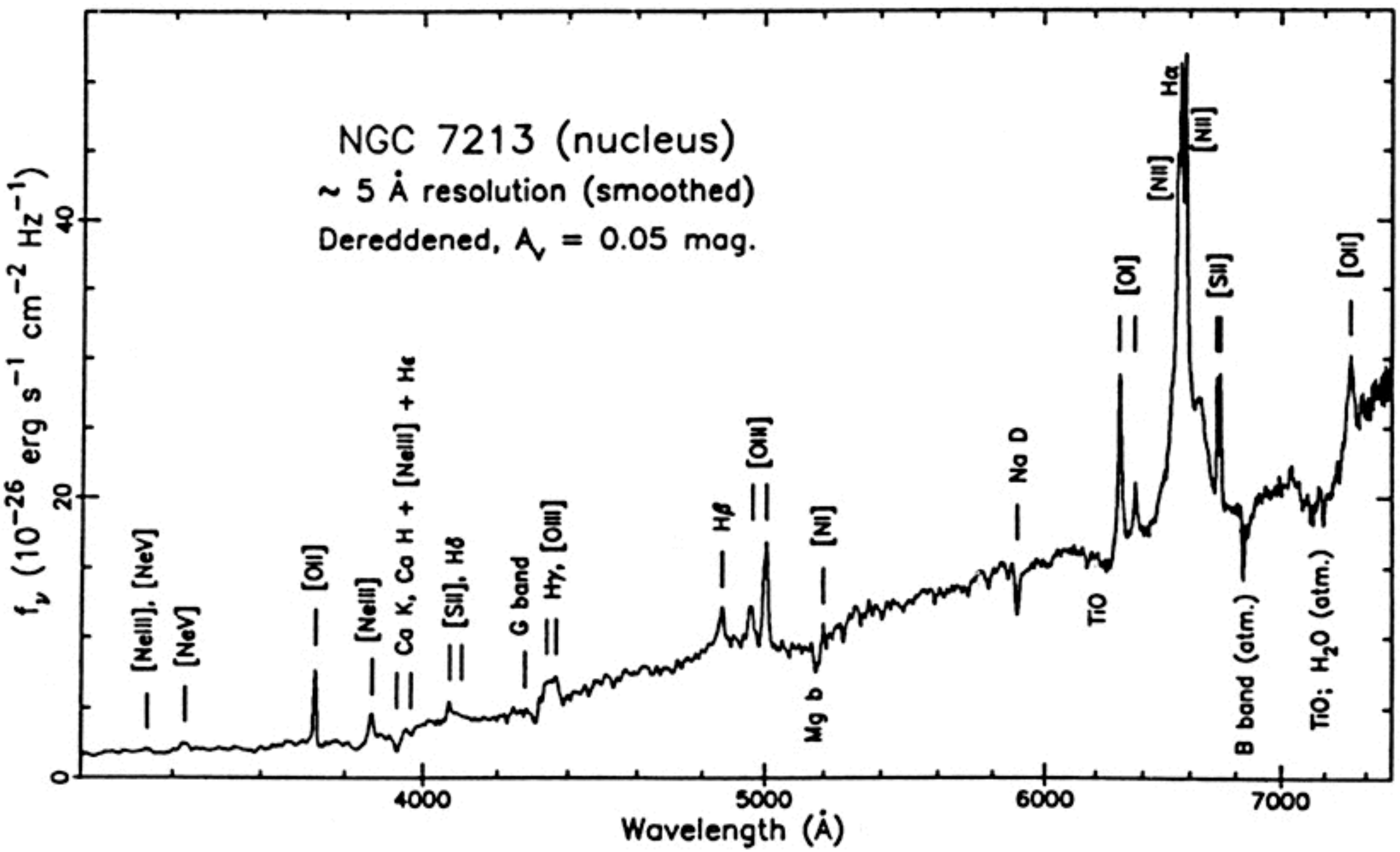




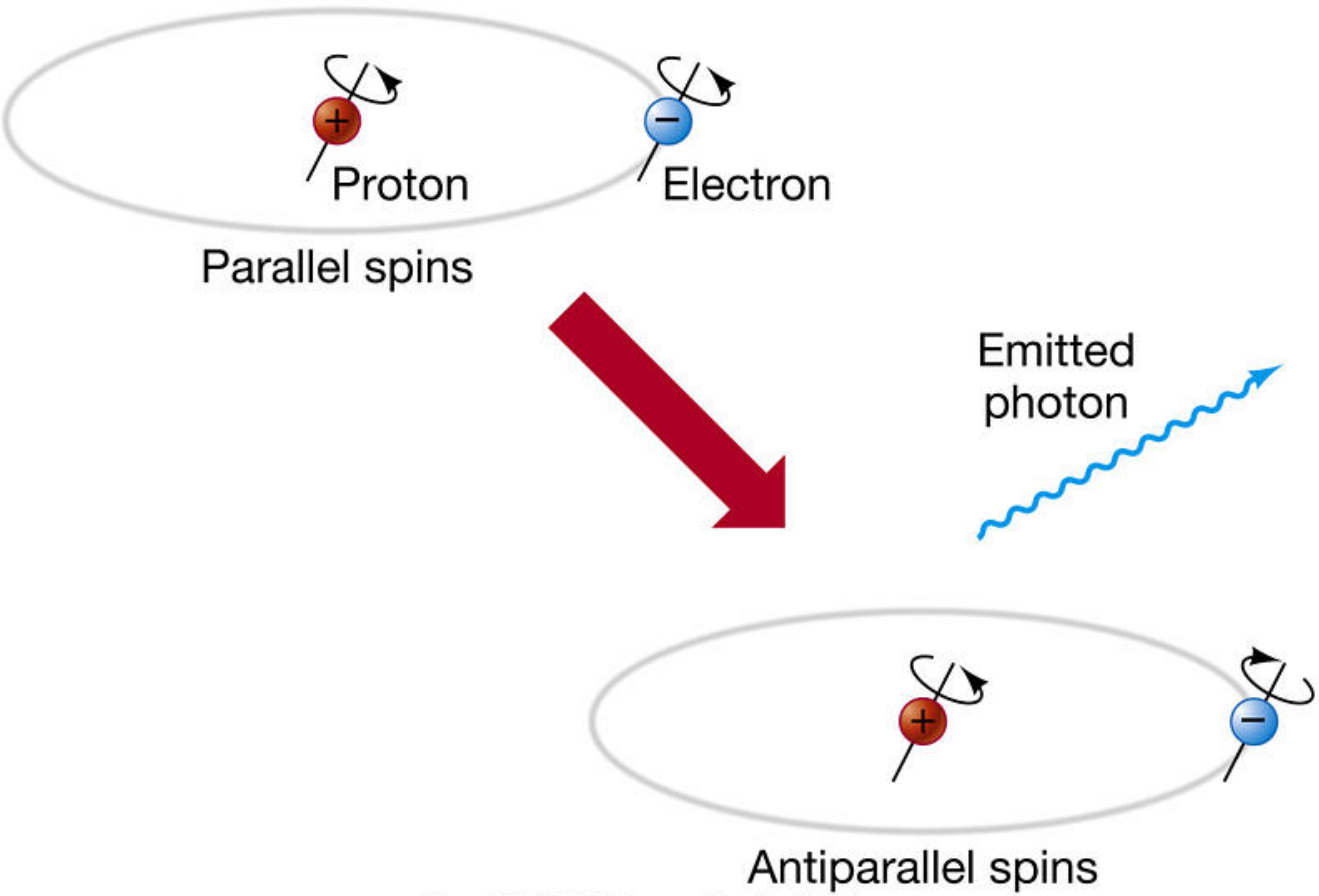


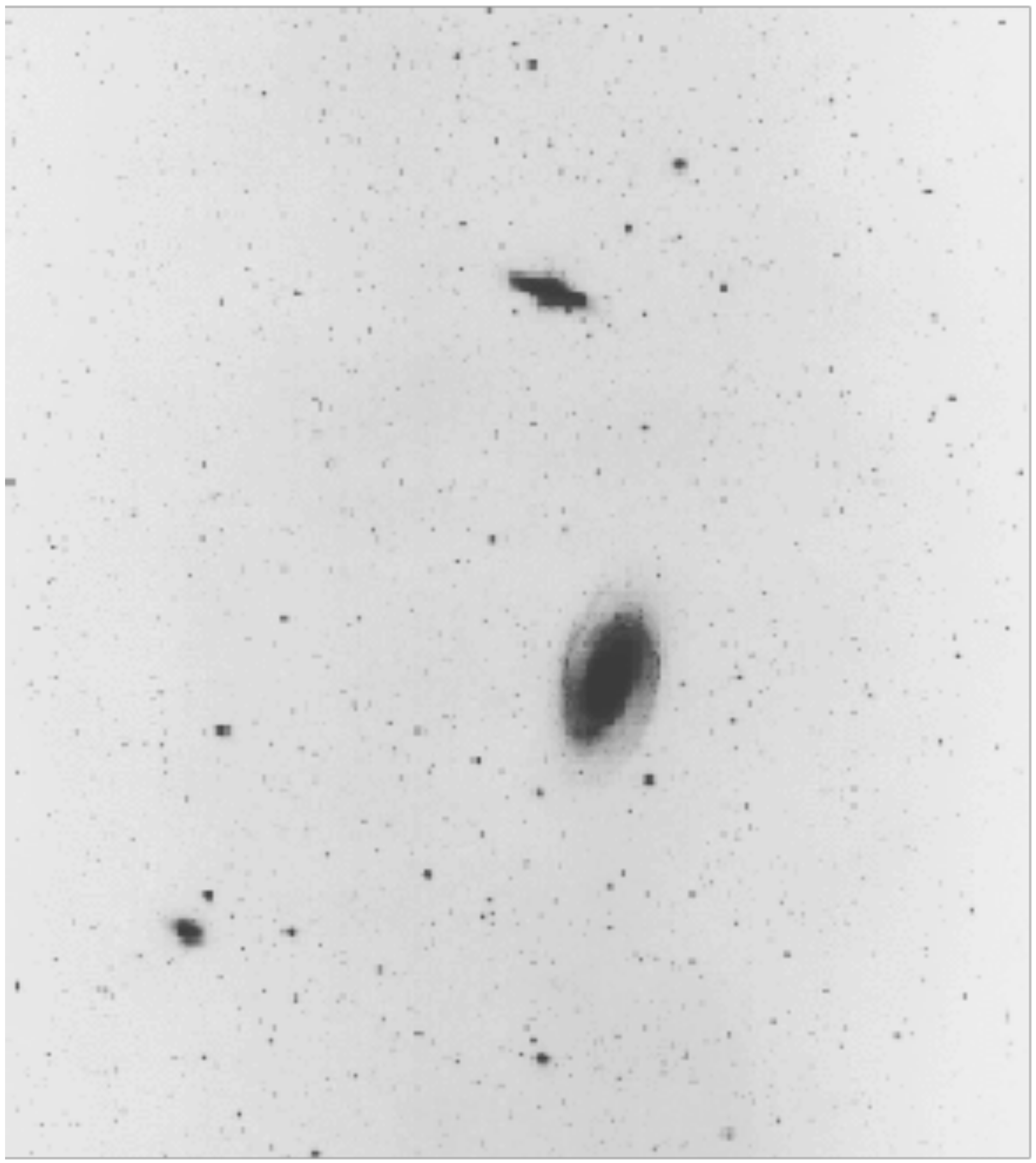
http://physics.uoregon.edu

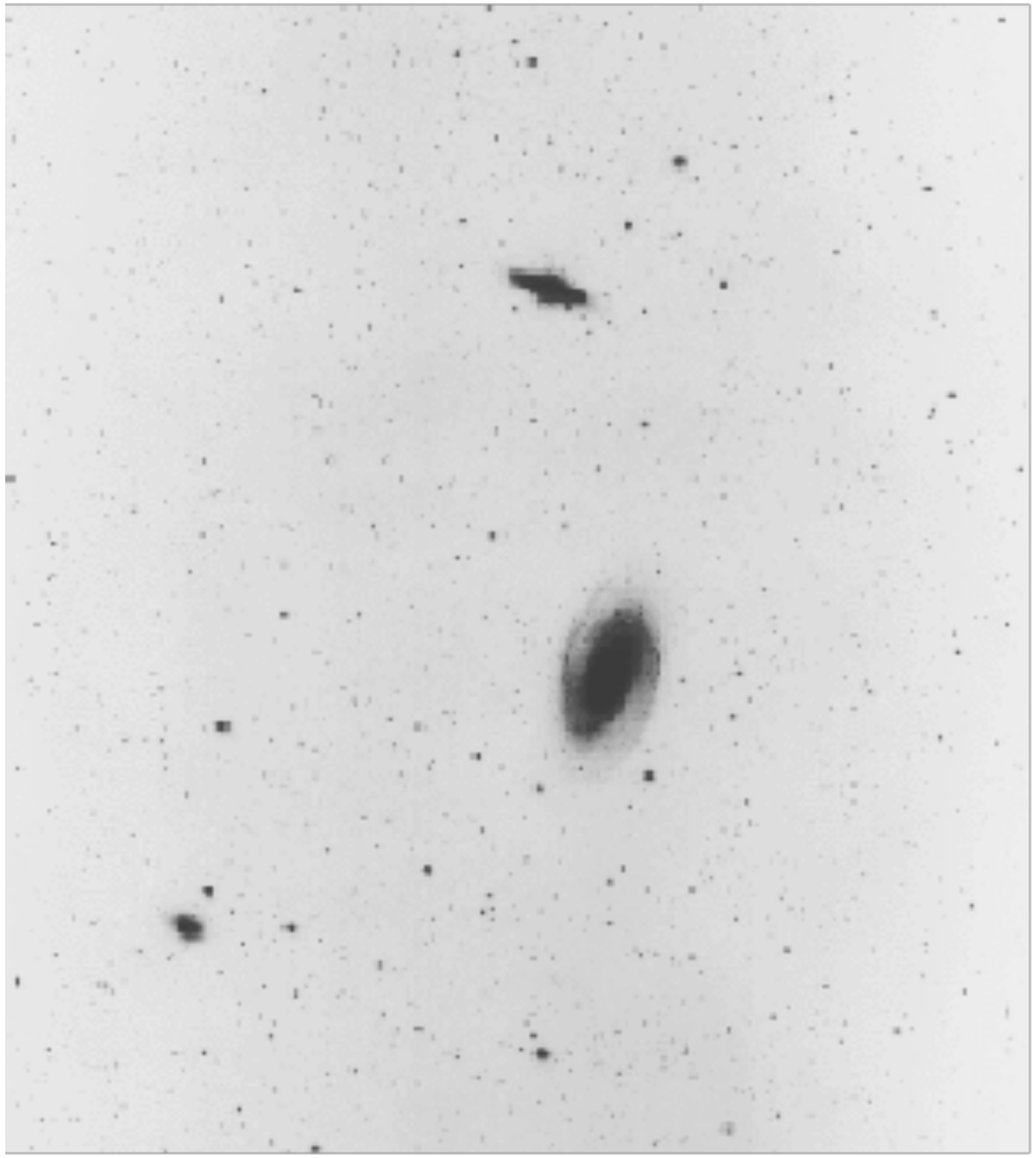
Spectral Lines

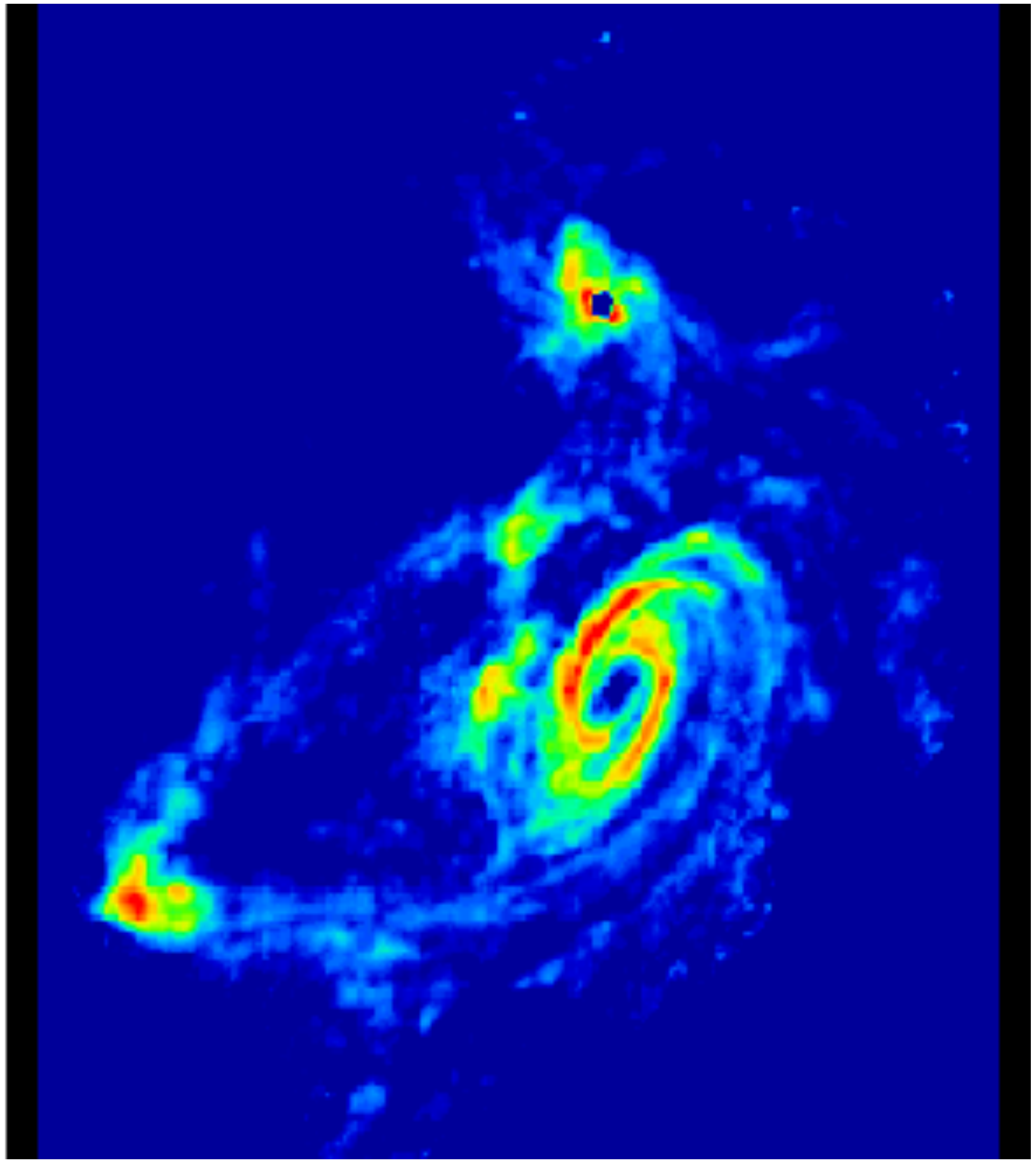


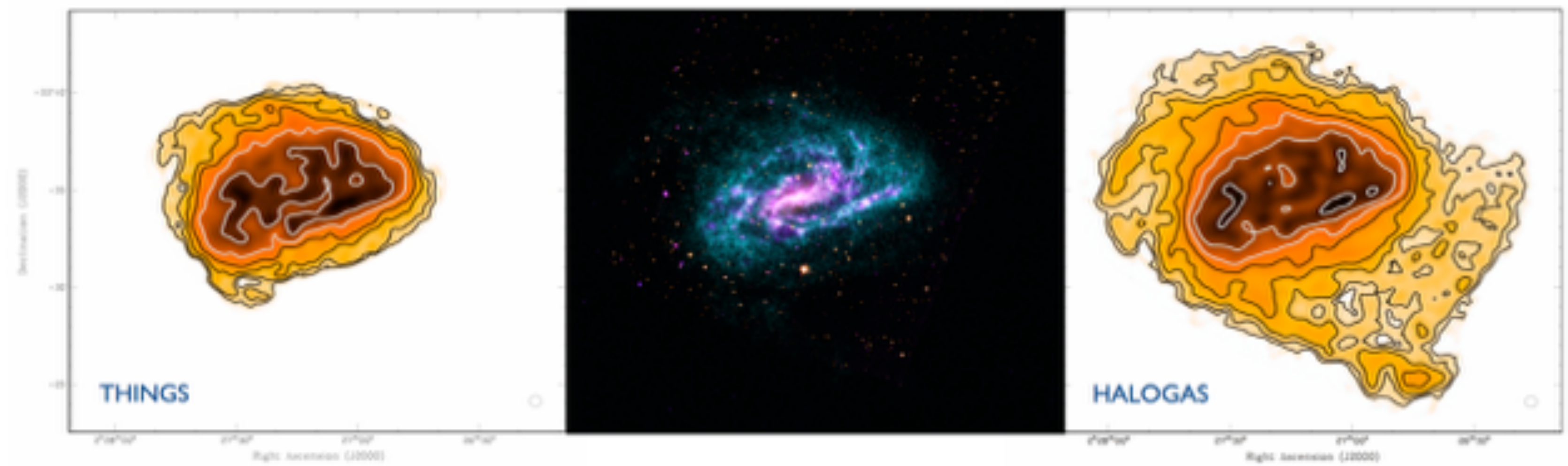
Neutral Hydrogen

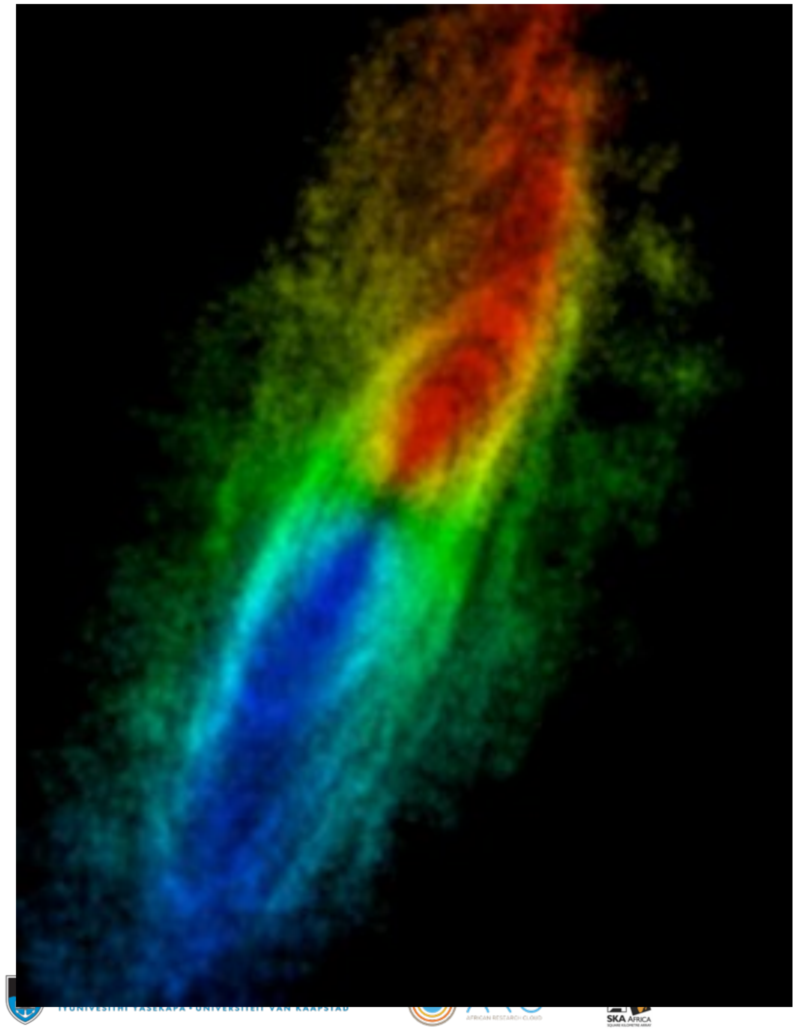


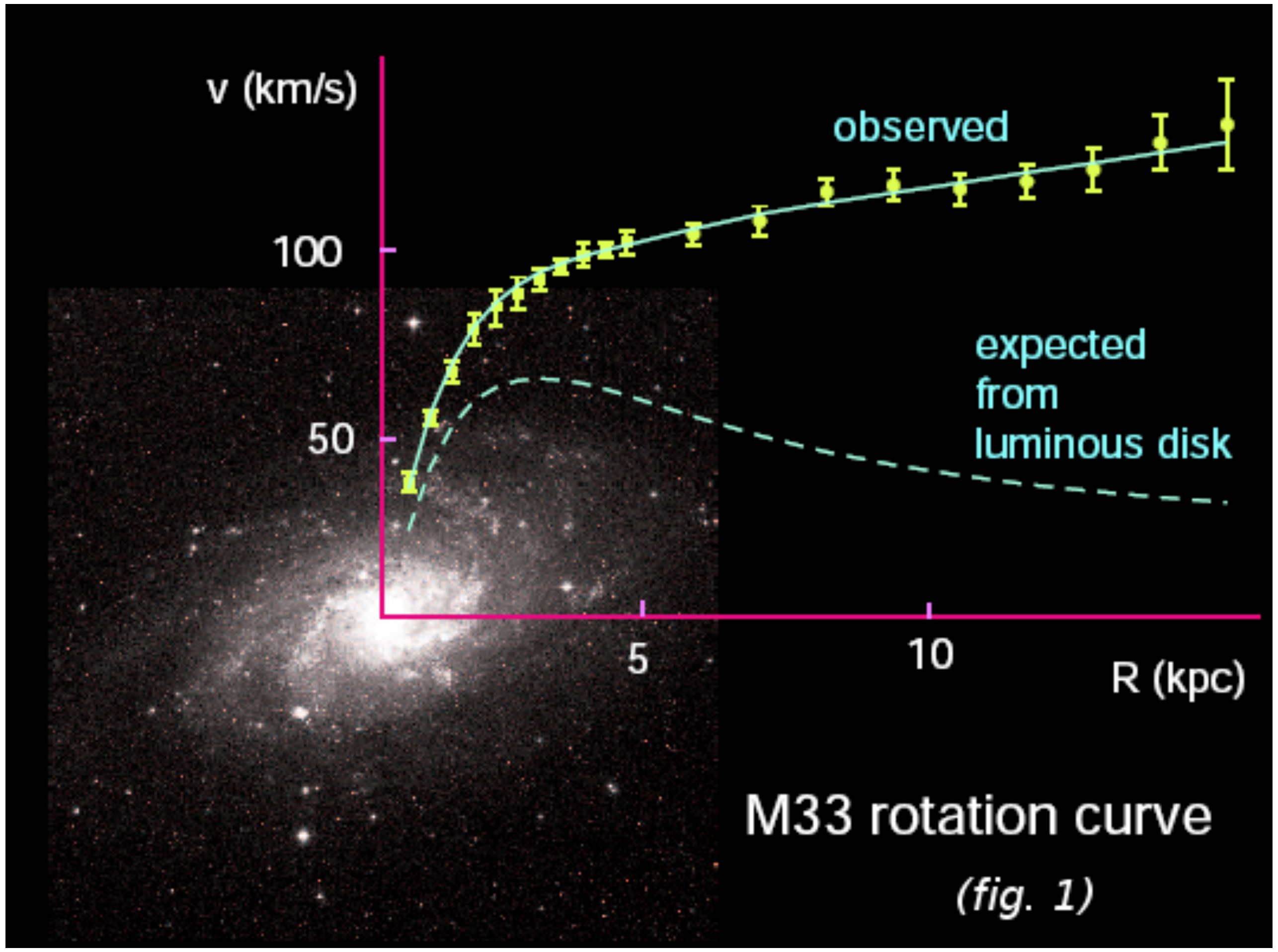


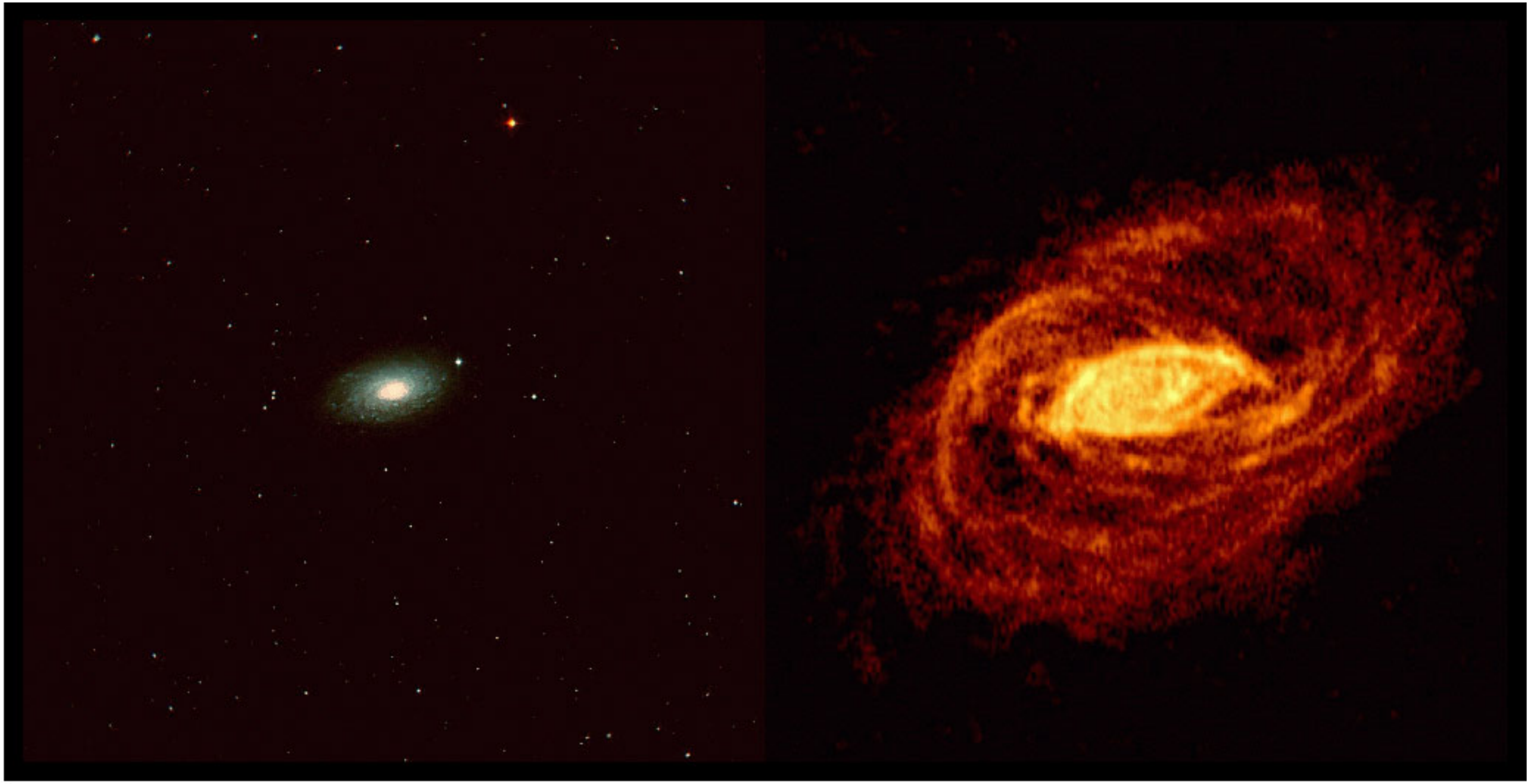




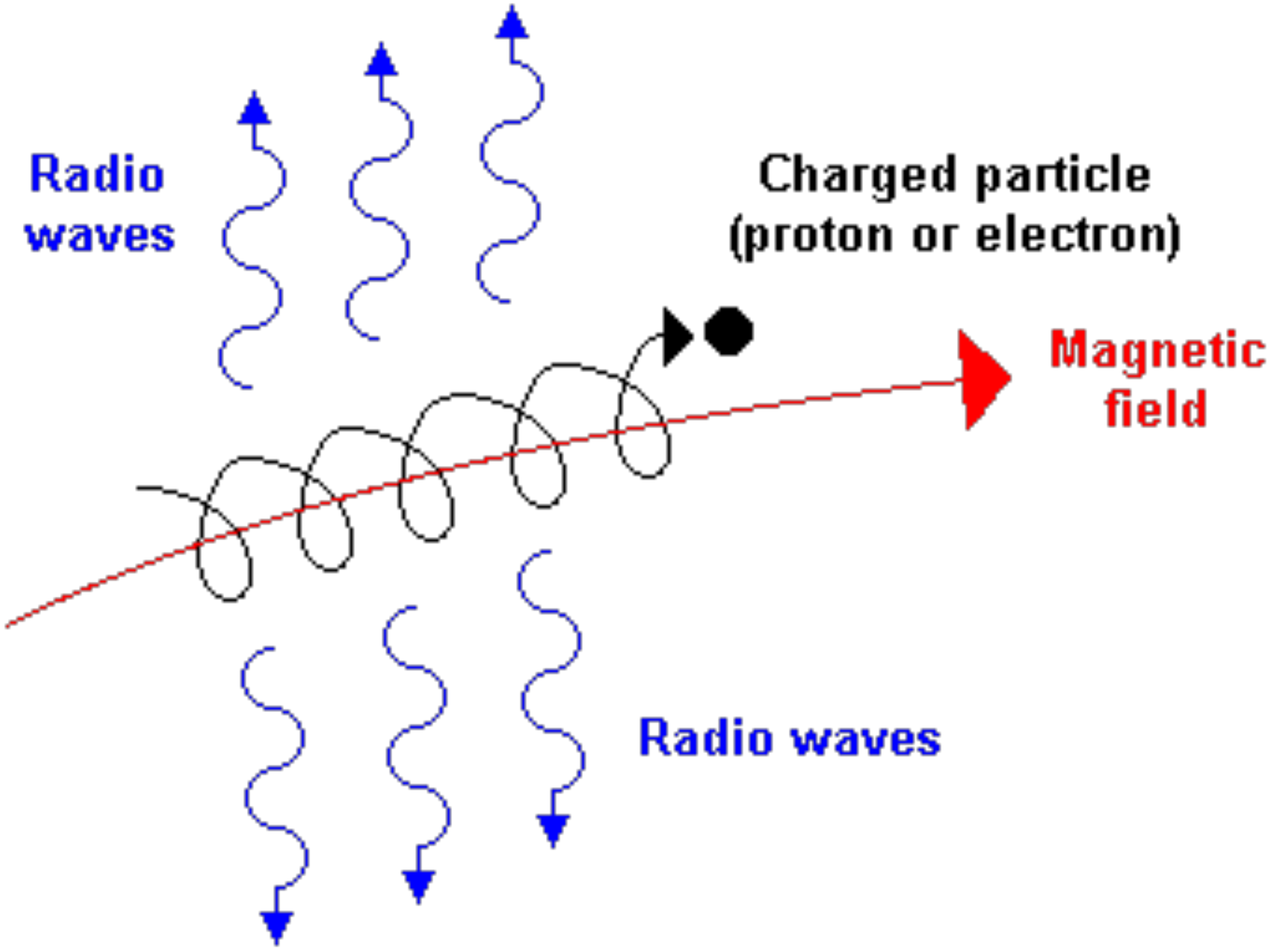

THINGS
HALOGAS

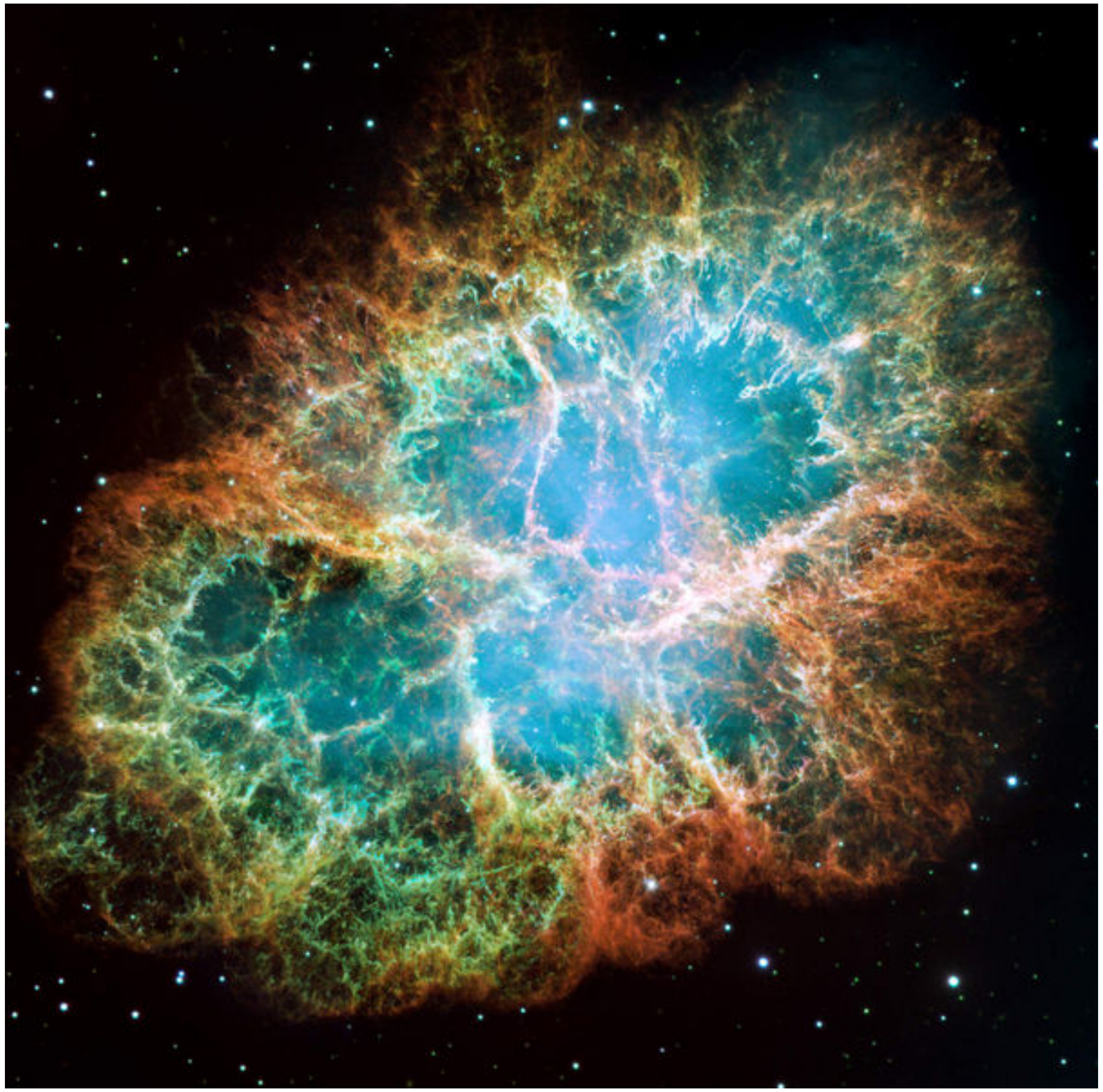


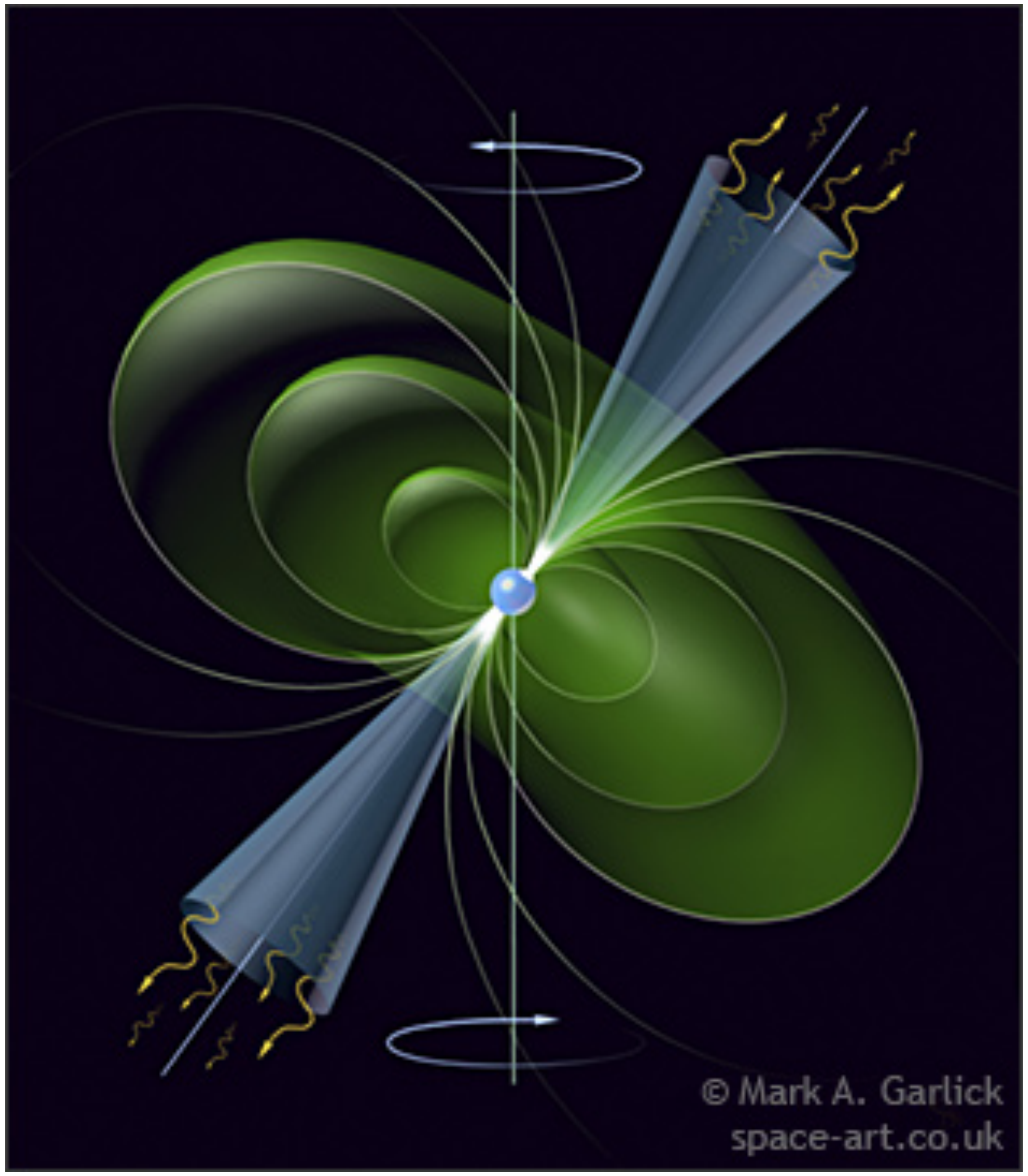

v (km/s)
observed
100
50
expected
from
luminous disk
