The ionosphere

Cathryn Mitchell

Professor in Electronic and Electrical Engineering,
NERC Knowledge Exchange Fellow &
Academic Director of the Doctoral College,
University of Bath

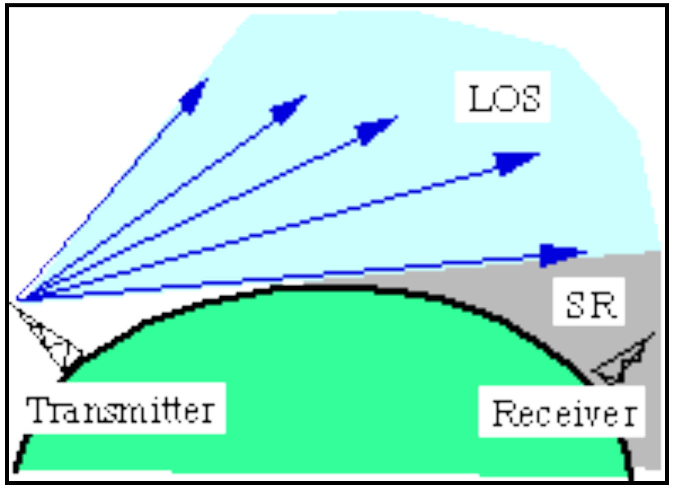
- **Introduction & discovery**
- Physics
- Instruments
- Storm conditions
- System impacts

Discovery of the ionosphere

1839 - Gauss speculated that small diurnal variations in measurements of the magnetic field of the Earth might be caused by “atmospheric electric currents.”

1901 - Marconi transmitted radio signals from Cornwall, England to St John’s Newfoundland in North America.

This transmission was very difficult to understand. Since radio signals were known to be a form of electromagnetic radiation they were expected to travel in straight lines (excluding diffractional effects). In this case the communication link should be prevented by the curvature of the Earth.



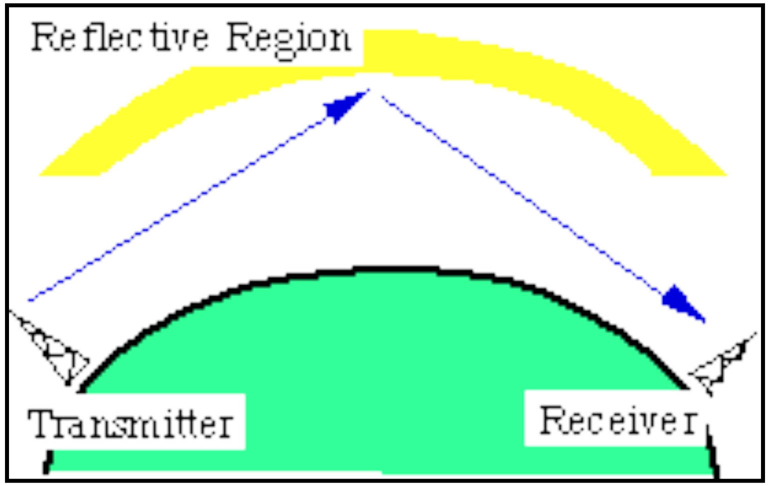
Clearly in Marconi’s trans-atlantic experiment something different was happening.

Discovery of the ionosphere

Early ideas

1902 - Heaviside (UK), Kennelly (USA) and Nagoaka (Japan) all independently suggested that there was an electrically conducting layer in the atmosphere.

"There may possibly be a sufficiently conducting layer in the upper air," - Heaviside



If such a layer existed it would dramatically extend the line-of-sight communications.

Many early experiments were conducted by radio-amateurs who found that frequencies above 2 MHz could be used for long-distance propagation.

Discovery of the ionosphere

Evidence from fading signals

- **1910** - Pierce proposed that radio waves reflected from this conducting layer might interfere with a direct ground-wave.

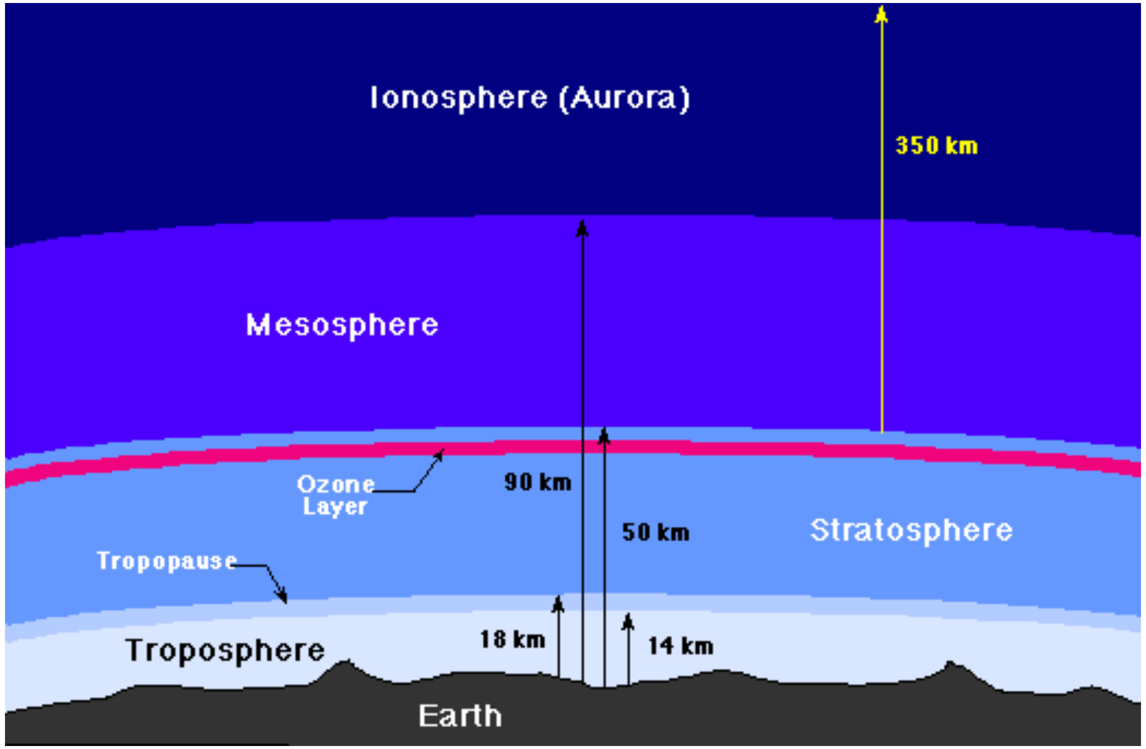
- **1912** - De Forest reported evidence of fading when he received a signal transmitted at 90kHz from Los Angeles to San Francisco.

- **1924** - Appleton and Barnet used a BBC transmitter at Bournemouth UK with a slowly changing frequency. The ground wave and the interfering reflected wave moved slowly in and out of phase. When the path difference was a whole number of wavelengths they combined to produce a strong signal. When it was an odd number of wavelengths they combined to produce a weak signal.

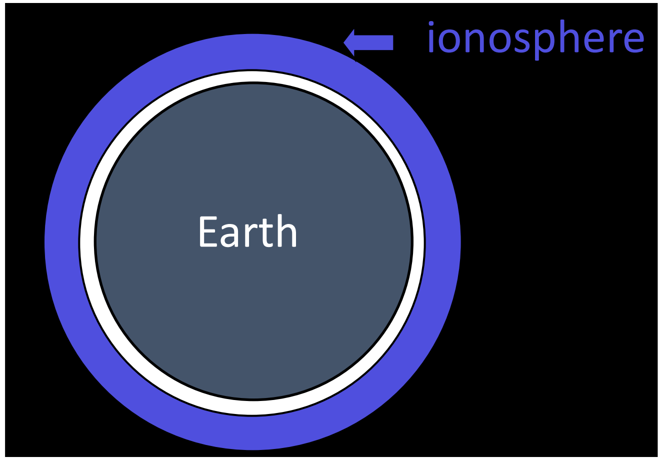
- **1925** - Breit and Tuve (USA) used a pulse sounder technique to directly measure the layer. This technique was a forerunner to modern radar.

- Introduction & discovery
- **Physics**
- Instruments
- Storm conditions
- System impacts

Physical Processes



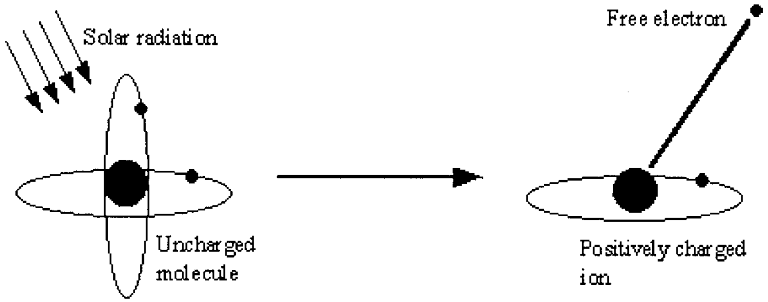
Physical Processes



Production and loss of the ionosphere

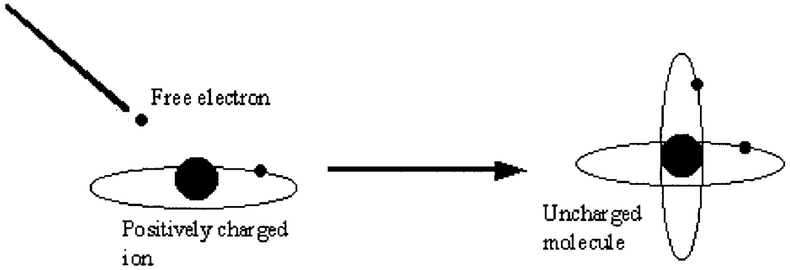
The ionosphere can be considered to be basically a thick shell of free electrons surrounding the Earth, starting at about 90 km altitude and extending to well beyond 700 km altitude.

The ionosphere is produced by the action of extreme ultra violet (EUV) light from the sun ionising neutral atoms of the Earth's atmosphere.



Loss of the free ionisation can be by combination with a positively charged ion.

Other recombination processes occur with more than one stage of charge transfer.