5
10
R (kpc)
M33 rotation curve
(fig. 1)



Synchrotron Emission






© Mark A. Garlick
space-art.co.uk

AGN

Evolution of Stars and Galaxies

- Detailed observations provide insights into how the physical universe evolves.
- Galaxies appear different because they *are* different — different gravitational potentials, stellar mass, gas mass, etc…
- Optical — stellar population, star formation rate.
- Radio — high-energy processes, AGN, gas conditions.
    - Doppler shifts of spectral lines — dynamical information about galaxies.

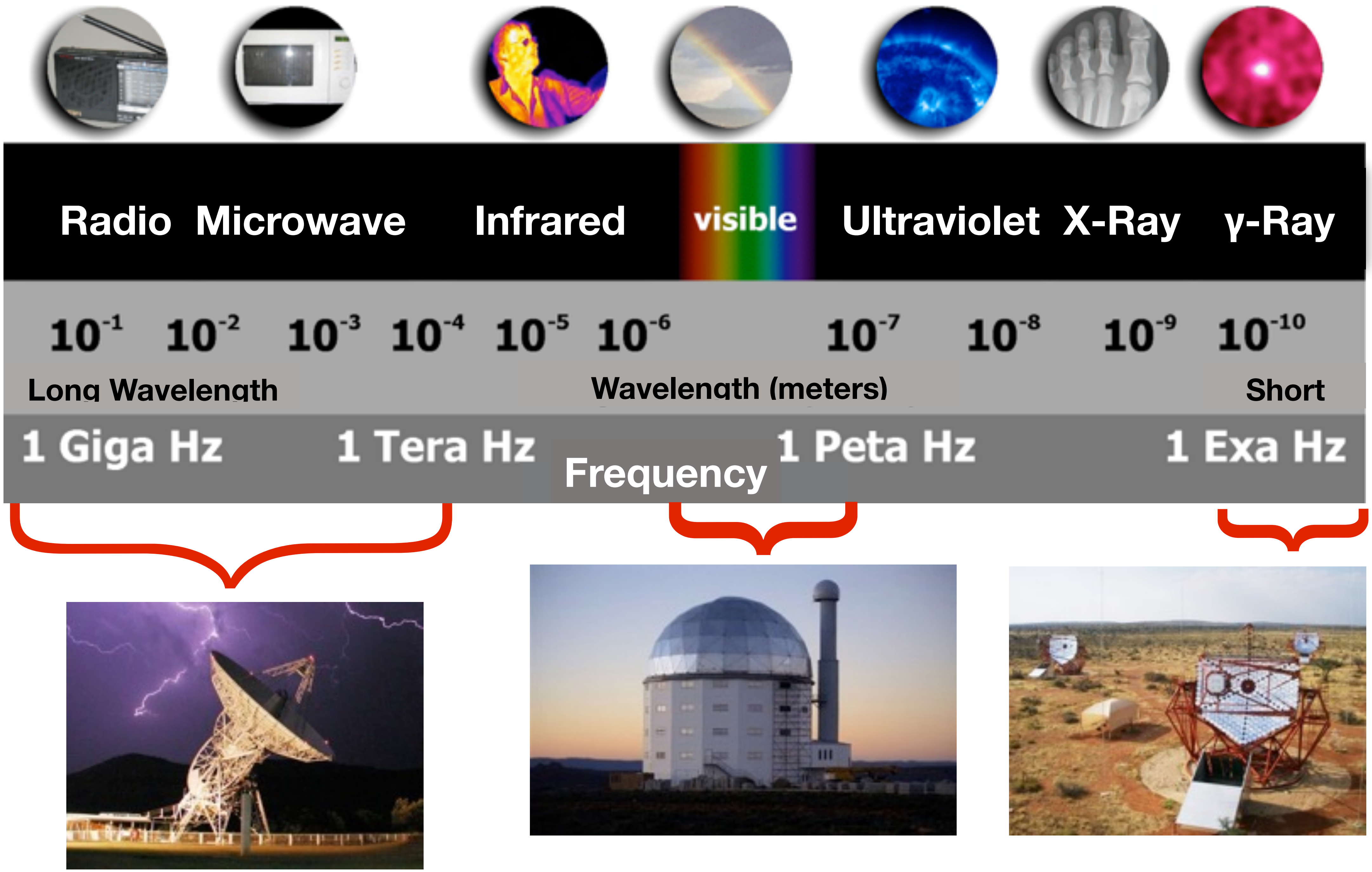
Telescopes

- Buckets — collectors of electromagnetic radiation.
- Focus, collect, detect and amplify astrophysical signals.
- In the optical, telescopes detect and count photons (CCDs).
- In the radio, telescopes detect waves (antenna receivers).
- This lecture — radio telescopes.
  - Optical telescopes/astronomy will be presented shortly.

Photons

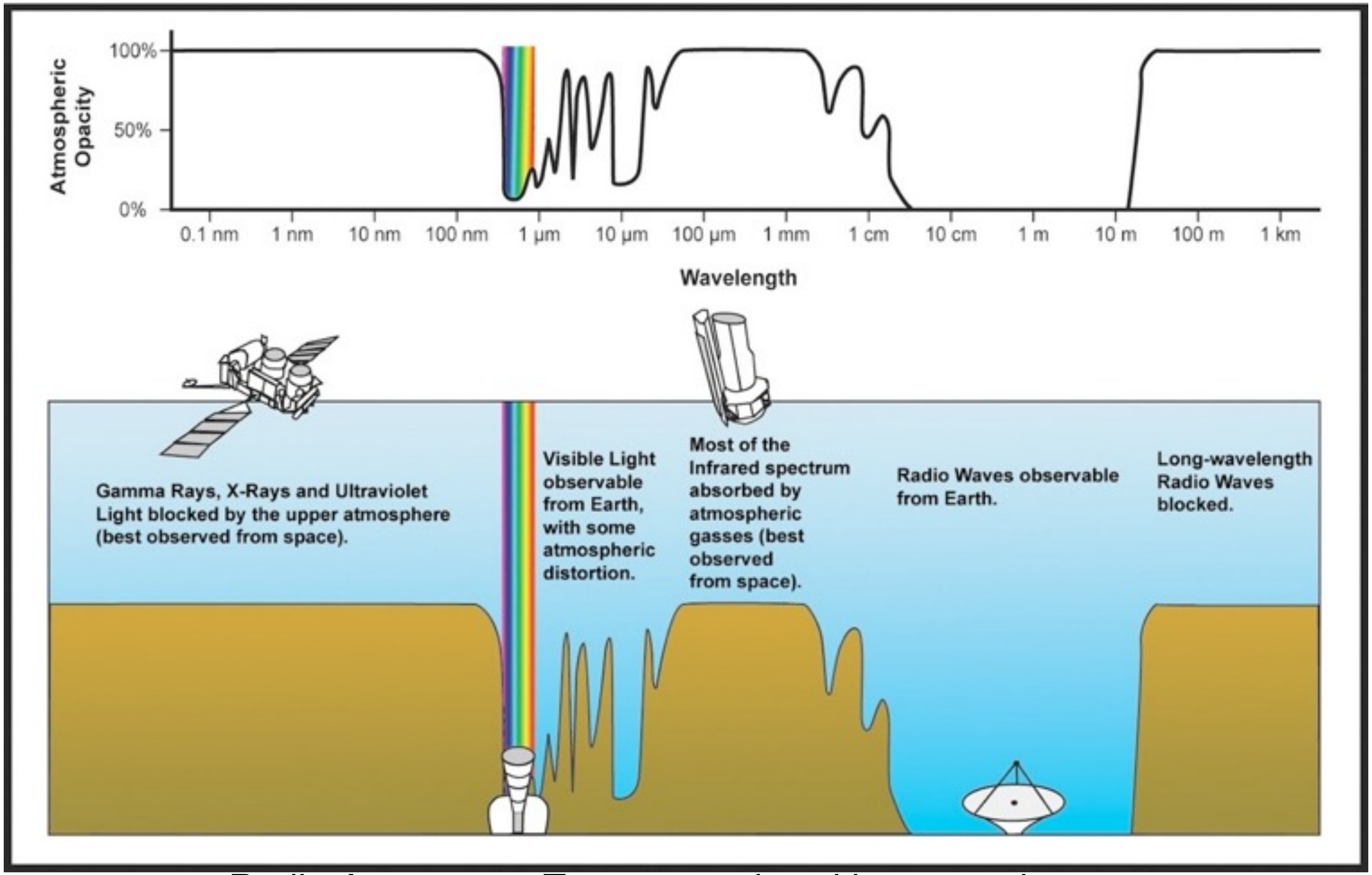
- Electromagnetic Radiation.
- Energy $E=h\nu$
  - h Planck's constant 6.63×10^{-34} J. s.
  - $\nu\lambda=c$, ν (frequency), λ (wavelength)
    - $c = 3\times10^{8}$ m. s.$^{-1}$
- Short wavelengths ($<10^{-6}$ m), photons behave like particles.
- Long Wavelengths ($>10^{-3}$ m), photons behave like waves.