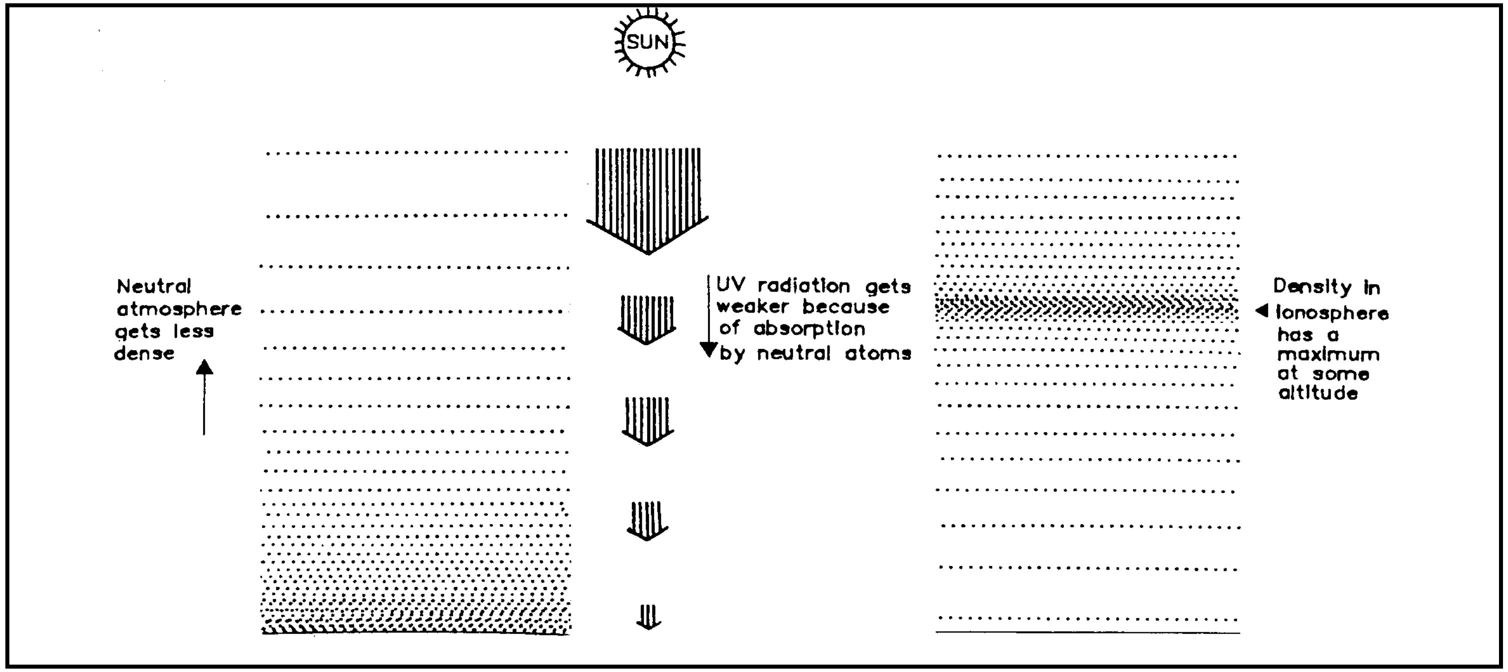


Physical Processes

Vertical structure of the ionosphere

The fact that the electrons form a layer at some altitude in the ionosphere is a combination of two opposing phenomena:

- The amount of EUV light increases at higher altitudes
- The density of the atmosphere decreases with altitude.



Physical Processes

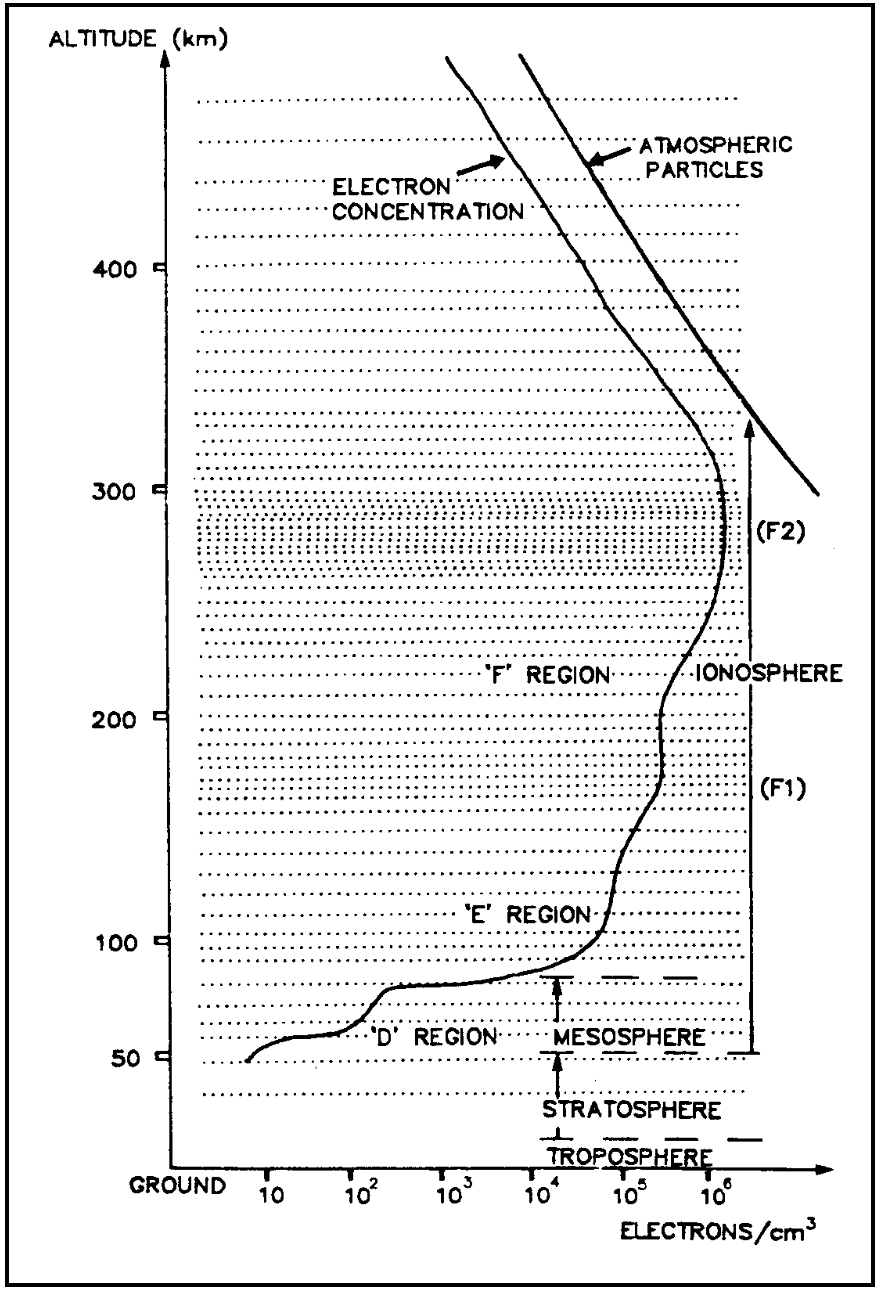
Vertical structure of the ionosphere

Two other factors must be considered:

- The strength of the EUV is not constant at all wavelengths. Different wavelengths of EUV ionised different types of molecules and atoms

- The composition of the atmosphere changes with altitude

These factors result in several layers of ionisation with maxima in the electron density at different altitudes.



Physical Processes

Regions of the ionosphere

During the **day** there may be four regions of the ionosphere present:

- **D region** - from 50 to 90 km
- **E region** - from 90 to 140 km
- **F1 region** - from 140 to 210 km
- **F2 region** - above 210 km

During the **night** only the **F2** region remains.

This is partly due to the fact that the recombination time is longer at higher altitudes where the density is lower.

The F2 region is most important for HF propagation as:

- it is present 24 hour a day
- it allows the longest propagation paths
- it usually refracts the highest frequencies

The ionosphere is not a stable medium which allows the use of one frequency over the year, or even over 24 hours. A frequency which may provide successful propagation now, may not do so an hour later.

Physical Processes

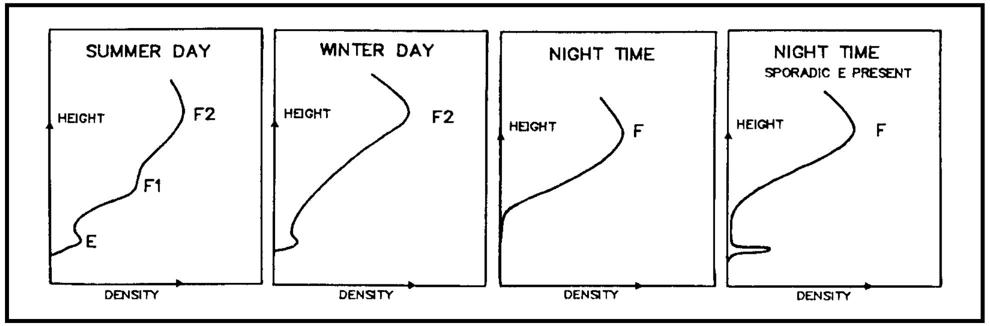
Variability of the ionosphere

The fact that the ionosphere is created by the sun tells us that it will vary with many factors - these include:

- Diurnal - throughout the day
- Seasonal - throughout the year
- Location - both geographic and geomagnetic factors
- Solar activity - 11 year solar cycle and disturbances
- Height - different layers

Examples of some vertical profiles of electron density against height are shown below.

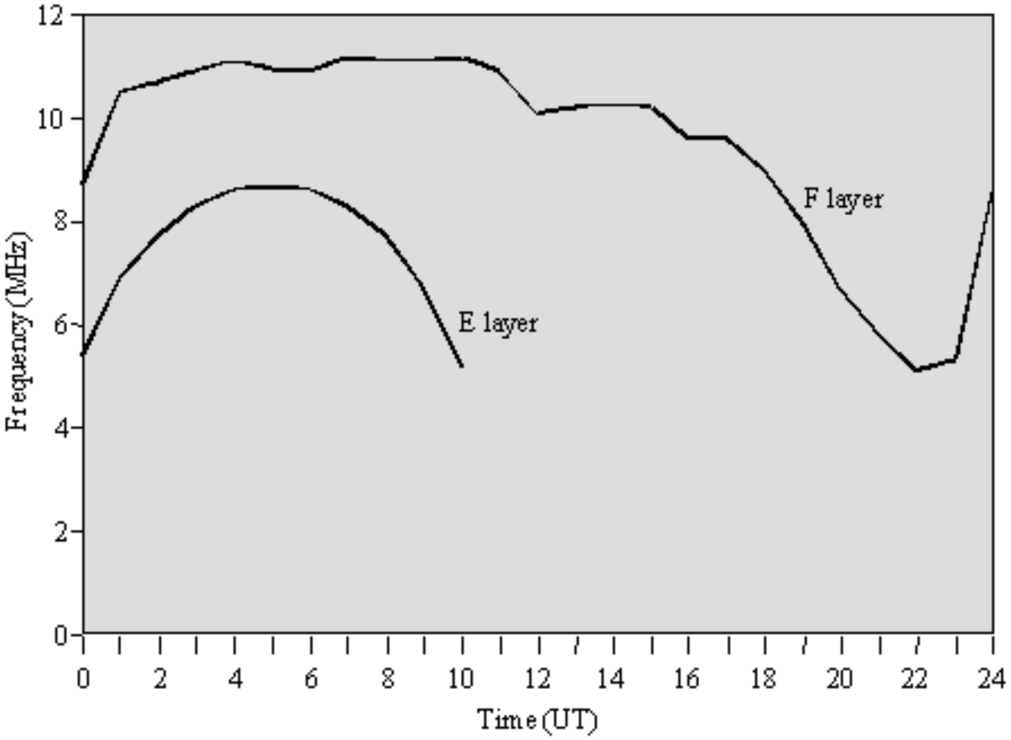
The electron density at a specific location at a specific time is not easy to model, particularly in the F2 layer. It depends on a number of complex physical interactions of **production** (EUV and impact ionisation), **loss** (recombination of electrons with neutrals and ions) and **transport** (neutral wind effects).



Physical Processes

Diurnal variations

Operating frequencies are normally higher during the day and lower at night. With dawn, solar radiation causes electrons to be produced in the ionosphere and frequenci increase reaching their maximum around noon. During th afternoon, frequencies begin decreasing due to electron loss and with evening, the D, E and F1 regions become insignificant. HF sky wave communication during the night therefore by the F2 region and absorption of radio waves lower because of the lack of the D region. Through the night, frequencies decrease reaching their minimum just before dawn.