EM Spectrum



Radio
Microwave
Infrared
visible
Ultraviolet
X-Ray
γ-Ray
10^-1 10^-2 10^-3 10^-4 10^-5 10^-6 10^-7 10^-8 10^-9 10^-10
Long Wavelength
Wavelength (meters)
Short
1 Giga Hz
1 Tera Hz
Frequency
1 Peta Hz
1 Exa Hz

Atmospheric Opacity

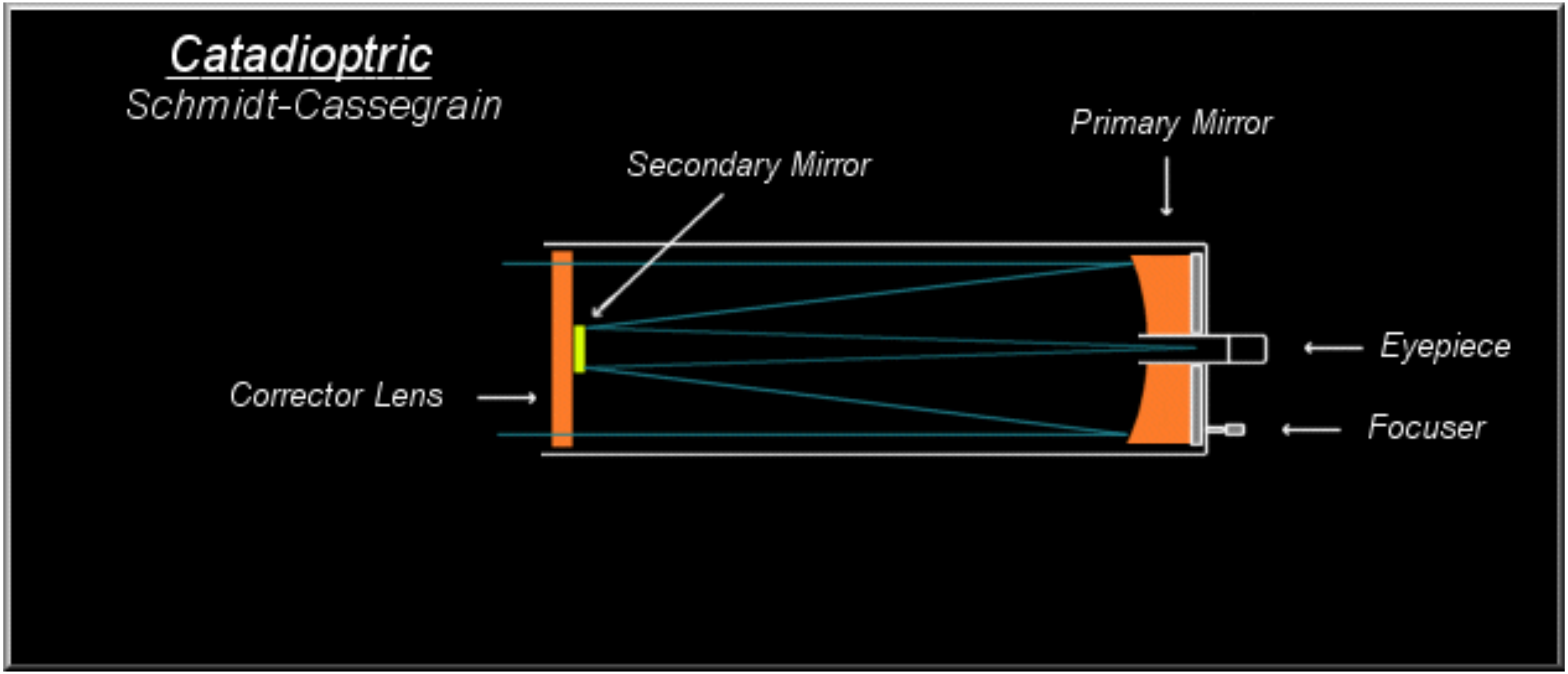


Radio Astronomy: Transparent band in atmosphere.

[cm] ~ [m] in wavelength.

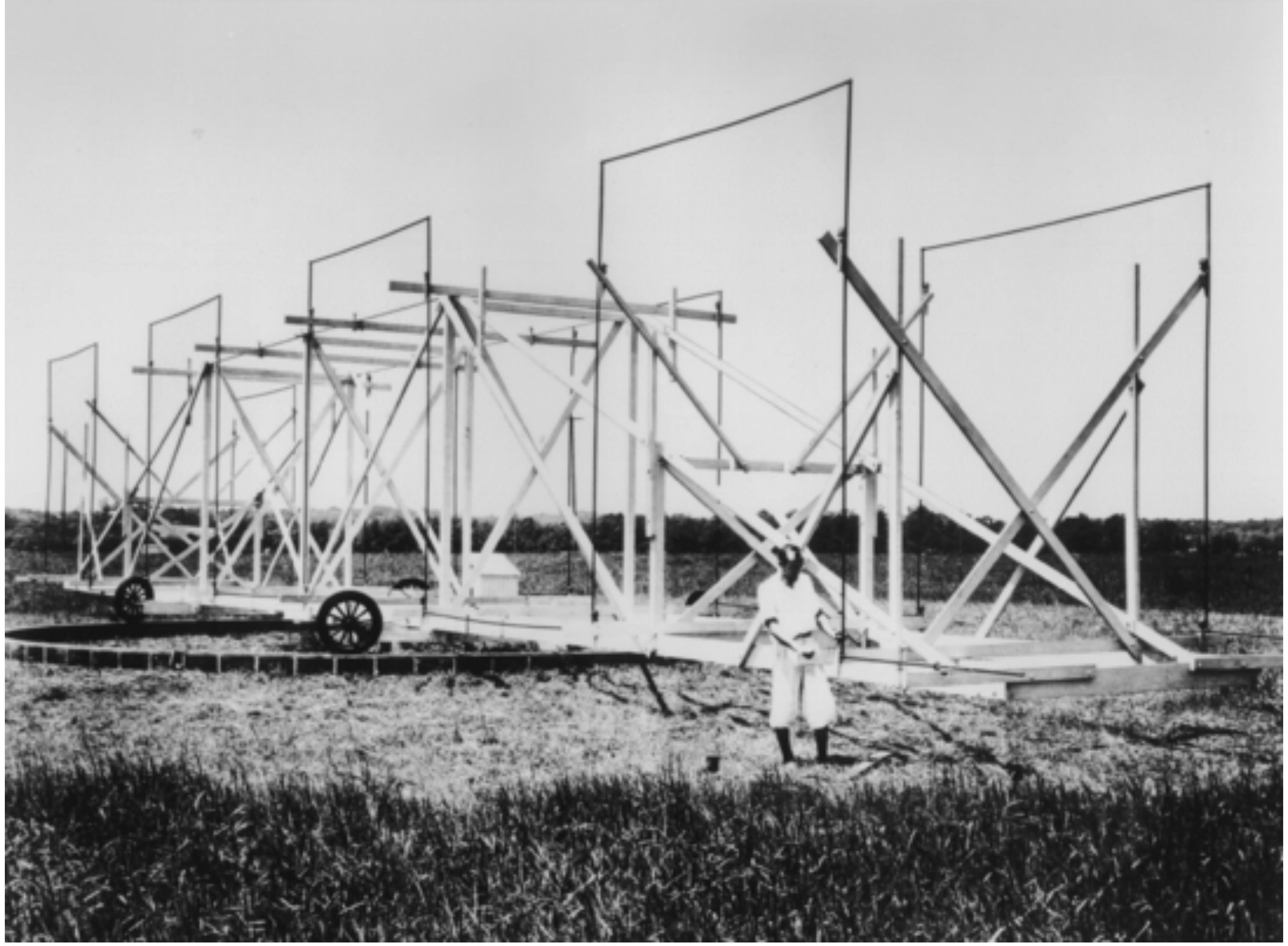
[GHz] ~ [MHz] in frequency.

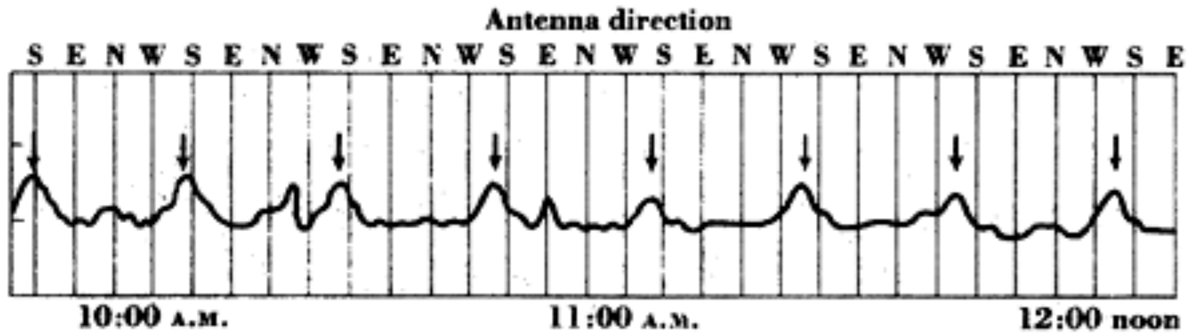
Simple Telescope Schematic



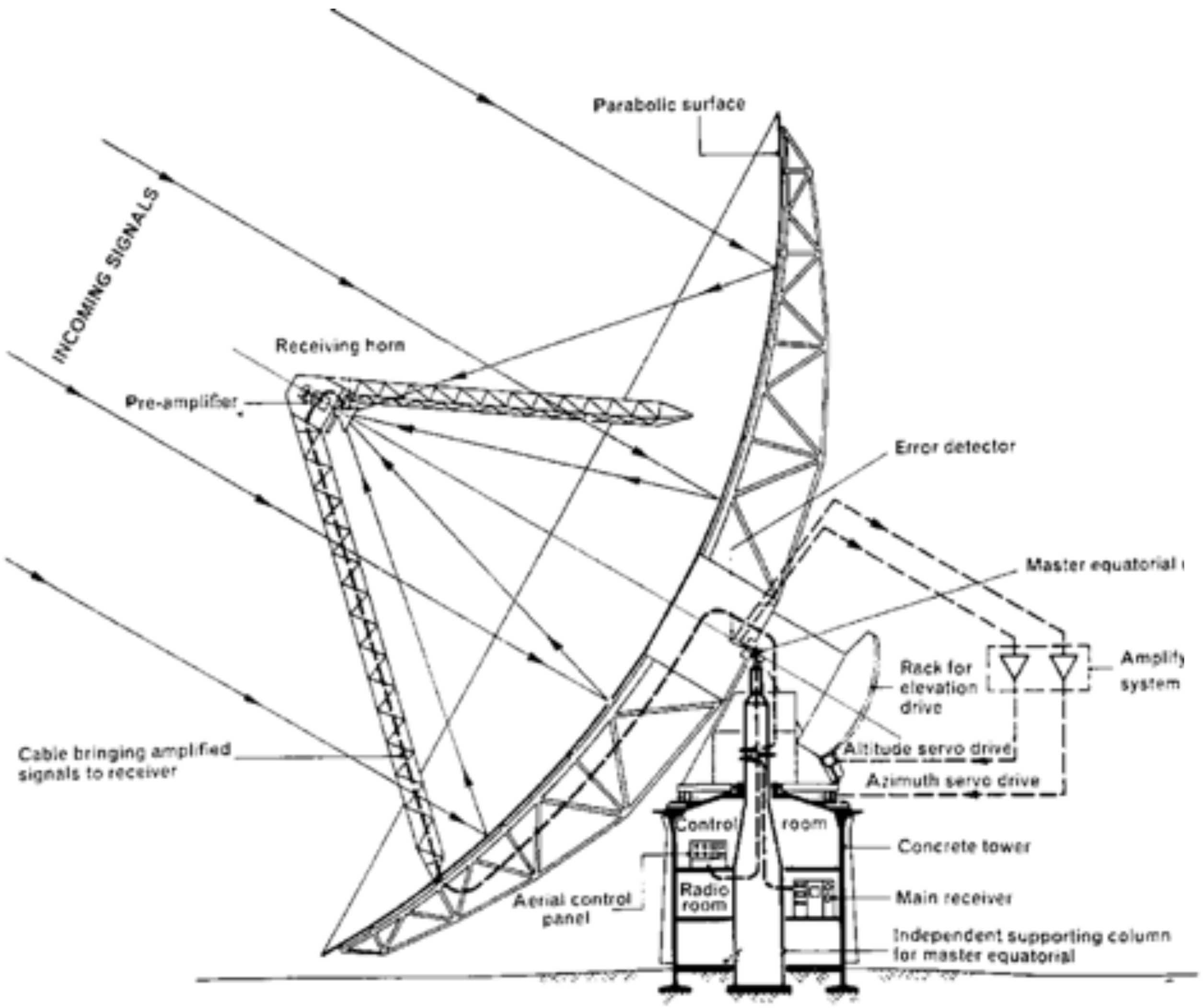
CFHT Telescope and Early Images

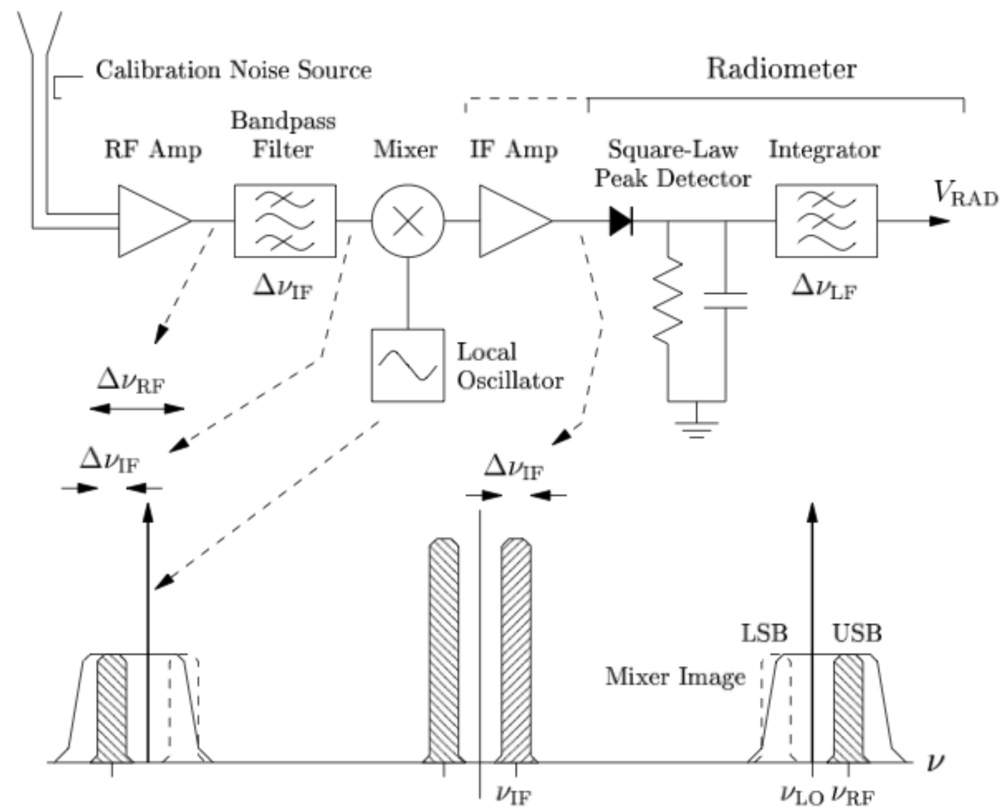
Karl Jansky and the Portrait of the Milky Way

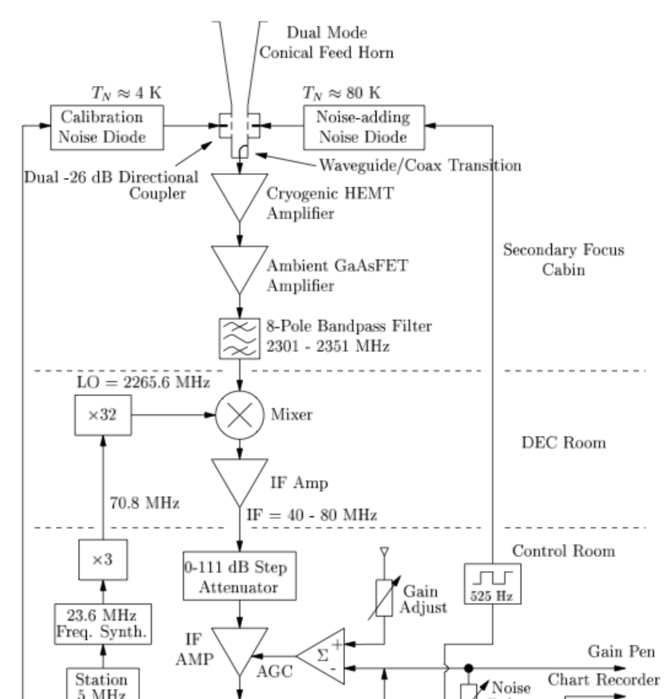




Radio Telescope Schematic







Radio Telescopes

- Bigger = Better for telescopes, in general.
- More sensitivity, i.e., collect more photos/EM energy.
- High resolution, i.e., better at resolving sharp features.
- Practical limitations to building large telescopes.
- Single structures are difficult/expensive to build/operate and maintain.

Radio Telescopes

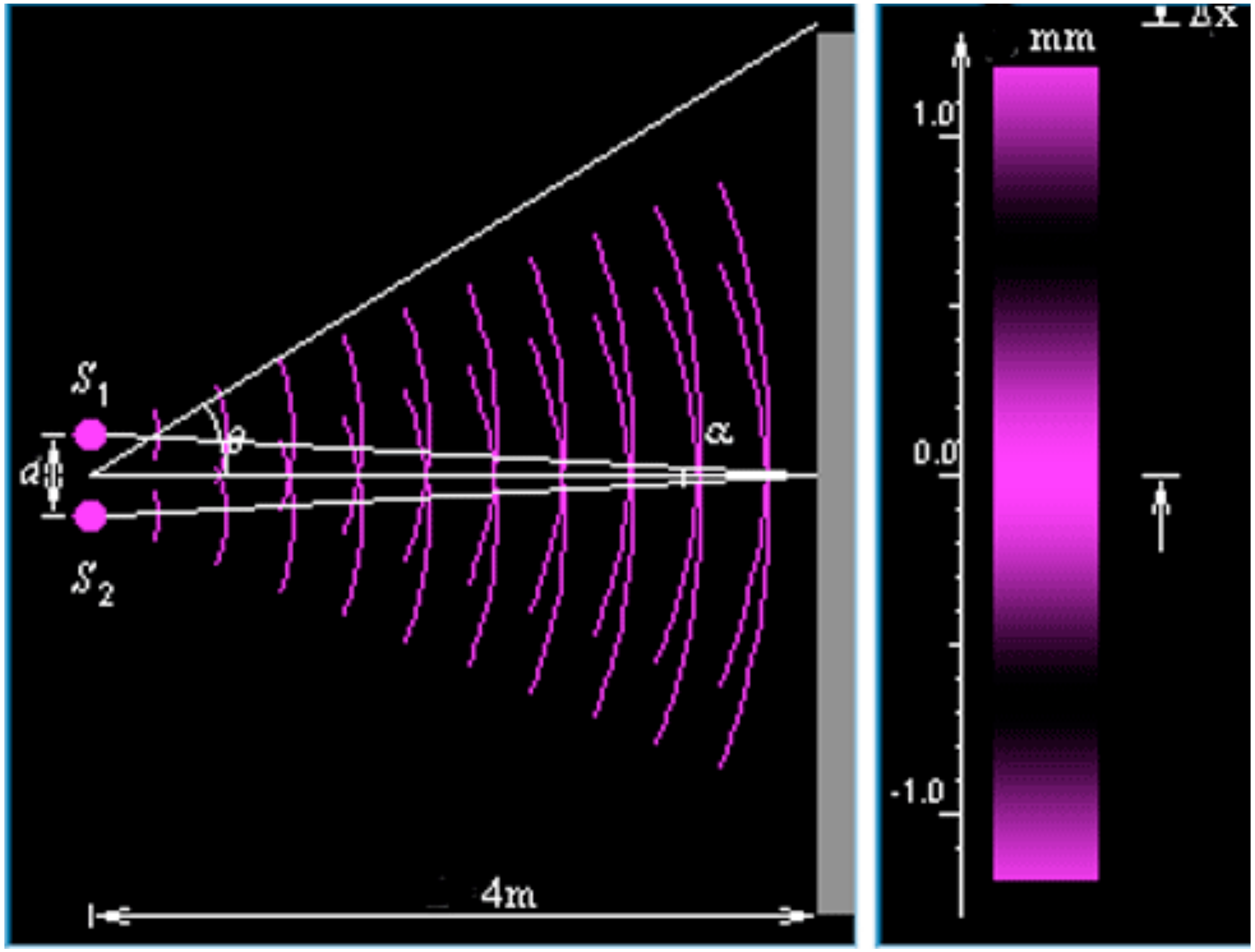
- Large optical telescopes (> 6m) can achieve excellent resolution and sensitivity.
- At long (radio) wavelengths, this becomes tricky:
  - Resolution is limited by the size of the telescope.
  - The human eye has a resolution of ~ 0.34′.
  - At 21cm, a radio telescope has be 2500m in diameter to have the same resolving power as the human eye.
  - Most interesting things are far away, or really small!

Radio Telescopes — Build ‘em Big
FAST — Five-hundred-meter Aperture Spherical

Build Lots of ‘em
VLA — Inteferometer

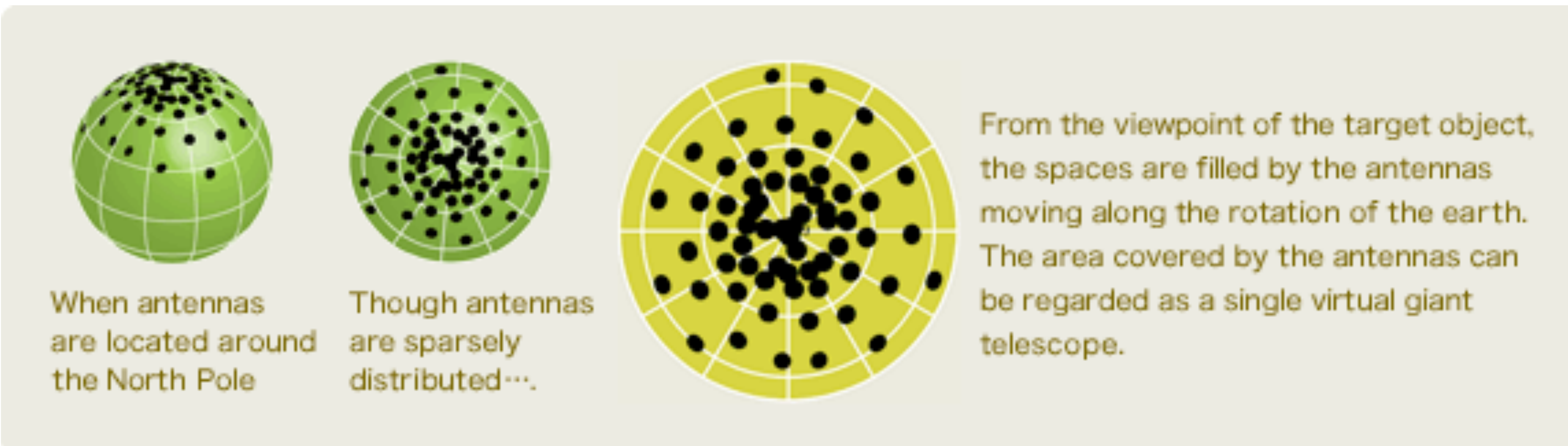
- Collecting Area = Sensitivity
- Diameter = Resolution

Young's Double Slit Interferometer



Radio Interferometer

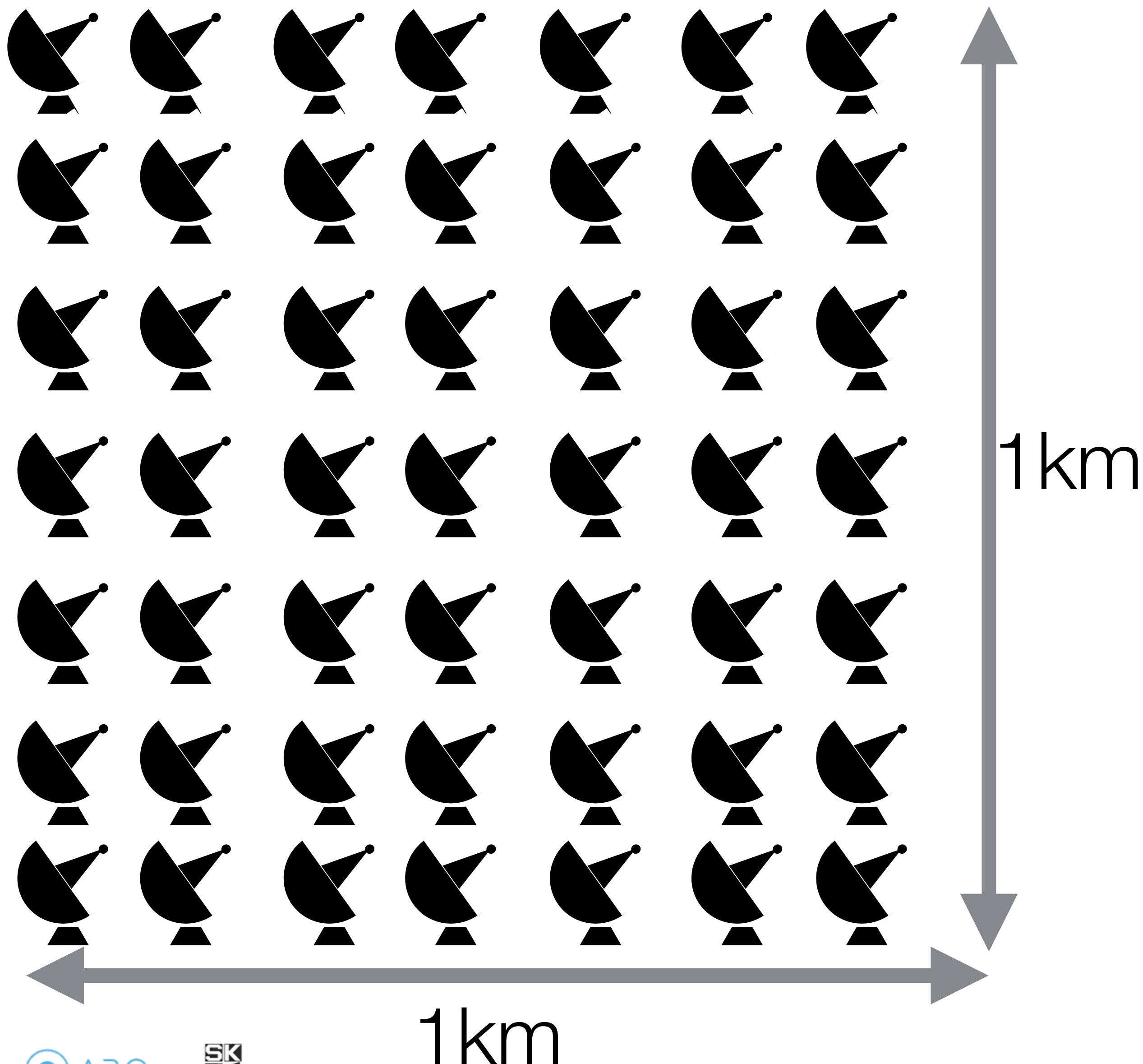
- Many radio telescopes (dishes/antennas) form multiple instantaneous “slits” or baselines.
- Many more baseline are created as the earth rotates.
- The “interference” pattern is measured - not the image!
- The physical image can then be reconstructed with a careful, mathematical analysis of the interference pattern.
- You can have lots of telescopes (sensitivity) scattered across large areas (resolution).



* The actual ALMA antenna location differs from the figure above. The figure is a conceptual illustration to explain the principle of the "aperture synthesis" technique (interferometric imaging method) in a very simple way.

The SKA

- How much collecting area would you need to detect the most distant neutral hydrogen in the universe?



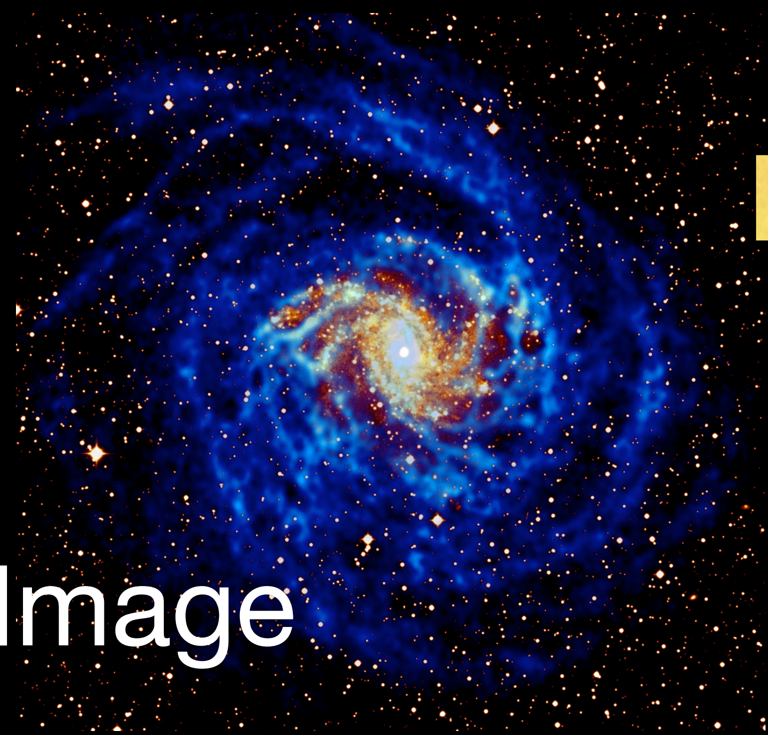
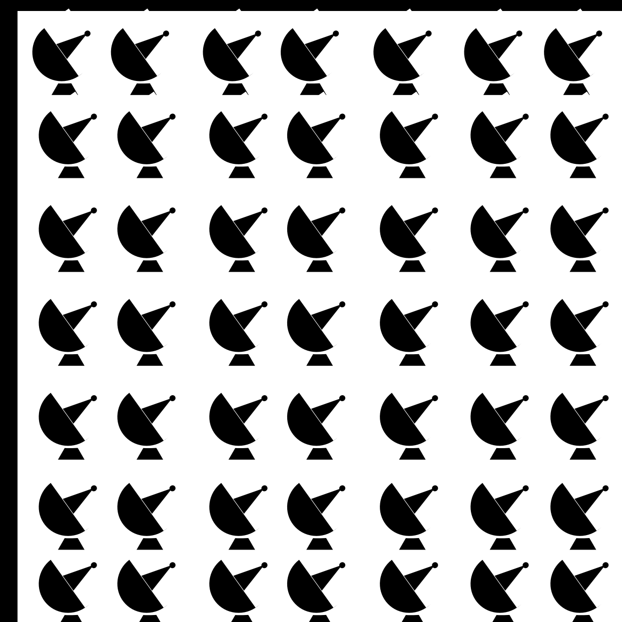
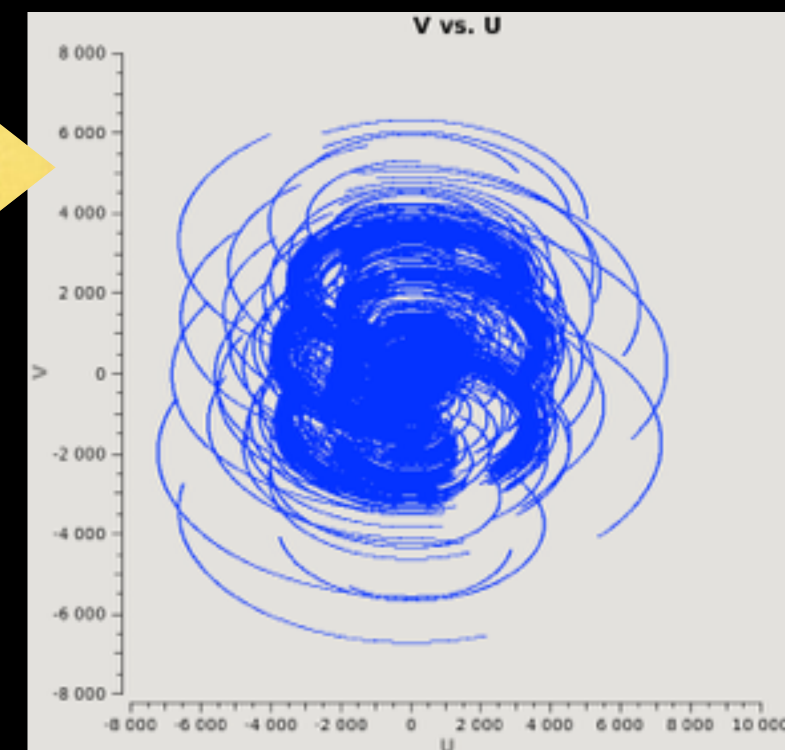


Image Domain



“Synthesized” Aperture



Fourier Domain

A brief history of South African SKA

- 2003: SKA - Square Kilometre Array, International Project started.
- 2004: Karoo Array Telescope - concept and design as part of SA bid.
- 2006: South Africa & Australia shortlisted as hosts. Planning and construction of pathfinders begin.
- 2006/8: MeerKAT - SA government commits to funding and design/development begins.
- 2012: Dual site agreed; Major part of dish-array goes to SA.
- 2017: MeerKAT under construction, MK16 producing data!