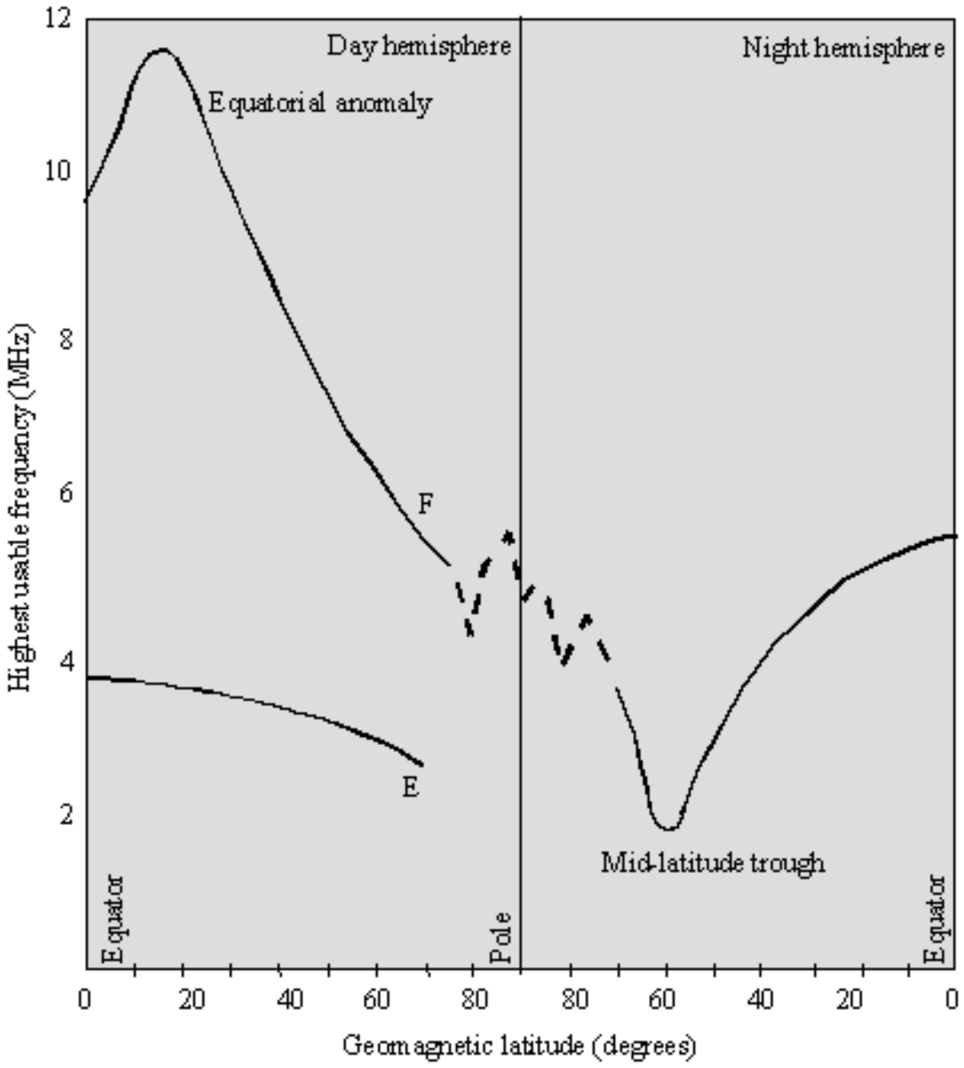
day *night*

Physical Processes

Variations with latitude

The figure indicates the variations in E and F region frequencies at noon and midnight from the poles to the geomagnetic equator.

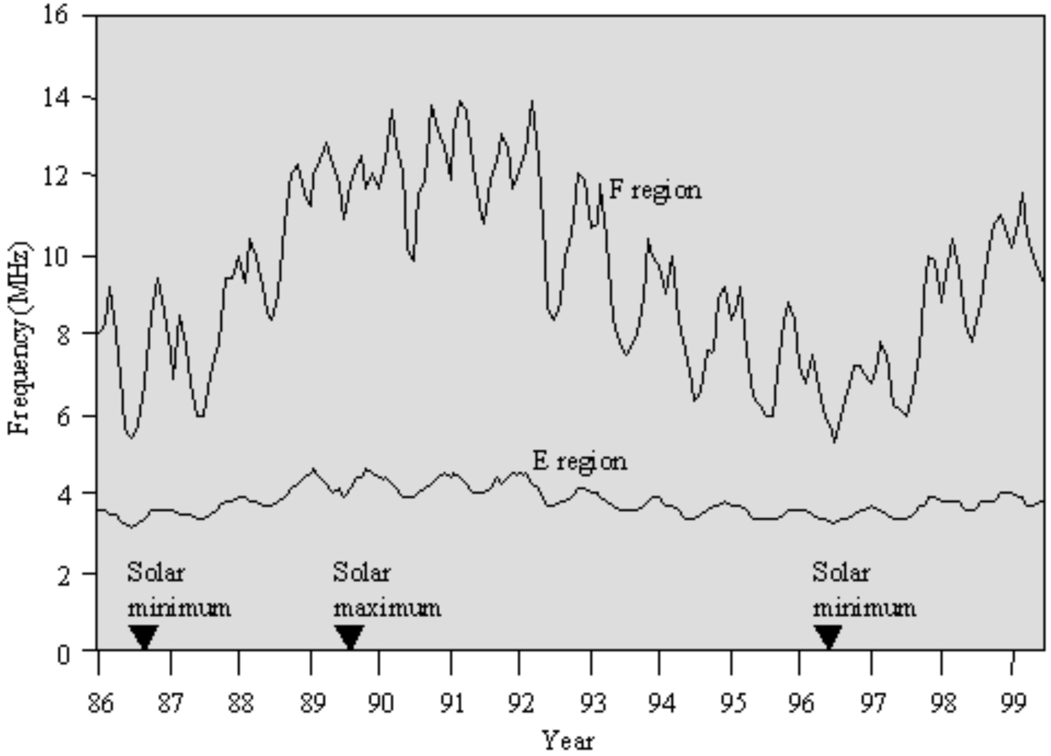
During the day and with increasing latitude, solar radiation strikes the atmosphere more obliquely, so the intensity of radiation and the electron density production decreases towards the poles.

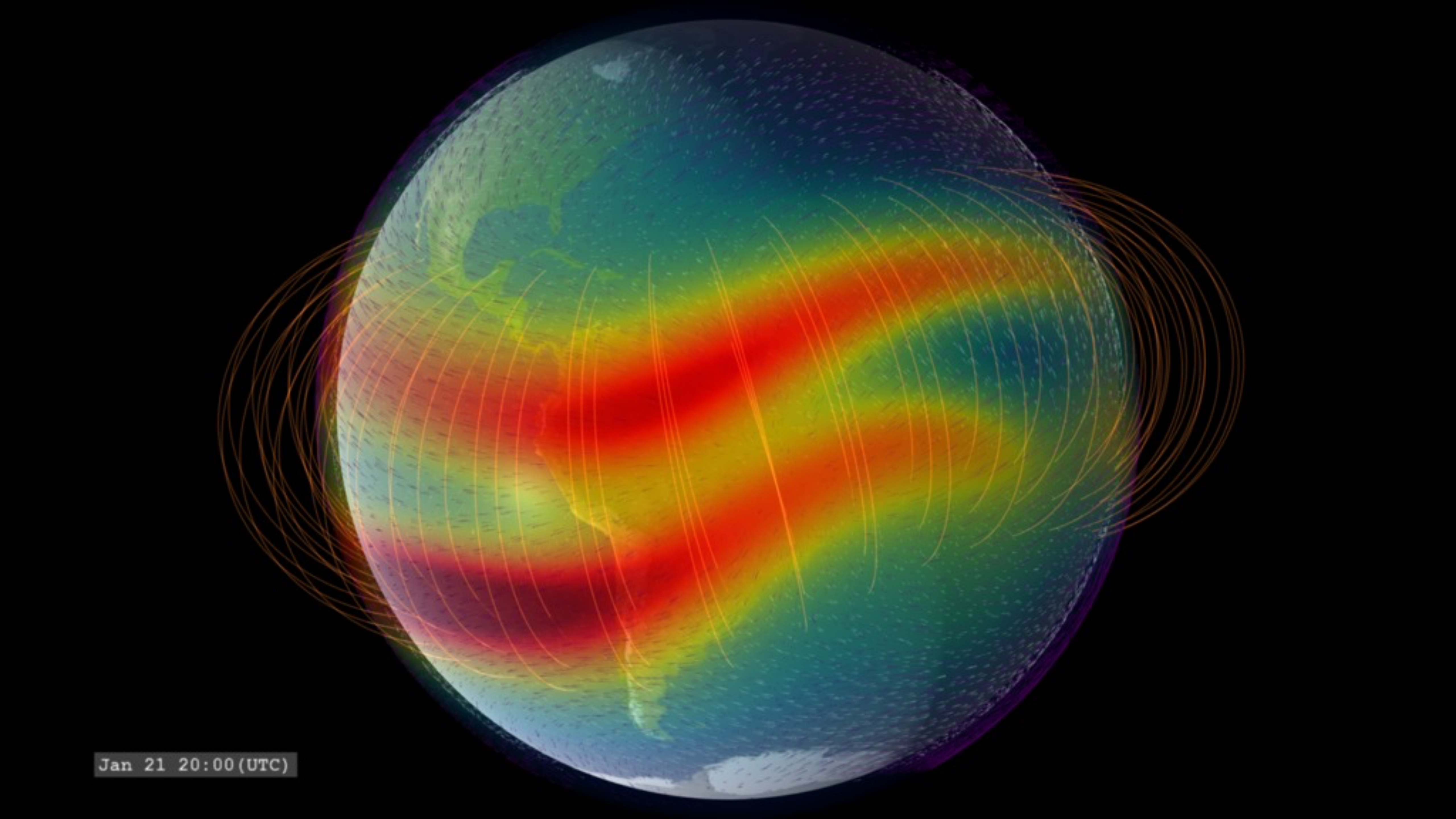


Physical Processes

Variations due to the solar cycle

The Sun goes through a periodic rise and fall in activity which affects HF communications; solar cycles vary in length from 9 to 14 years (nominally 11 years). At solar minimum, only the lower frequencies of the HF band will be supported by the ionosphere, while at solar maximum the higher frequencies will successfully propagate. This is because there is more radiation being emitted from the Sun at solar maximum, producing more electrons in the ionosphere which allows the use of higher frequencies.




Jan 21 20:00(UTC)

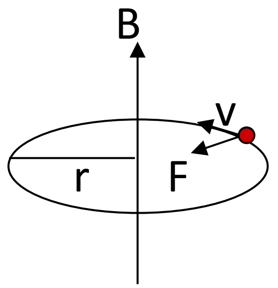
Fundamental ideas

Electron gyrofrequency

In a magnetic field particles tend to go round in circles. The rate of this gyration of called the gyrofrequency.

If the magnetic flux density is B, the charge on the particle is e and its velocity is v then the Lorentz force on the particle is

$$F=Bev$$



This is for an electron. Note that a positively charged ion would travel in the opposite direction.

The particle is constrained to move in a curved path. The gyroradius is

$$r_B = mv/Be$$

The period of revolution is $P=2\pi r_B/v = 2\pi m/Be$ and the angular frequency is

$$\omega_B = 2\pi/P = Be/m \textit{ radians per second}$$

$$\mathbf{f_B=Be/2\pi m}$$

Fundamental ideas

Plasma Frequency

If we consider a slab of ionosphere in which electrons and positive ions are ‘displaced’ and then released, they would oscillate with simple harmonic motion.

The higher the ionospheric density the higher the radio frequency that can be reflected.

The **critical frequency** of a layer is the maximum frequency that can be reflected from it. Radio waves at higher frequencies will simply pass straight through.

Fundamental ideas

Radio waves in an ionised medium

The refractive index of an ionised medium is given by the Appleton-Hartree equation:

$$\mu^2 = 1 - \frac{X}{1 - jZ - \left[\frac{Y_T^{\ 2}}{2(1 - X - jZ)}\right] \pm \left[\frac{Y_T^{\ 4}}{4(1 - X - jZ)^2} + Y_L^{\ 2}\right]^{1/2}}$$

$$\text{where } \mathrm{X} = \omega_N^{\ 2}/\omega^2, \mathrm{Y} = \omega_B/\omega, \mathrm{Y_L} = \omega_L/\omega, \mathrm{Y_T} = \omega_T/\omega, \mathrm{Z} = v/\omega$$

ω_N is the angular plasma frequency, ω_B electron gyrofrequency (where ω_L and ω_T are the longitudinal and transverse components)

We can deal with this equation by looking at specific cases.

- If absorption of the radio wave is small we can put Z=0.
- If we choose to ignore the magnetic field of the Earth Y=0.

Then we can use for refractivity $\boldsymbol{\mu^2 = 1 - \frac{\omega_N^{\ 2}}{\omega^2}}$

Fundamental ideas

Plasma Frequency

If we consider a slab of ionosphere in which electrons and positive ions are 'displaced' and then released, they would oscillate with simple harmonic motion with angular frequency ω_N.

$$\omega_N = \frac{Ne^2}{\varepsilon_0 m_e}$$

where N is the density (concentration) of free electrons, m_e is the mass of one electron, e is the charge on an electron and ε_0 is the permittivity of free space.

Using $f_N=2\pi/\omega_N$ and putting in the values for the constants we find that

$$\mathbf{f_N^{\ 2} = 80.5N}$$

where f_N is in units if Hz and N is in units of electrons m^{-3}.

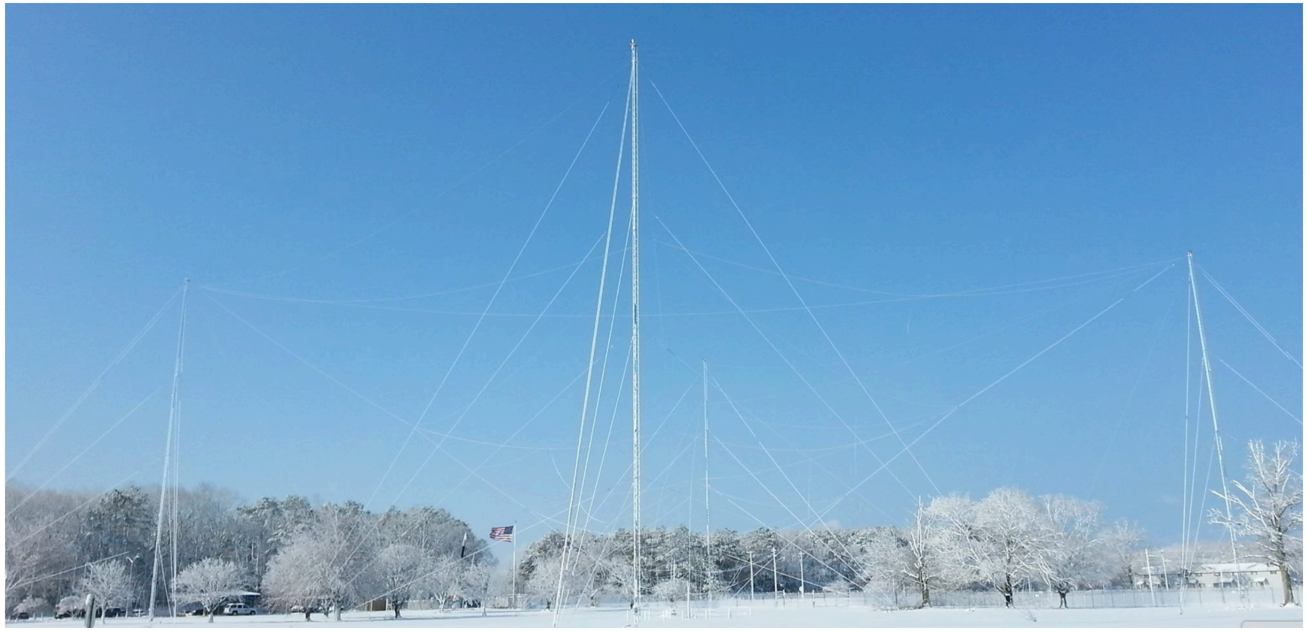
This is a very useful equation because it tells us the frequency of a radio wave (transmitted at vertical incidence) that will reflect from a layer of the ionosphere with a particular electron density.

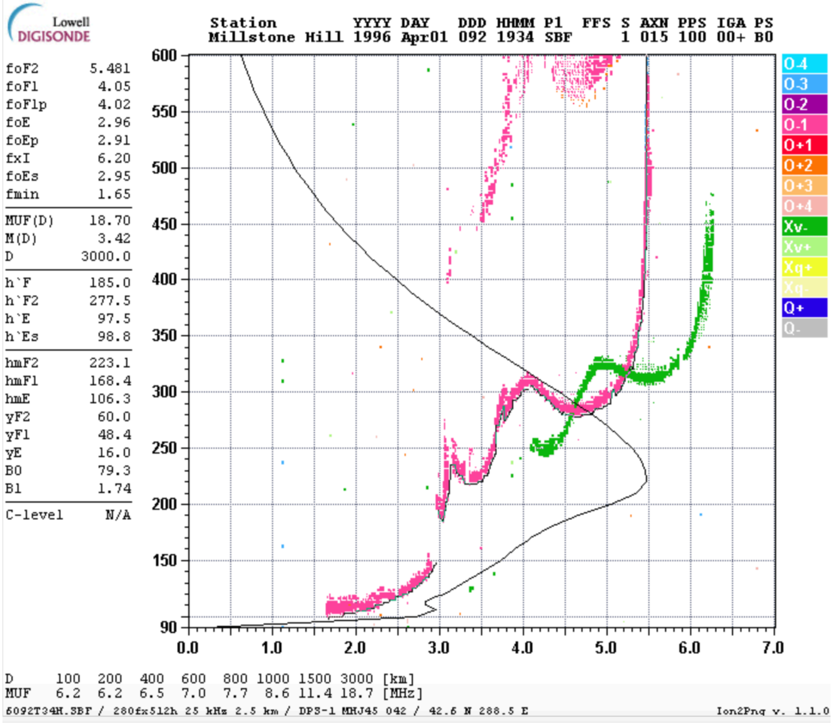
The **critical frequency** of a layer if the maximum frequency that can be reflected from it. Radio waves at higher frequencies will simply pass straight through.

- Introduction & discovery
- Physics
- **Instruments**
- Storm conditions
- System impacts

Instrumentation

Ionosondes

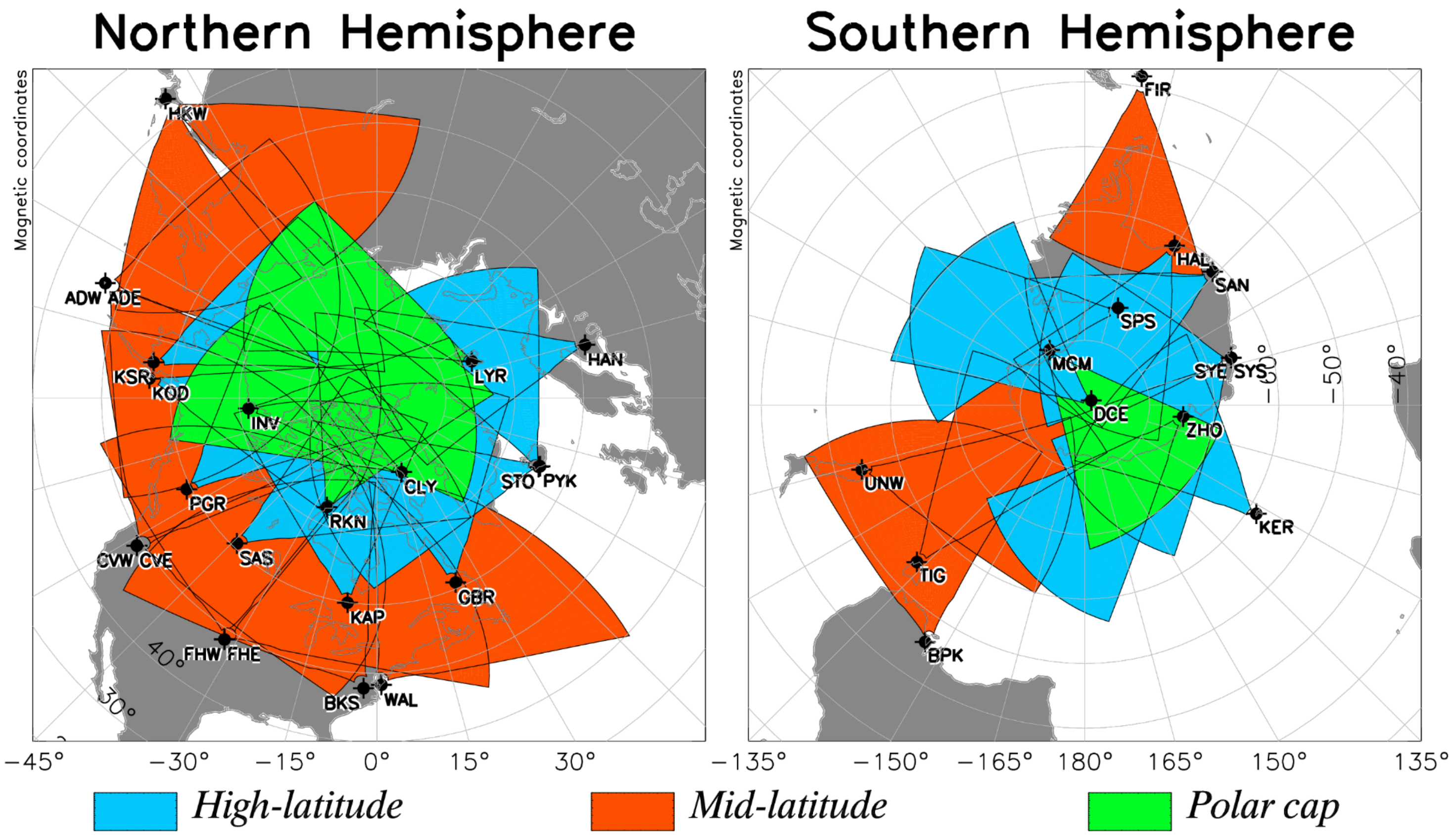




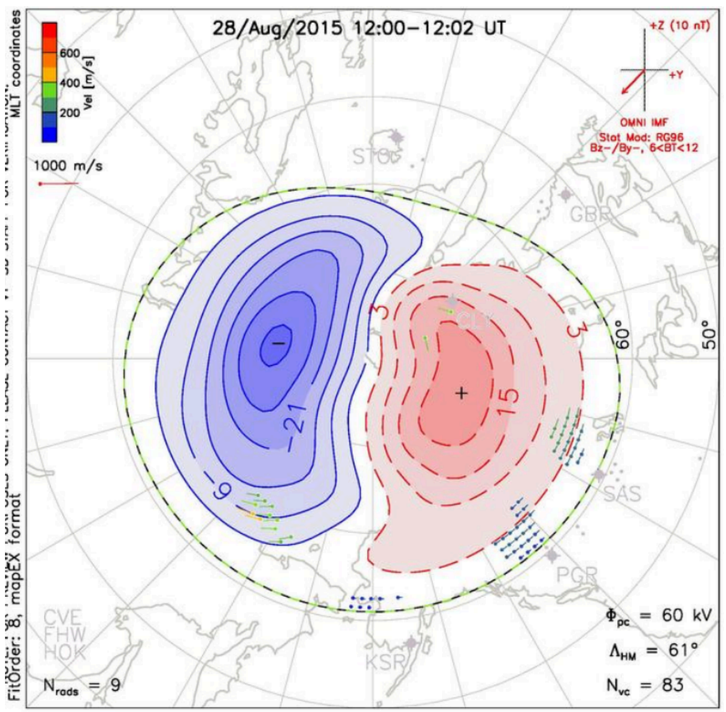
http://www.digisonde.com/index.html

Instrumentation

HF/VHF Backscatter (coherent scatter)