12/08/2016 14:00


12/08/2016 14:42


12/08/2016 15:24


12/08/2016 14:59

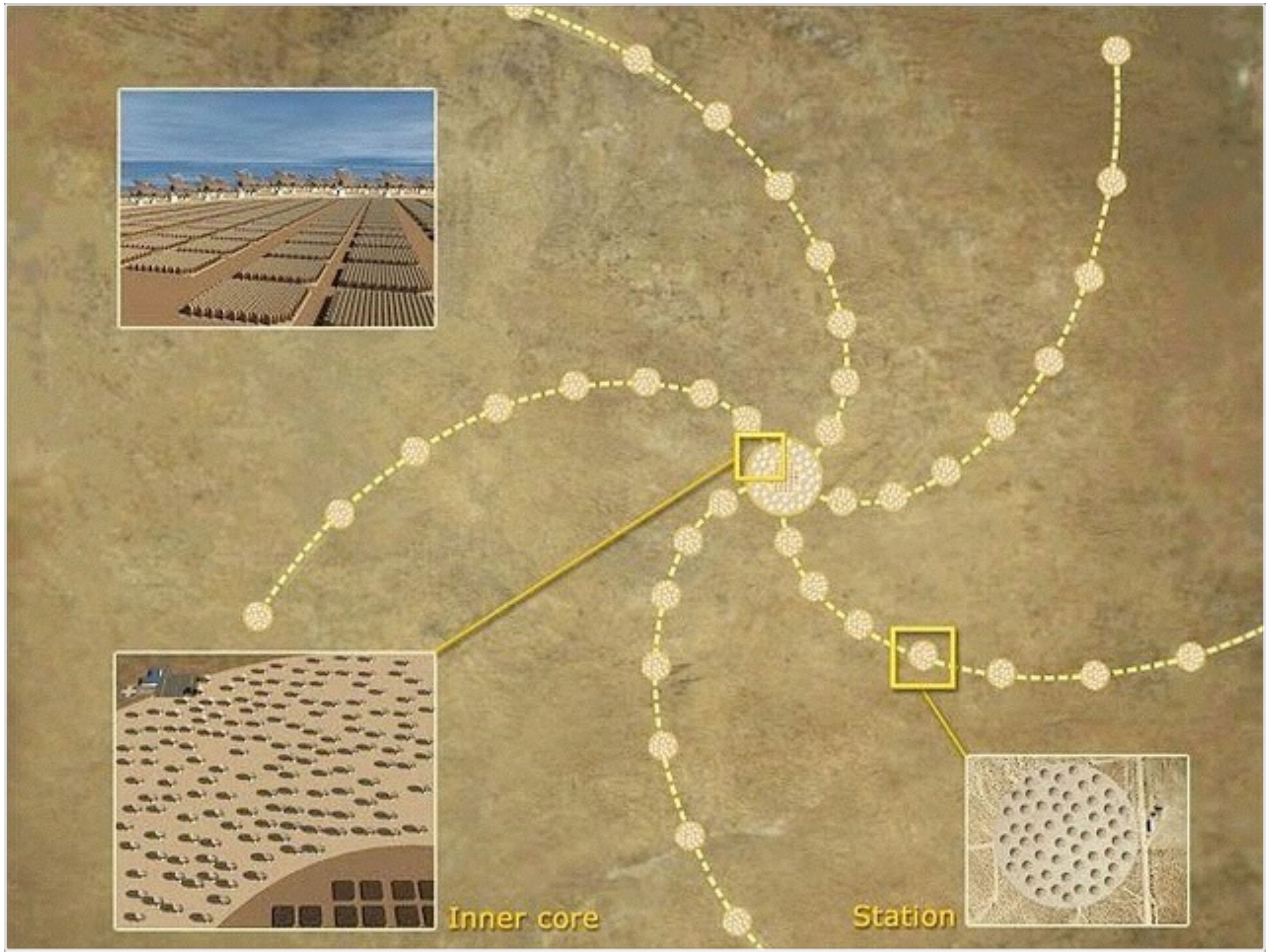

12/08/2016 12:23


12/08/2016 12:23

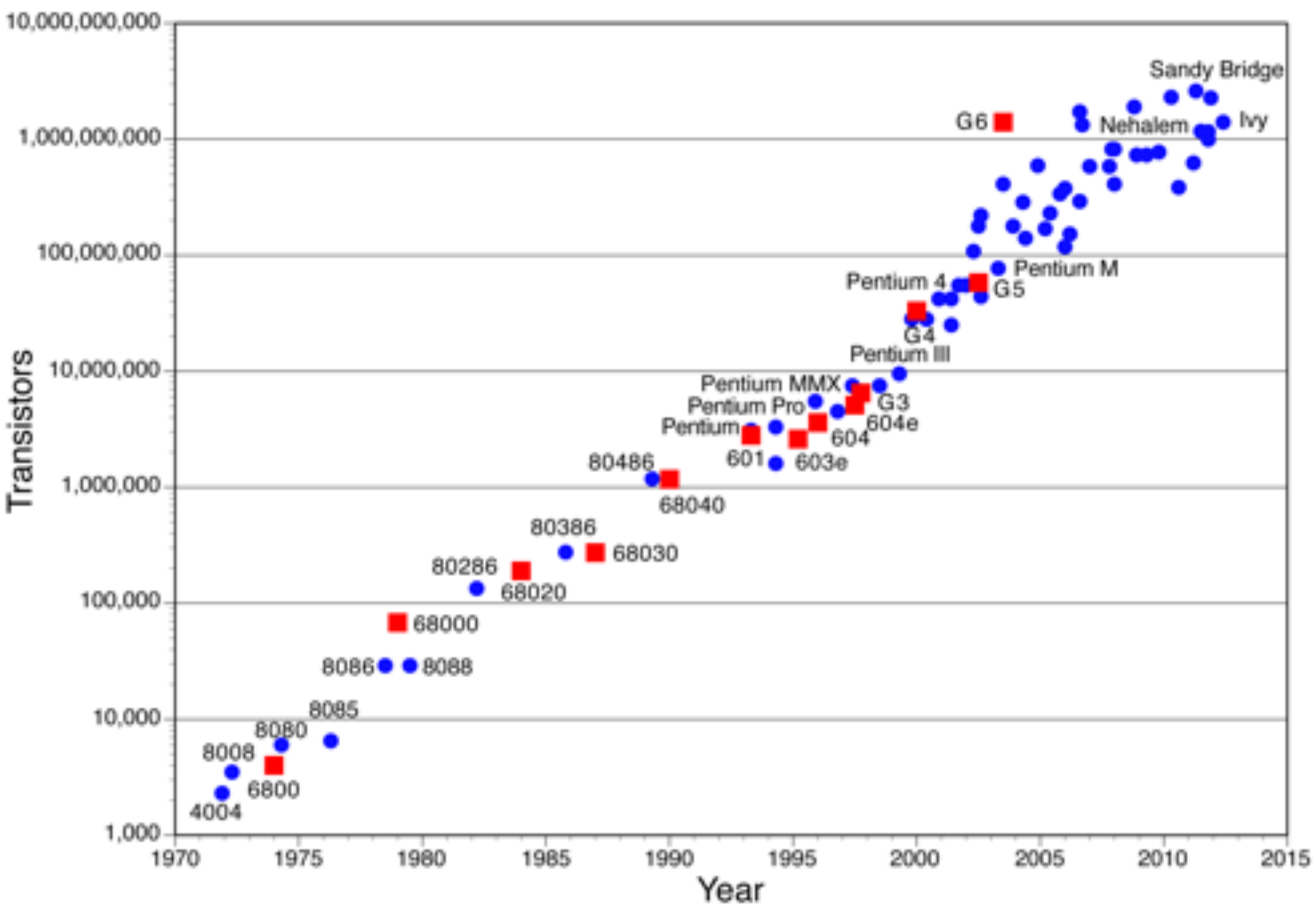

12/08/2016 13:22

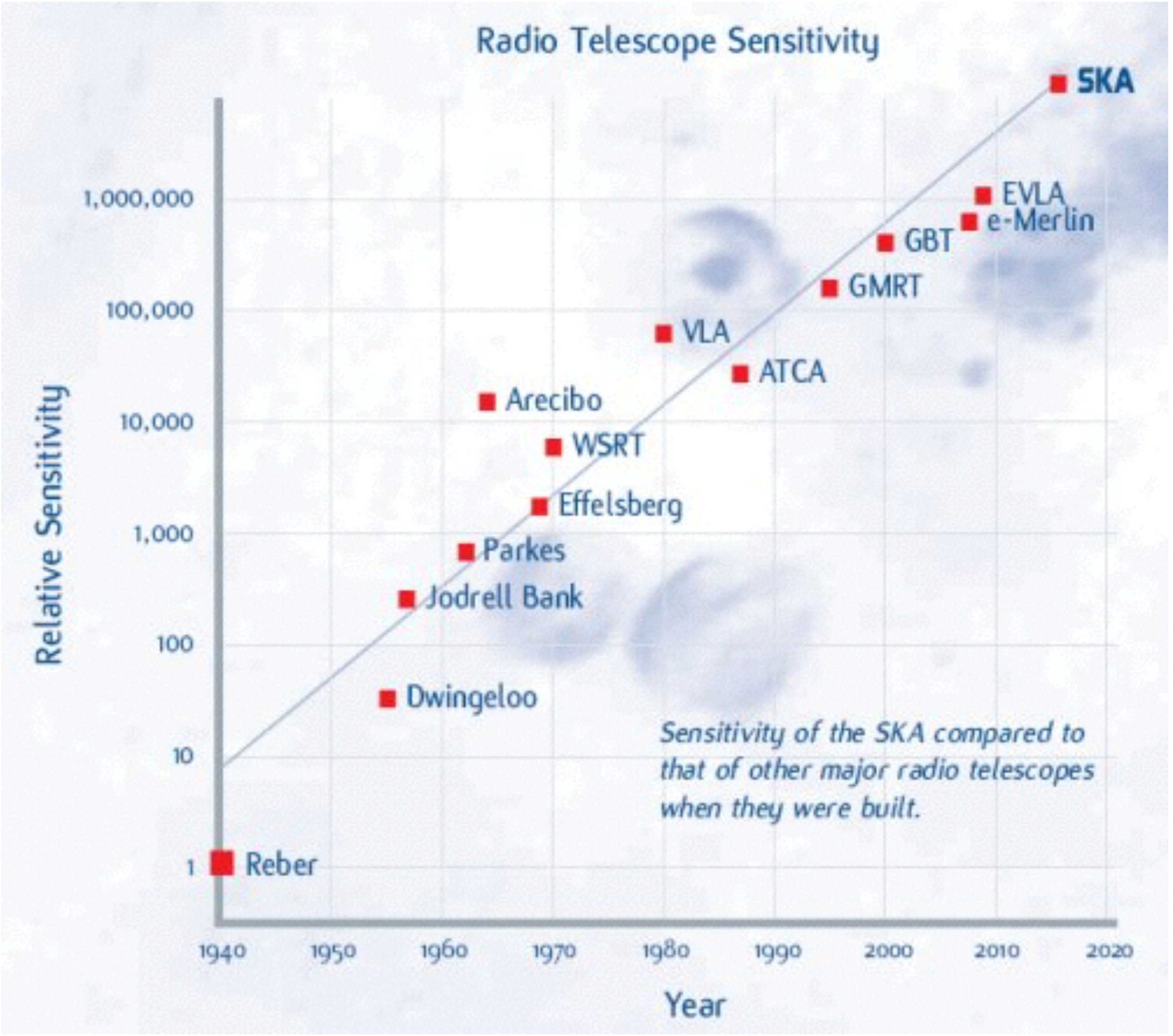

MAX SWL 9000KG
12/08/2016 12:16


EURO-DIESEL
K12815-AA
EURO-DIESEL
KS-VISION


Inner core
Station

Moore's Law and Telescope Sensitivity

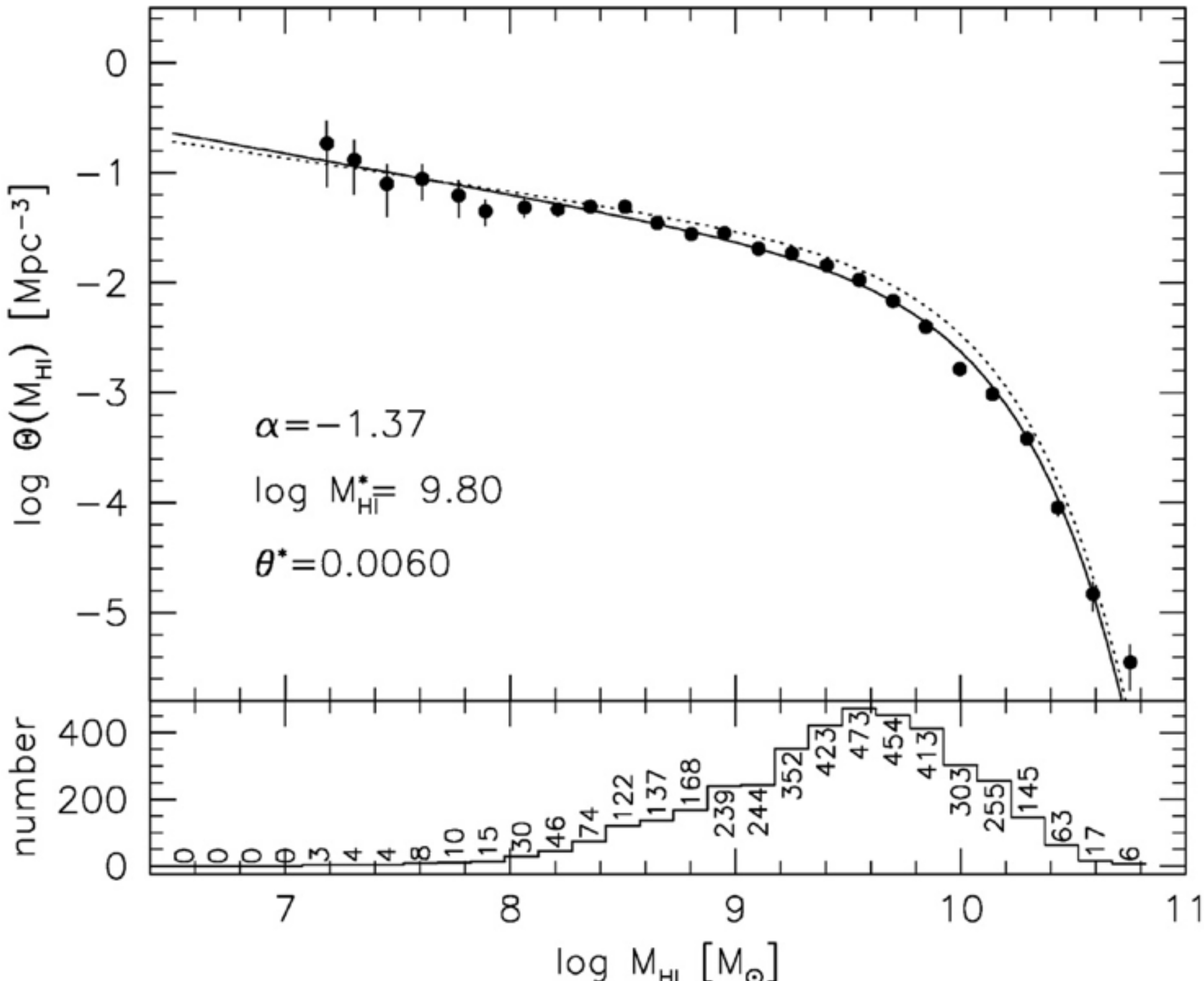




Astronomy and Big Data

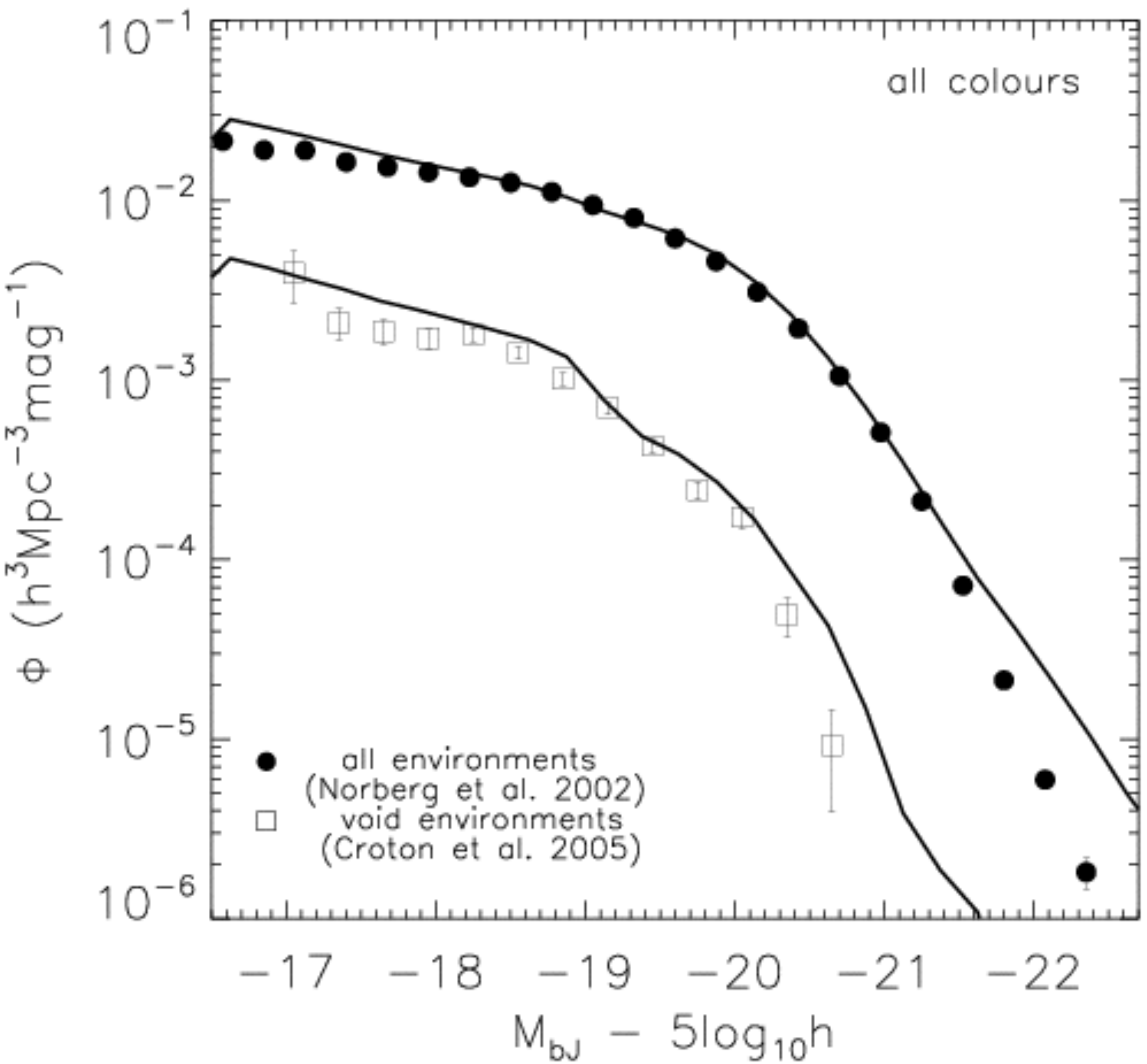
- In the era of precision astronomy. Telescopes are more precise and sensitive than ever before.
- Not just single objects, but complex systems comprising billions of entities.

HI Mass Function



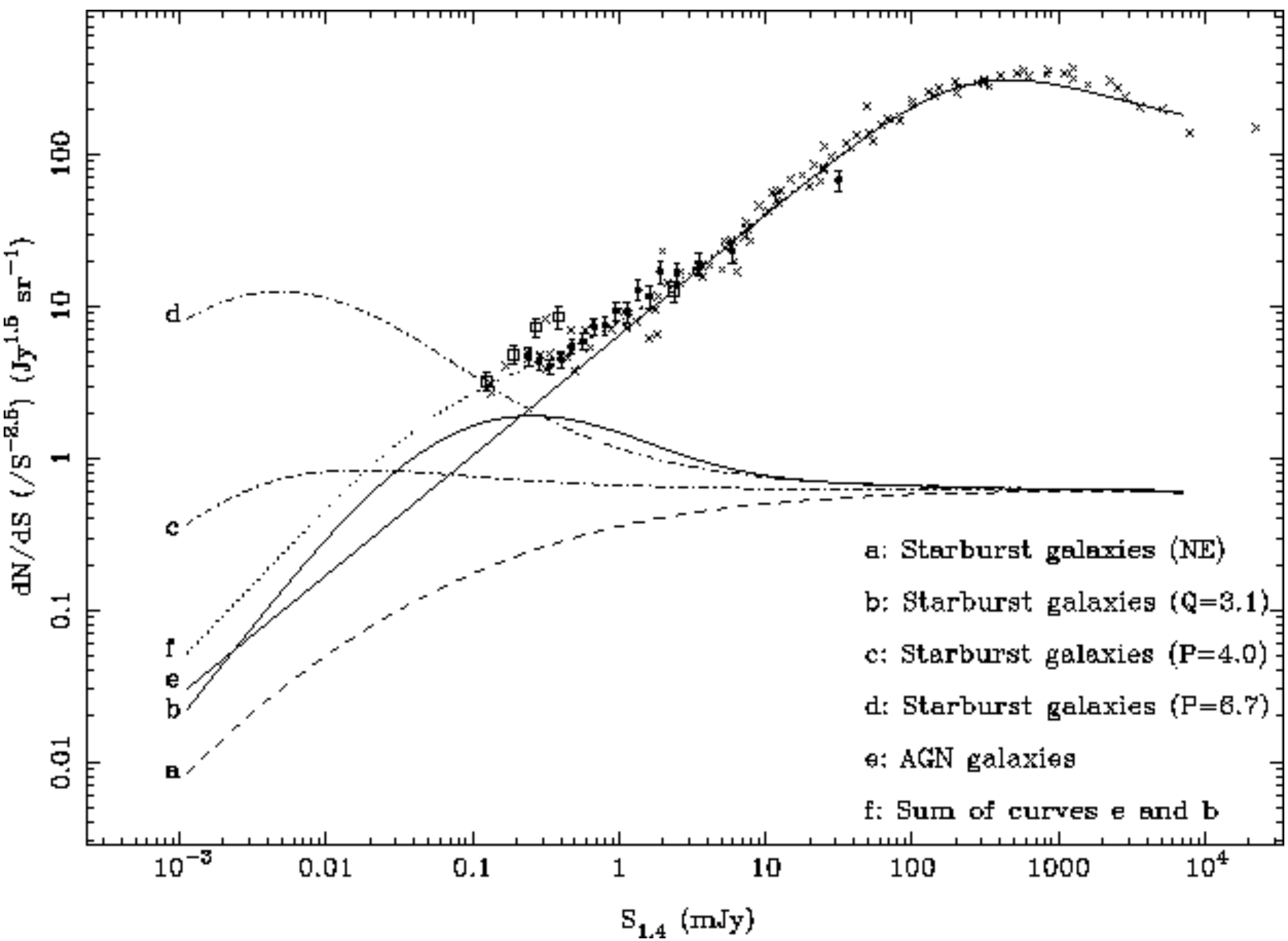
http://astrosun2.astro.cornell.edu/

Optical Luminosity Function



https://inspirehep.net/record/777685

Radio Source Counts

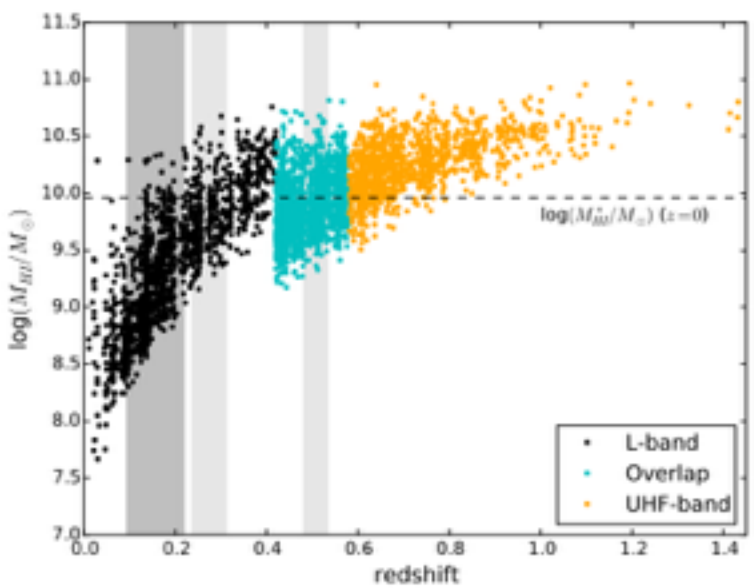


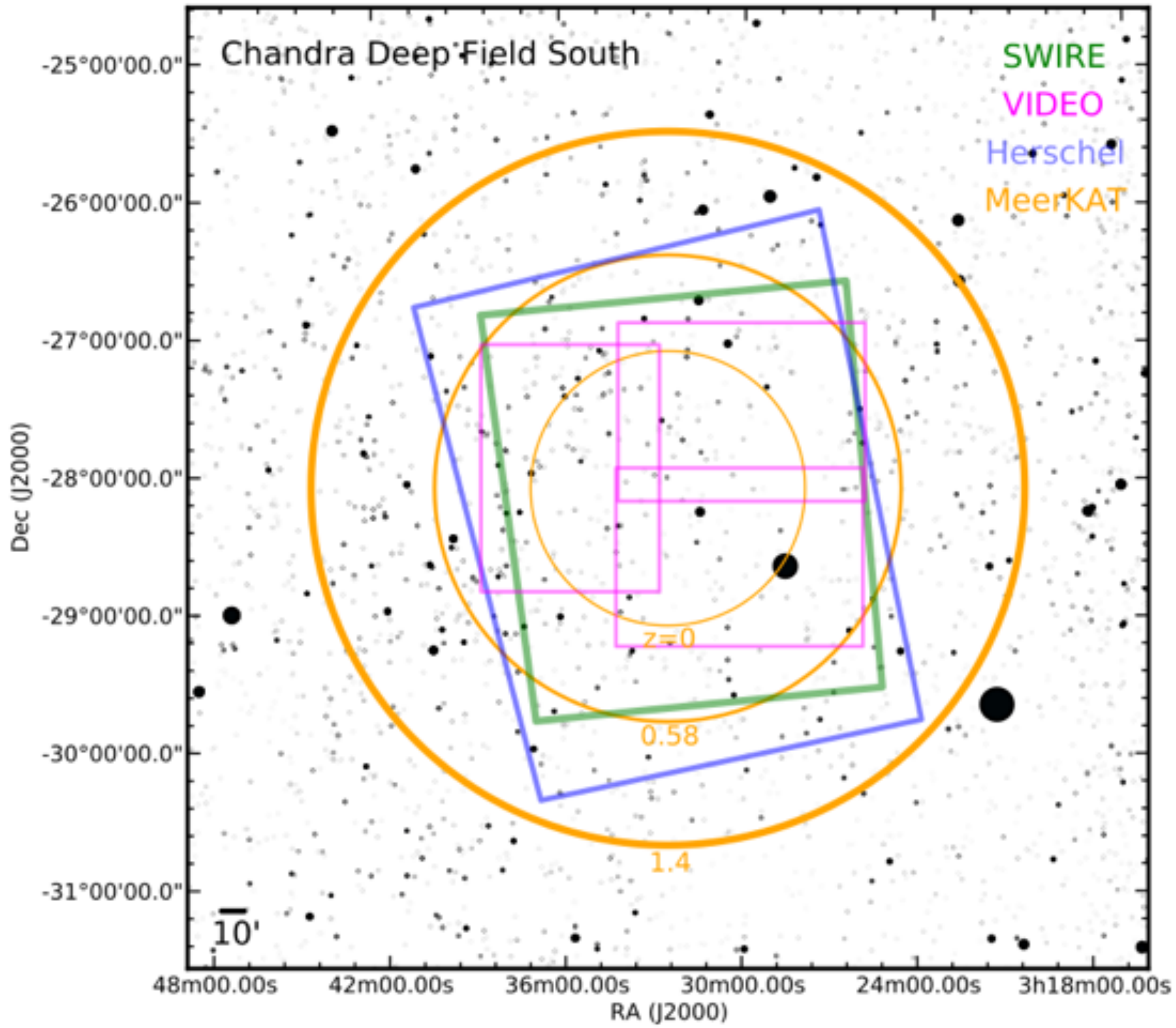
MeerKAT Large Surveys

- Galaxy Evolution
- As a function of cosmic time (z)
- Radio Continuum.
  - Star formation.
  - AGN/Quasar activity.
- HI
  - Dynamics - Content and Shape of Dark Matter Halo.
  - Gas accretion through cosmic web.
  - Neutral gas supply and star formation.
  - Conditions of galaxies in different environments - isolated, groups, clusters.

LADUMA

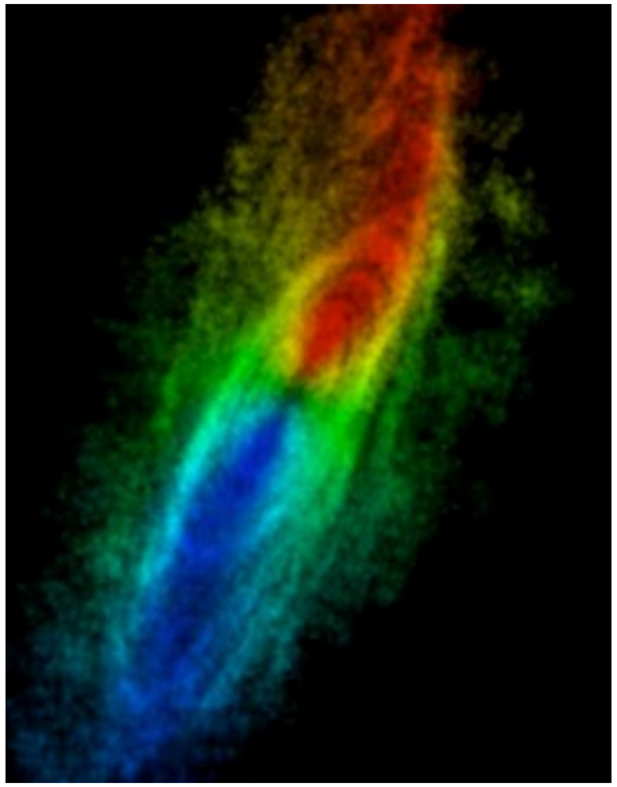
- PIs - Sarah Blythe (UCT), Andrew Baker (Rutgers), Benne Holwerda (Leiden)
- 5000-hours on a single field, looking for high redshift (really distant) HI, evolution of galaxies, statistical studies of gas content in galaxies.