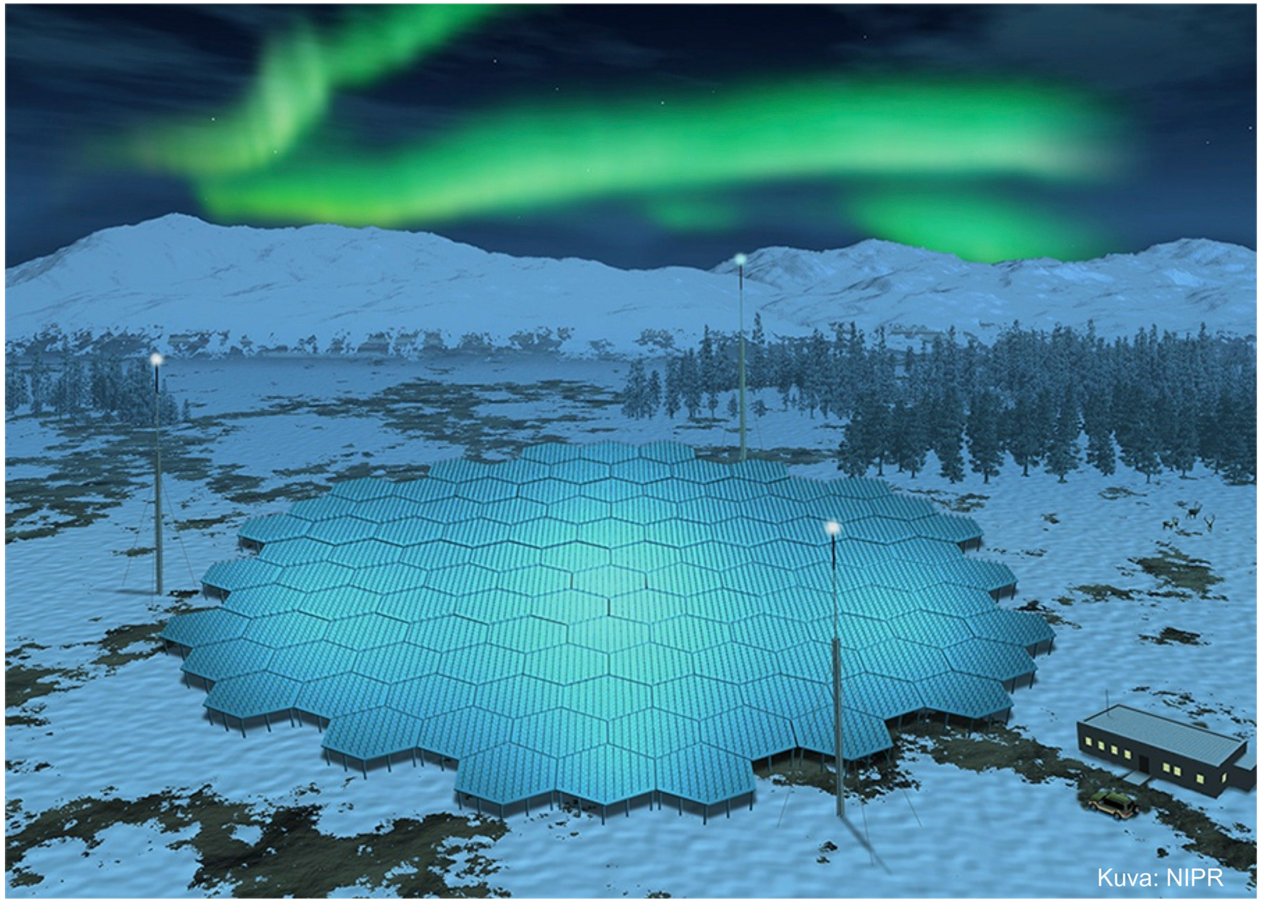
Example map

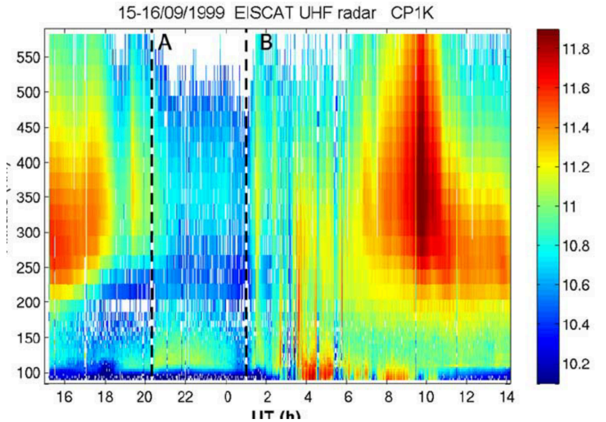


http://vt.superdarn.org/tiki-index.php?page=Radar+Overview

Instrumentation

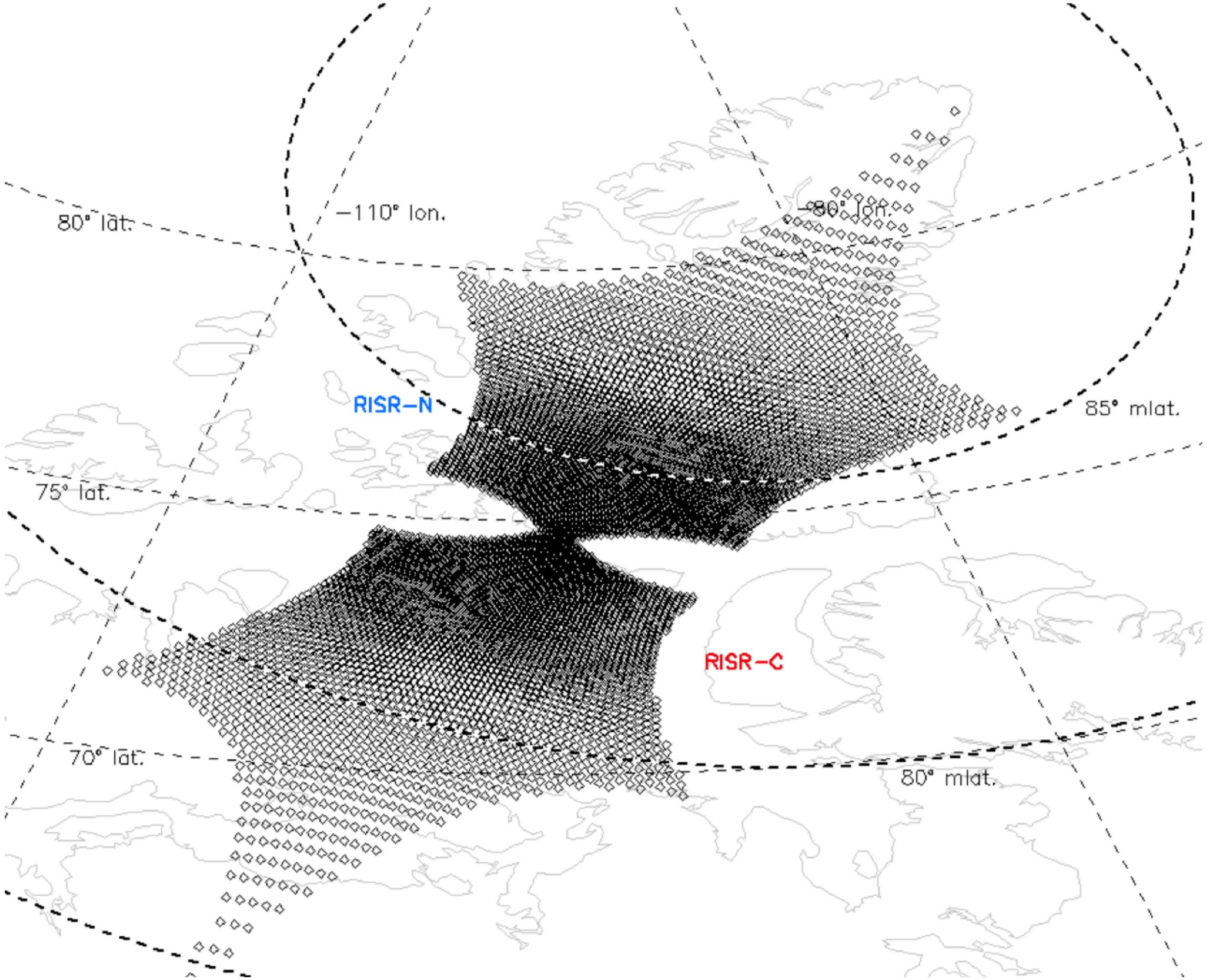
Incoherent scatter





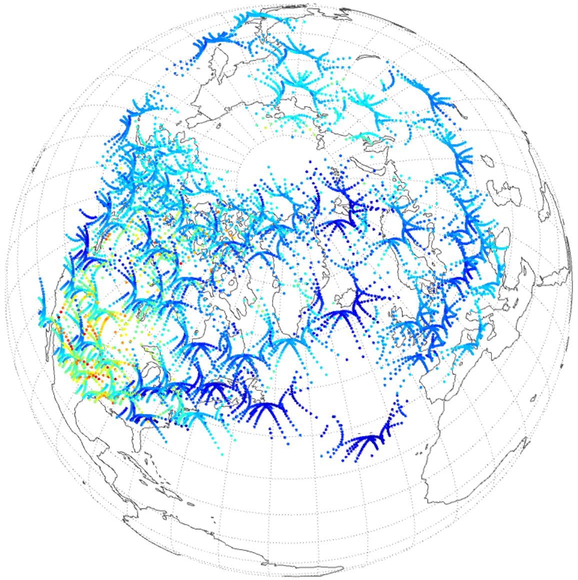
https://www.haystack.mit.edu/atm/mho/instruments/isr/isTutorial.html

Instrumentation

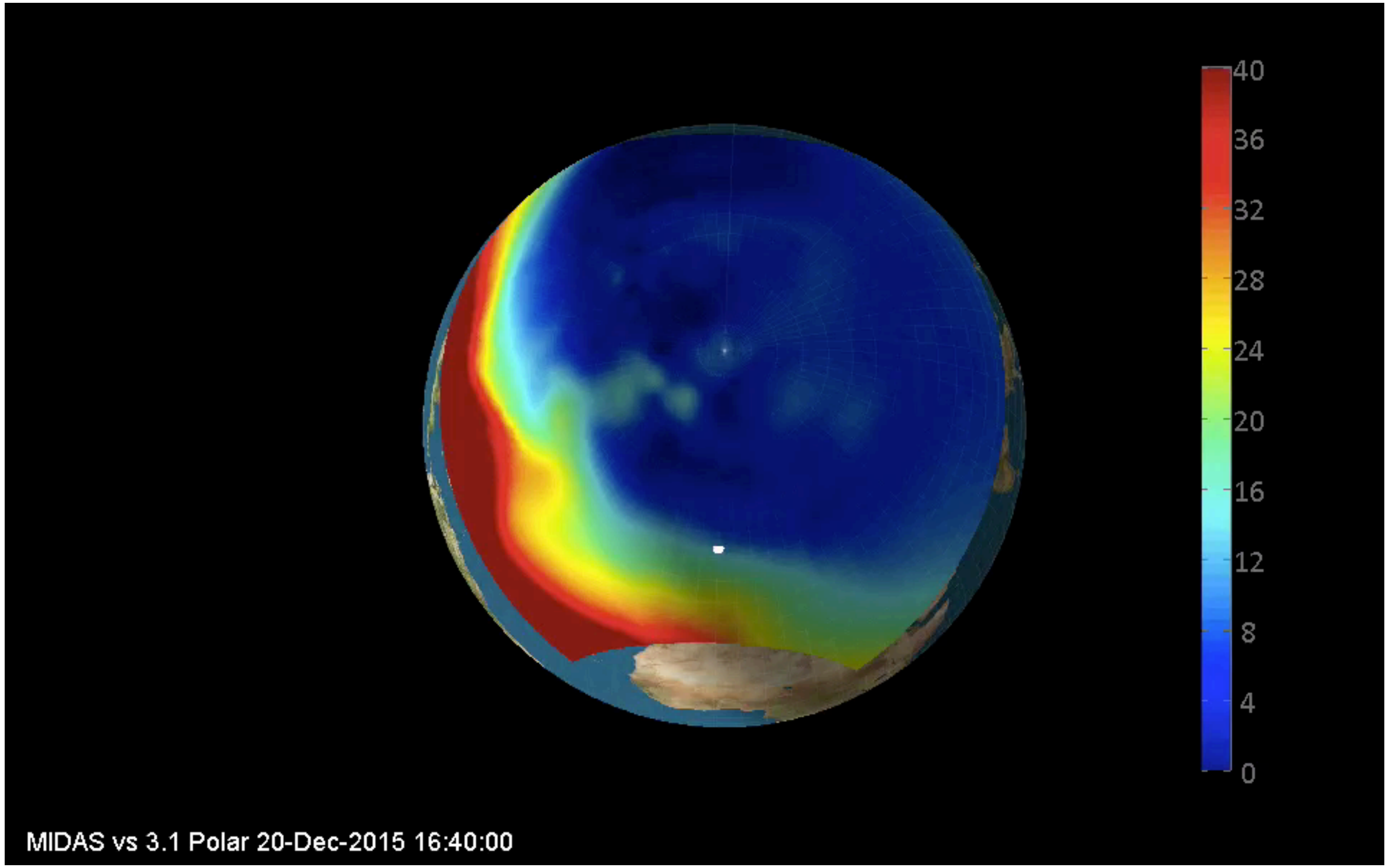


Incoherent scatter

Instrumentation



GPS tomography & data assimilation



- Introduction & discovery
- Physics
- Instruments
- **Storm conditions**
- System impacts

Storms

The aurora

An aurora is an atmospheric effect produced when charged particles from the radiation belts surrounding a planet like the Earth are accelerated down the magnetic field lines near the north and south magnetic poles. They collide with atoms of gas in the atmosphere such as nitrogen and oxygen, and the collisions cause the atoms to emit light at very specific wavelengths. This causes aurora to have red and blue colours. All of this happens between 70 and 300 kilometres above the ground and requires very low gas densities.

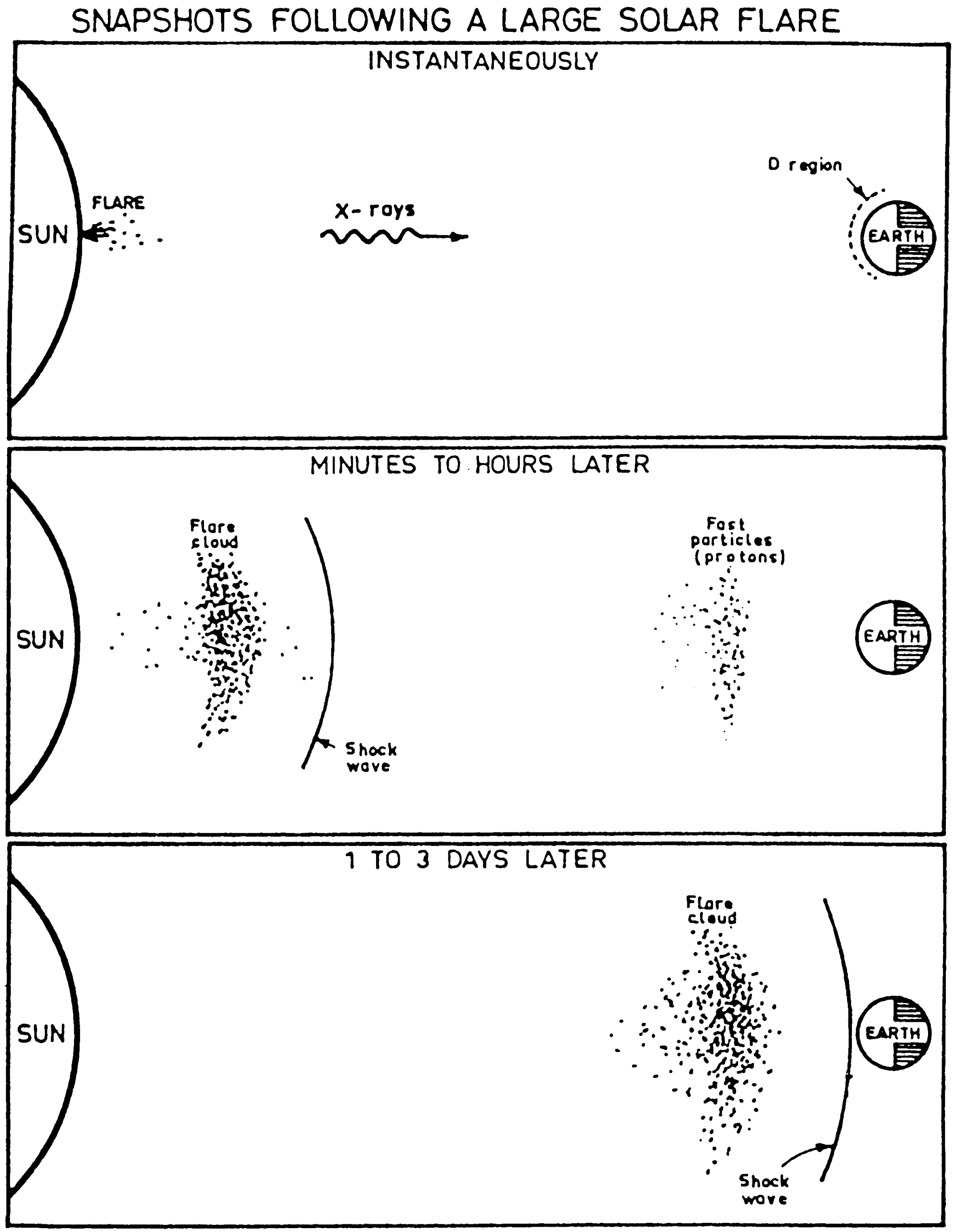
Aurora at Longyearbyen, Svalbard

(80 degrees north).

Storms

Summary of effects of a solar flare

- X-rays
- Protons
- Plasma cloud



Storms

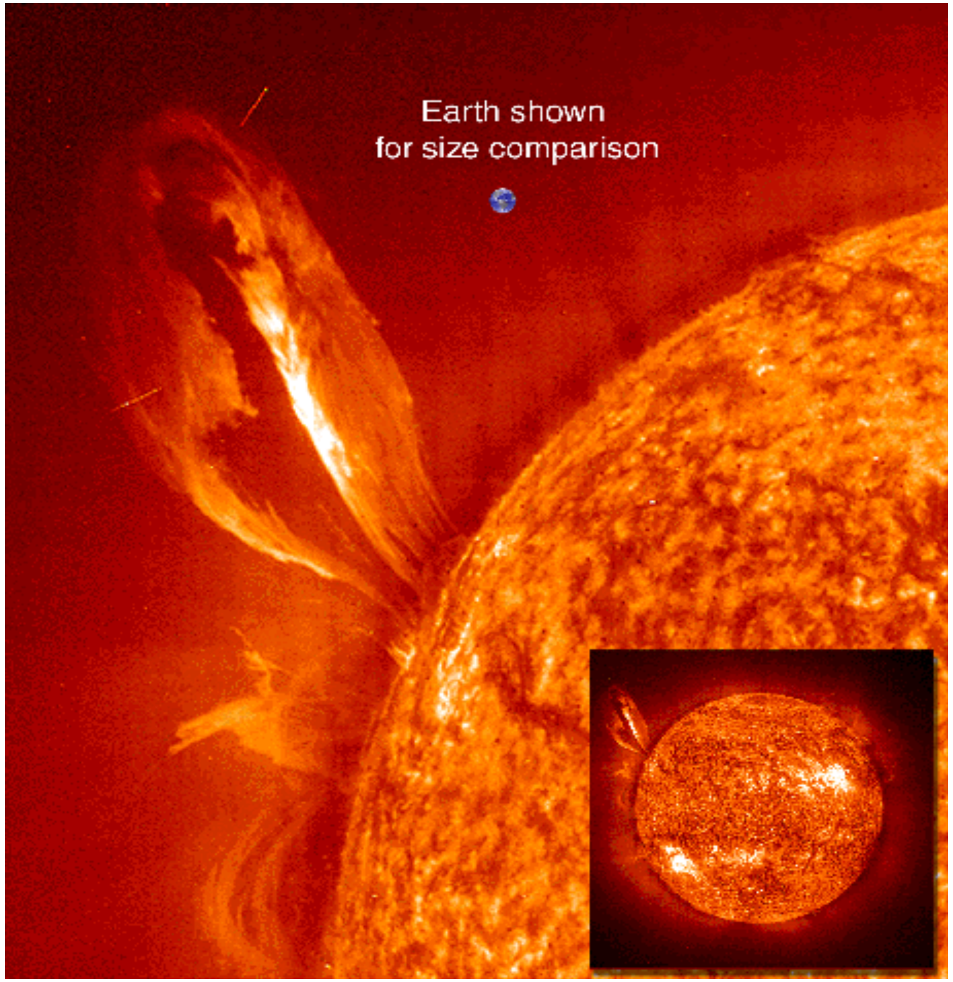
Solar flares

A solar flare is a large explosion on the sun, apparently caused by a sudden release of magnetic energy. Flares have a range of sizes (500-1000 millionths of the sun's disk) and can last for varying lengths of time (up to several hours). Their intensity has to be measured from satellites because the Earth's atmosphere absorbs the X rays, which is the best indicator of the energy in a flare.

Flares occur throughout the solar cycle, but most often at high solar activity. Only a few per cent will have an effect on the Earth's ionosphere.

Flares can have three different effects on the ionosphere - depending on the emissions from the flare and whether the geometry of the sun-Earth system in favourable or not.

- X-rays
- Protons
- Plasma cloud

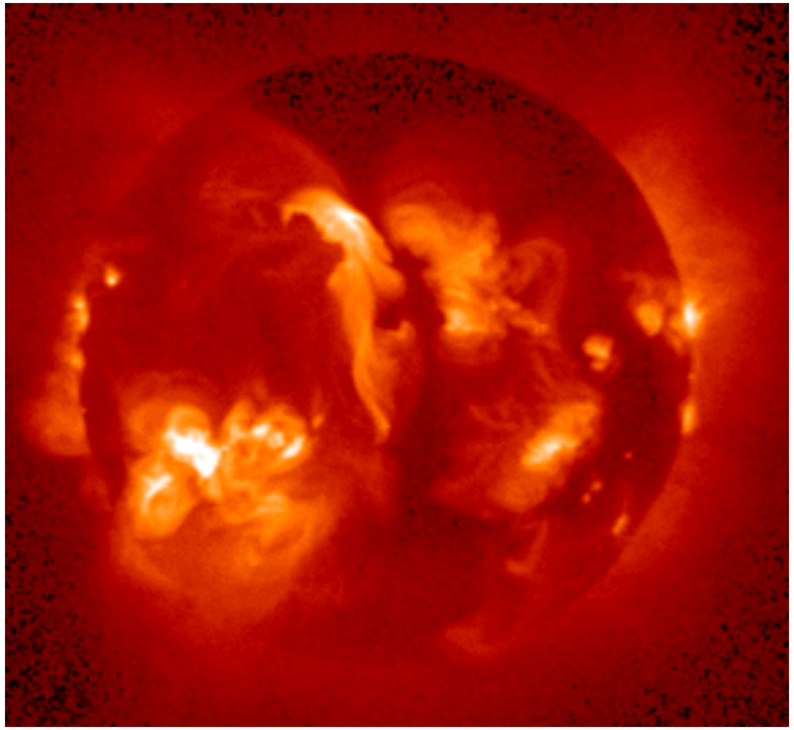


Large, eruptive prominence in He II at 304Å, with an image of the Earth added for size comparison. This prominence from 24 July 1999 is particularly large and looping, extending over 35 Earths out from the Sun.