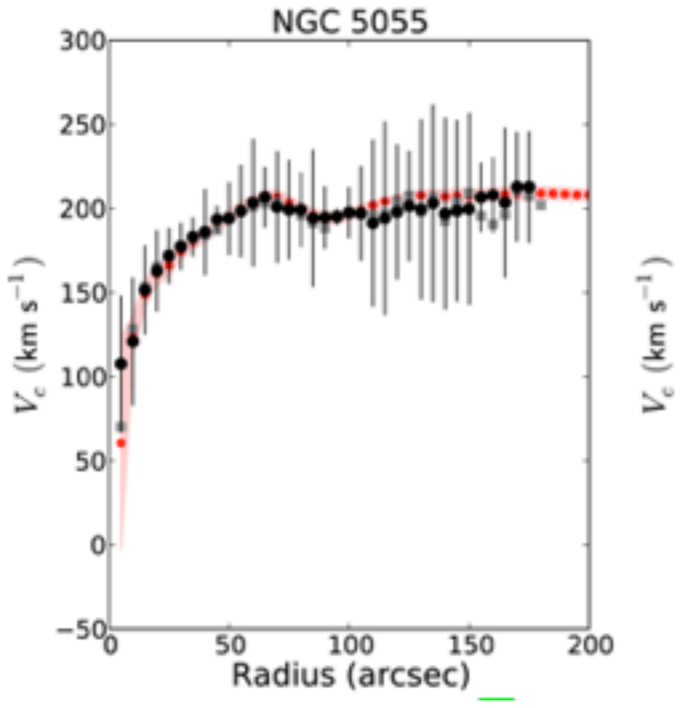


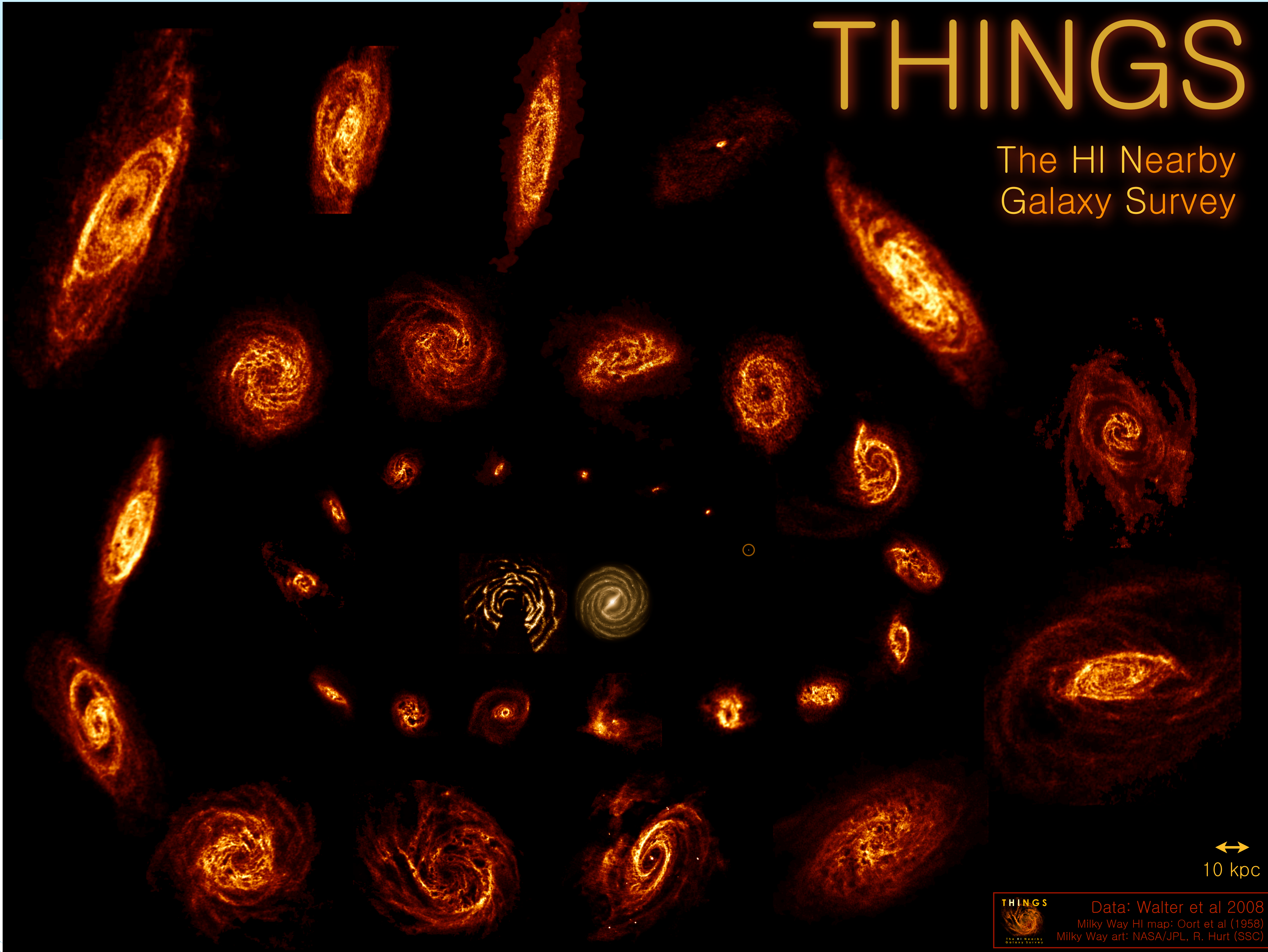


MHONGOOSE

- PI: Erwin de Blok (ASTRON)
- 30 Galaxies
- High resolution / high sensitivity.
- ~sub-kpc resolution.

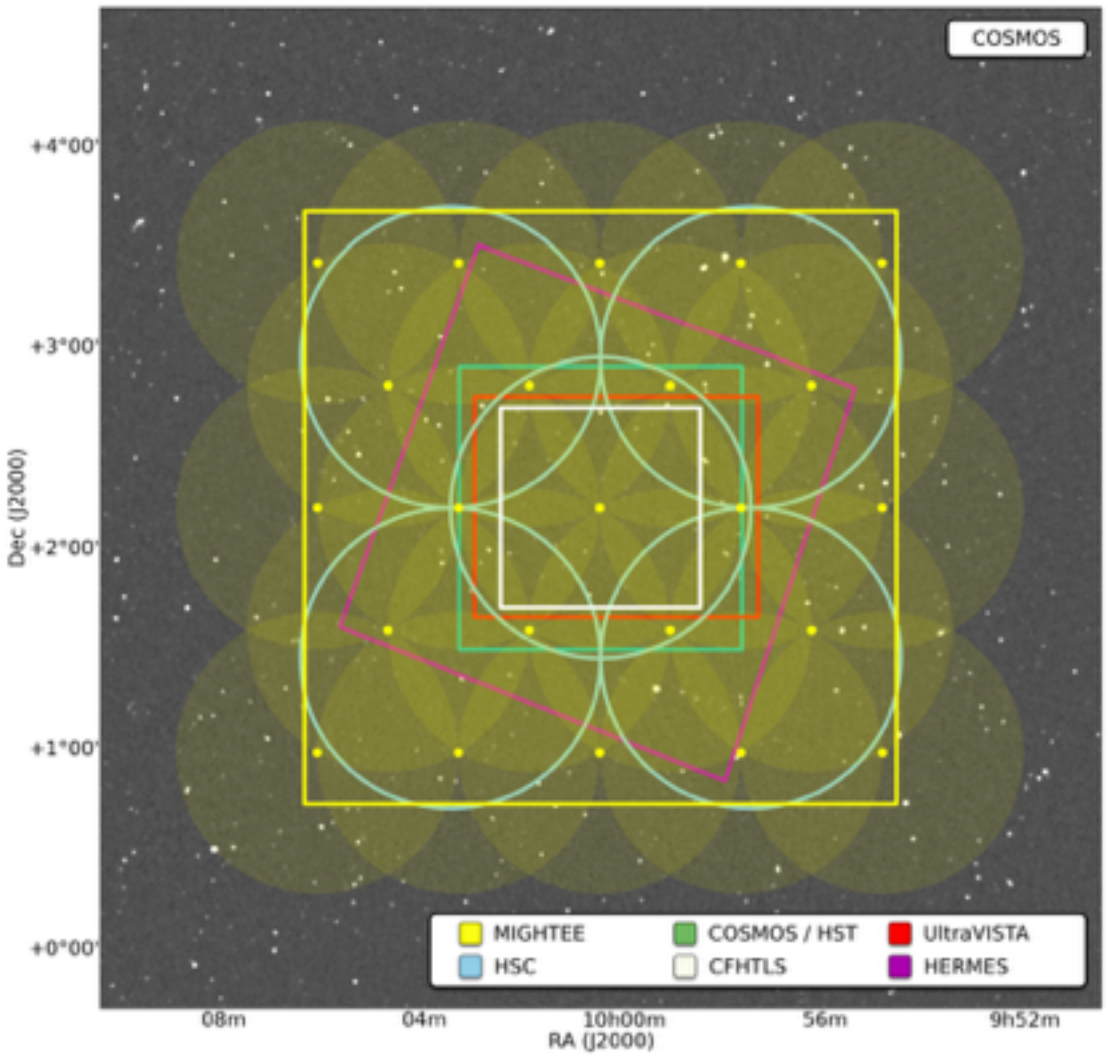






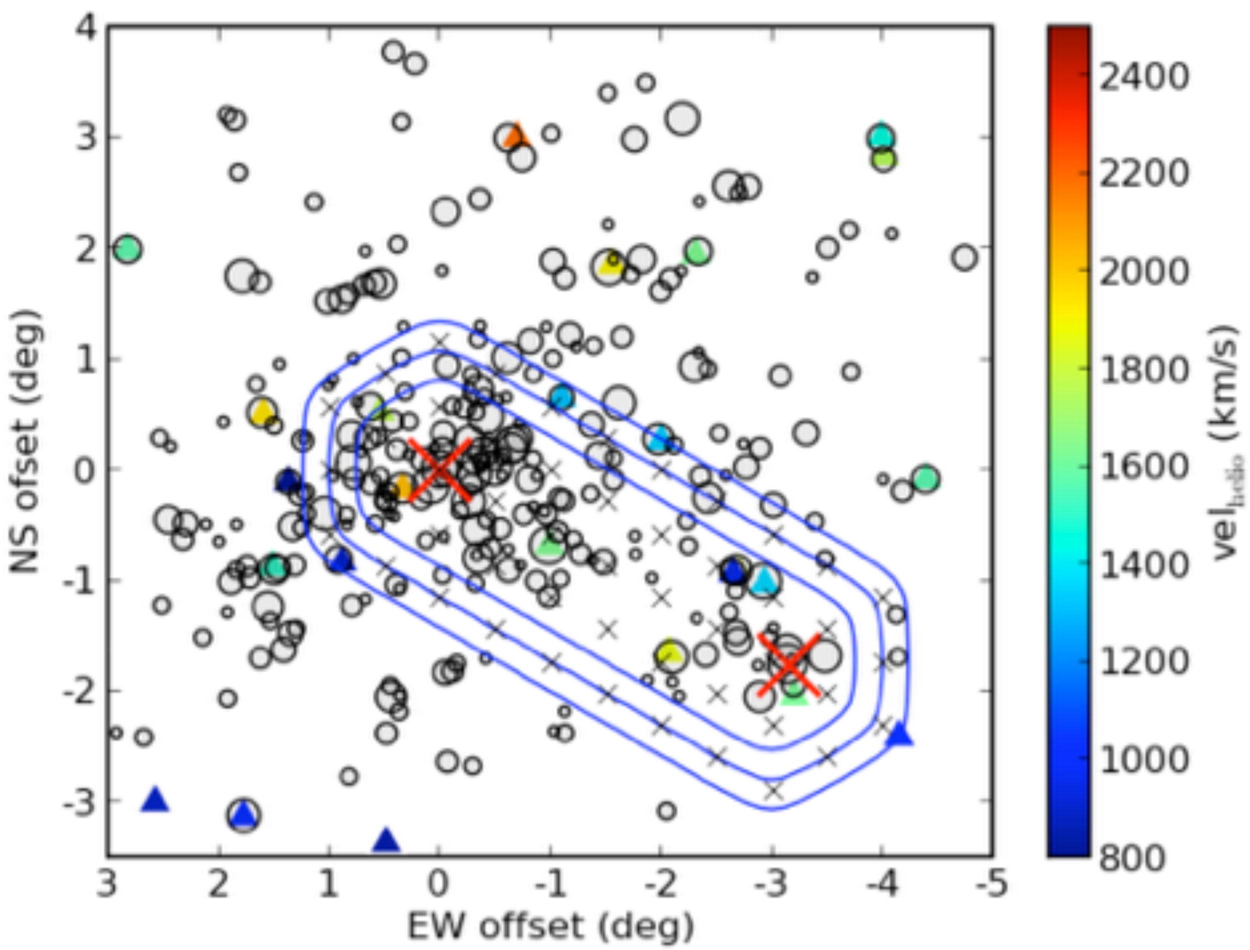
THINGS
The HI Nearby
Galaxy Survey
10 kpc
THINGS
The HI Nearby Galaxy Survey
Data: Walter et al 2008
Milky Way HI map: Oort et al (1958)
Milky Way art: NASA/JPL, R. Hurt (SSC)

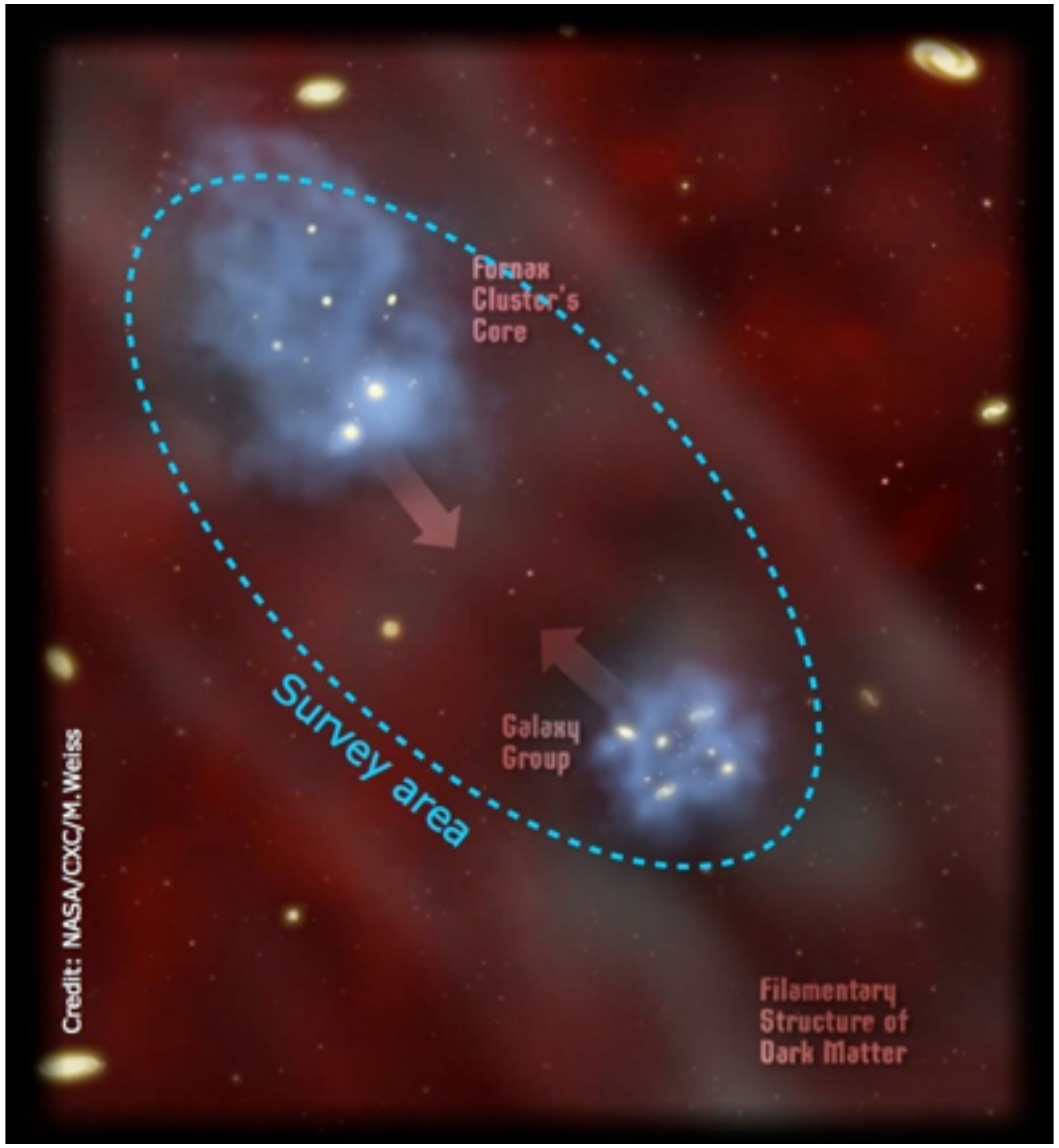
- PIs: Matt Jarvis (Oxford), Russ Taylor (UCT/UWC)
- ~1500 hours on for fields, ~20 deg2.
- AGN/Quasars:
- Radio loud vs Radio quiet.
- Lifecycle of AGN activity - relationship with SF.
- Cold gas supply (HI):
- as a function of AGN type/activity.
- Morphology/Color (red/dead, blue/star-forming)
- Environment - Clusters, Groups, Voids.



The MeerKAT Fornax Survey

- PI: Paolo Serra (INAF, Cagliari)
- ~2000-hours on 11 deg^2 strip of Fornax Cluster





Stripping of neutral hydrogen from galaxies

10^{19} cm^{-2} sensitivity @ 1 kpc resolution

Massive clusters: hydrodynamics

Fornax: gravity?