Storms

Effects of X-rays

If the flare is sufficiently energetic and occurs on the part of the sun facing the Earth then some of the X-rays will encounter the Earth's atmosphere (after 8 minutes travel time) and photo-ionise as low as the D-region. A large flare can increase the ionisation in the D-region by a factor of 10. This then means that there are 10 times as many electrons to take energy from the radio waves and lose it in collisions with the neutral atoms. Often this can result in all of the energy of the radio wave absorbed in the D-region. This is known as **shortwave fadeout**.



This effect can last as long as a flare (up to a couple of hours). It effect the lowest frequencies most strongly, so we should attempt to use higher frequencies of communication during this type of event.

Since SWF is caused by X rays which travel in straight lines, it only effects the side of the Earth facing the sun. For this reason it is sometimes called 'daylight fadeout.'

They effect the equatorial regions most severely where the sun us most directly overhead.

Storms

Effects of solar protons

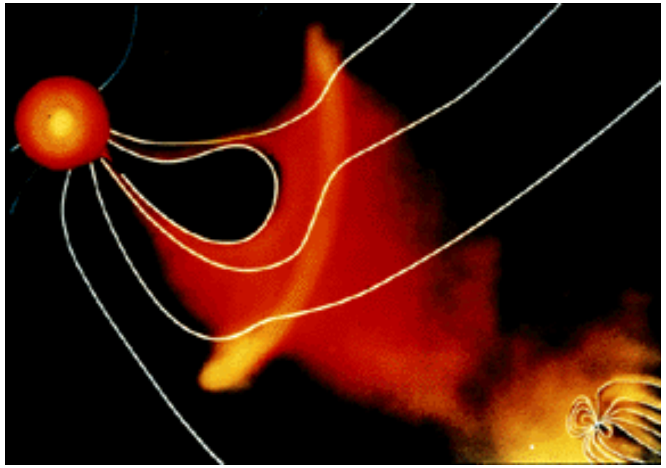
Some very energetic flares eject a stream of protons which can hit the Earth. The protons are produced within the flare site by the ionisation of hydrogen. On their way to Earth these protons (at speeds up to 0.8 c) can cause damage to unshielded satellites and astronauts.

The protons can arrive at the Earth between 10 minutes and a few days after the start of the flare. When they arrive at the Earth they encounter the Earth's magnetic field and gyrate around the lines of force. At low latitudes the magnetic field is horizontal and therefore these regions are protected from proton's penetrating into the ionosphere.

At higher latitudes the field lines are almost vertical and the electrons can penetrate right into the ionosphere. In the D-region they can cause a dramatic increase in the electron density by collision ionisation. This results in severe absorption around the polar cap regions and is known as a **polar cap absorption event** (PCA). The effects can last for several days, up to a maximum of about one week.

The HF communicator must wait until the stream of protons has declined.

These events are rare and occur mostly at solar maximum (about 8 per year).



Storms

Effects of a plasma cloud

An ionospheric storm is analogous to an atmospheric storm. Conditions for HF propagation are dramatically changed, especially the foF2.

If a large energetic flare occurs near the centre of the sun as seen from the Earth, and if the cloud of plasma hits the Earth it has severe effects. The electric field in the ionosphere is changed and many factors such as the time of day, local time, season and latitude will determine the effect of the storm on the F region. Effects are greater in equinox and summer and at higher latitudes.

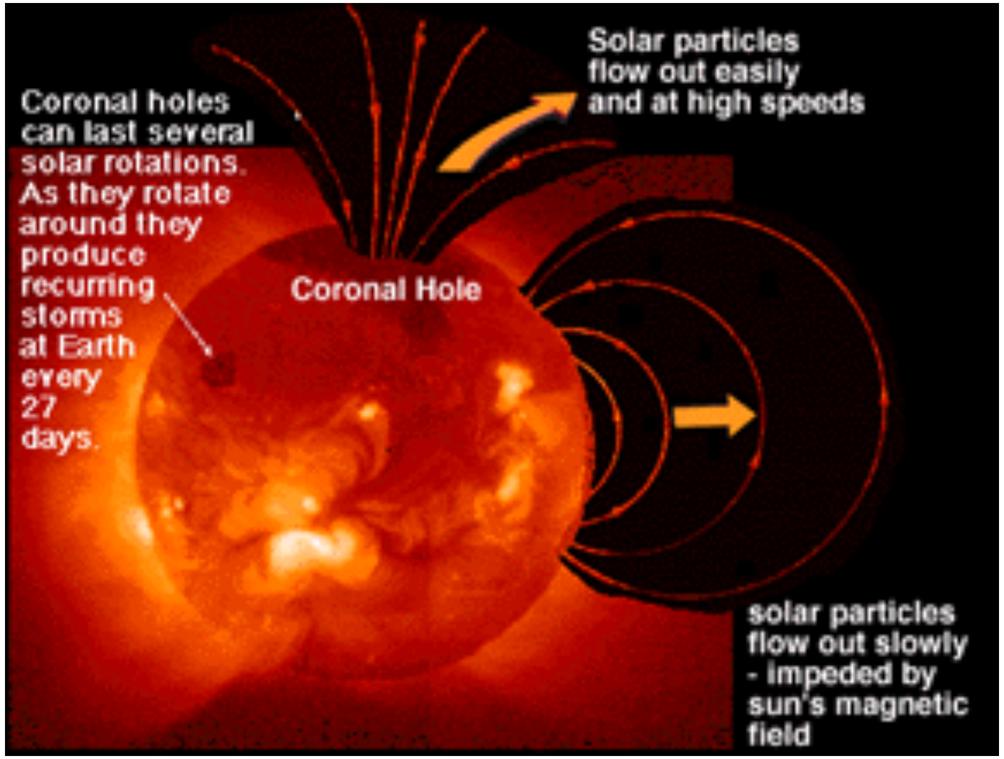
The storm commences about 2-4 days after the flare. Increases in F-region critical frequency (common in winter storms) will not be noticed by the HF communicator. However, typical effects of storms usually include a decrease in foF2 by a factor of around 2.

Storms

High Speed Solar Wind Stream

Coronal holes are ‘cool’ open structures from which the lines of force of the sun’s magnetic filed stretch out into space. Ionised material can pour out from these and travel along the field lines in a ‘**high speed solar wind stream’** or HSSWS. This can travel from the sun to the Earth in about 4 days (300- 500 km per second).

Effect are similar to a plasma cloud, although the resulting ionospheric storm is less severe but may be longer lived.

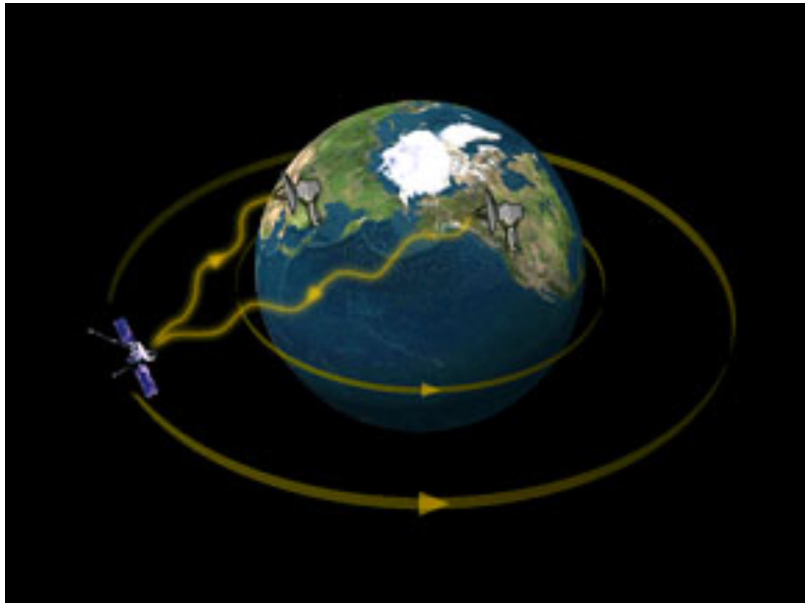


Polar coronal holes almost always exist. However, the fast streams that are associated with the polar coronal holes never reach the Earth. During some intervals, coronal holes are formed at lower solar latitudes. In this location, the coronal holes spray the Earth with high speed plasma streams, like a powerful garden sprinkler, and are responsible for generating space weather storms that recur in intervals of 27 days as the coronal hole rotates back over the limb of the Sun.

Storms

Satellite to ground signals

Shorter wavelength radio waves (such as VHF, UHF and microwaves) don't curve around the Earth or reflect from the ionosphere as they travel. The fact that they penetrate the ionosphere makes them very useful for long-distance communications. These signals can be transmitted from ground stations to satellites in orbit around the Earth. The satellites receive, boost and re-transmit the signal to other ground stations.



During periods of disturbed space weather, the ionosphere can be filled with small-scale size irregularities. Radio signals traversing the disturbed ionosphere are disrupted by these irregularities which cause phase and amplitude fluctuations in the signals.

Storms

Ionospheric Scintillations

There are 3 major sectors of scintillation activity. These are the equatorial region and the north and south polar regions.

Irregularities in the polar region are caused by precipitating high velocity auroral particles. The auroral particles violently collide with atmospheric particles, knocking electrons free, and creating enhanced densities. The fluxes of high-velocity precipitating particles are very structured in space and create irregular structures in the ionosphere. These same particles are responsible for the auroral lights. In addition, the flow of ionospheric plasma from the dayside to the nightside and vice-versa at high-latitudes results in the formation of large-scale plasma blobs and density troughs. The steep edges of these structures are unstable and soon generated intense regions of small-scale density irregularities that produce severe scintillation effects. These types of irregularities vary in severity and geographic location during space weather disturbances.

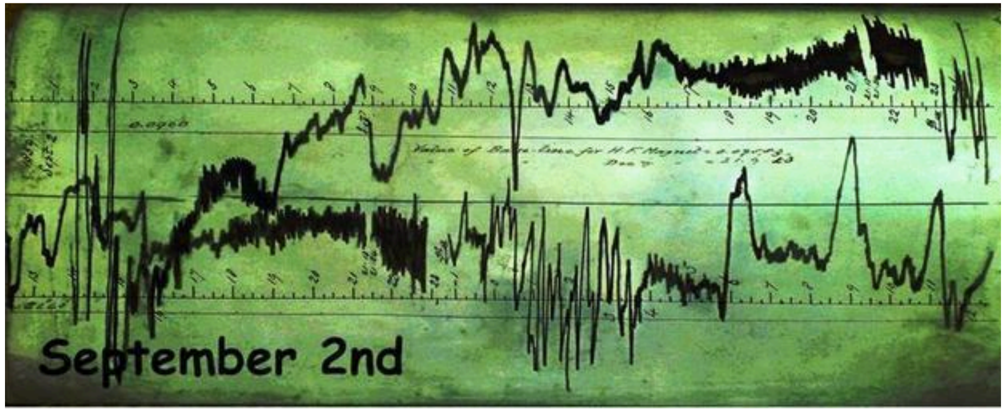
In the equatorial regions, irregularities result from bubbles that form at the bottom of the F region ionisation layer and percolate upward through the topside ionosphere, emerging just after sunset, distorting into plumes. The steep edges of the plumes are unstable and smaller-scale irregular density structures develop along these edges. The small-scale irregularities cause intense scintillation effects. Individual patches of irregularities have lifetimes of 2-3 hours.

- Introduction & discovery
- Physics
- Instruments
- Storm conditions
- **System impacts**

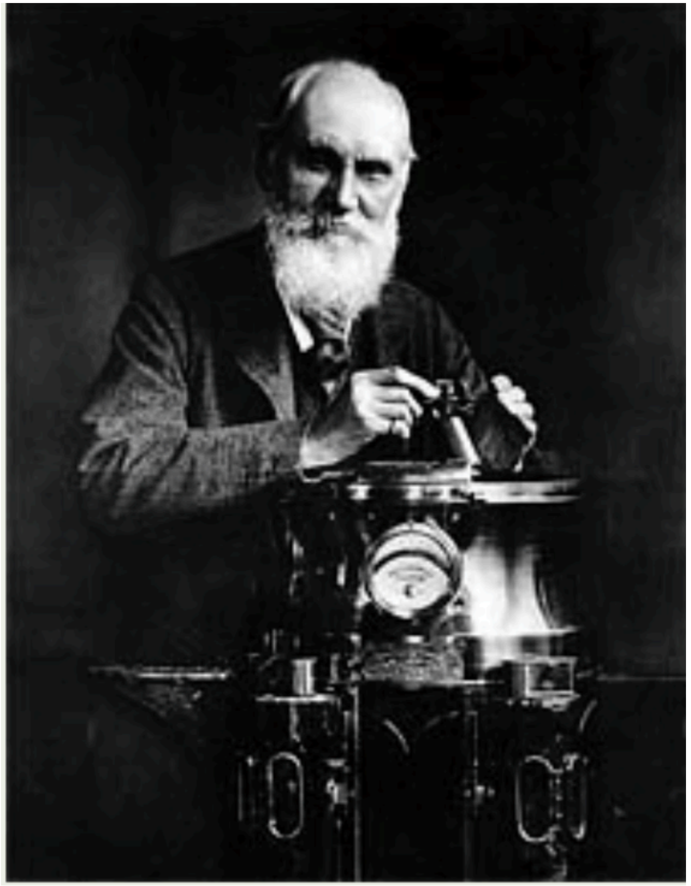
Carrington event

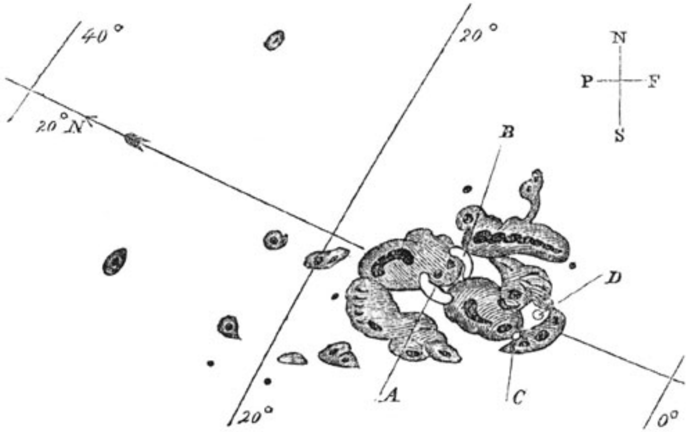
Richard Carrington – British solar astronomer

- 1859 on Sept 1st - Carrington saw flash of light on the Sun - a large solar flare
- Kew observatory – subsequent magnetic disturbance



- Suspected a solar-terrestrial connection



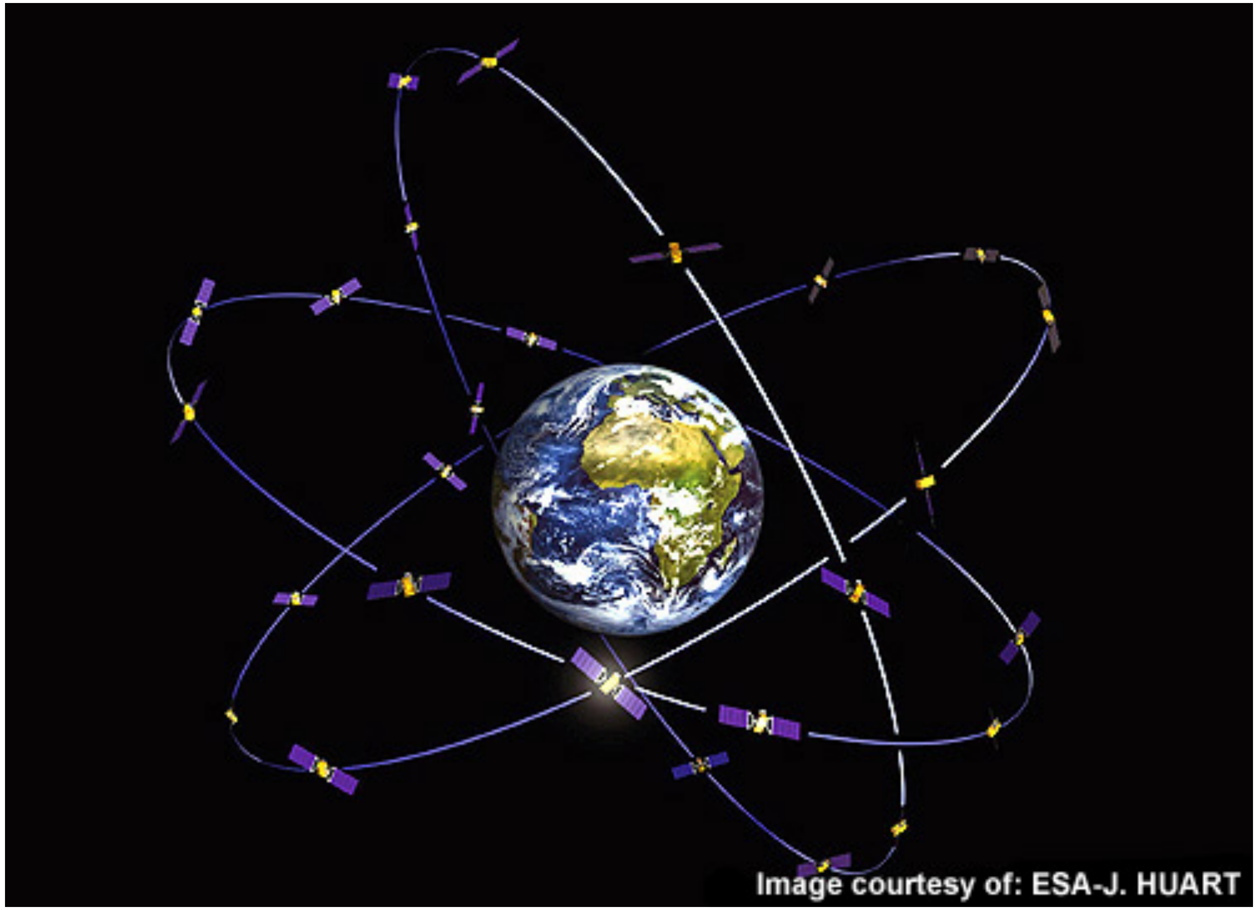


- Boston, Friday, Sept. 2 1859

“There was another display on the aurora last night, so brilliant that at about one o'clock ordinary print could be read by the light. The effect continued through this forenoon considerably affecting the working of the telegraph lines. The auroral currents from east to west were so regular that the operators on the Eastern lines were able to hold communication and transmit messages over the line between this city and Portland, the usual batteries being discontinued from the wire. The same effects were experienced upon the Cape Cod and other lines”

Reference: New York Times, September 3, 1859

System Impacts

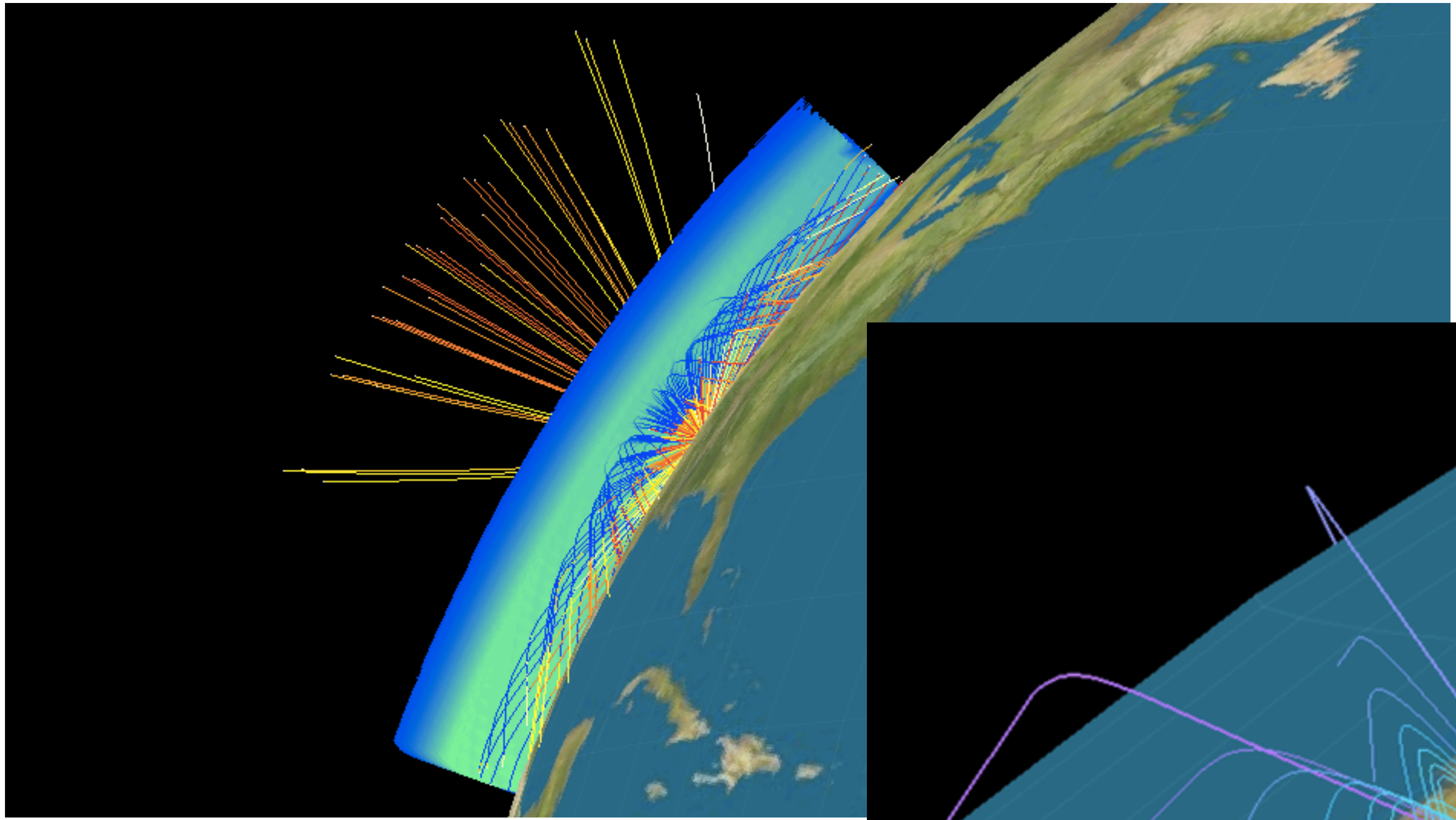


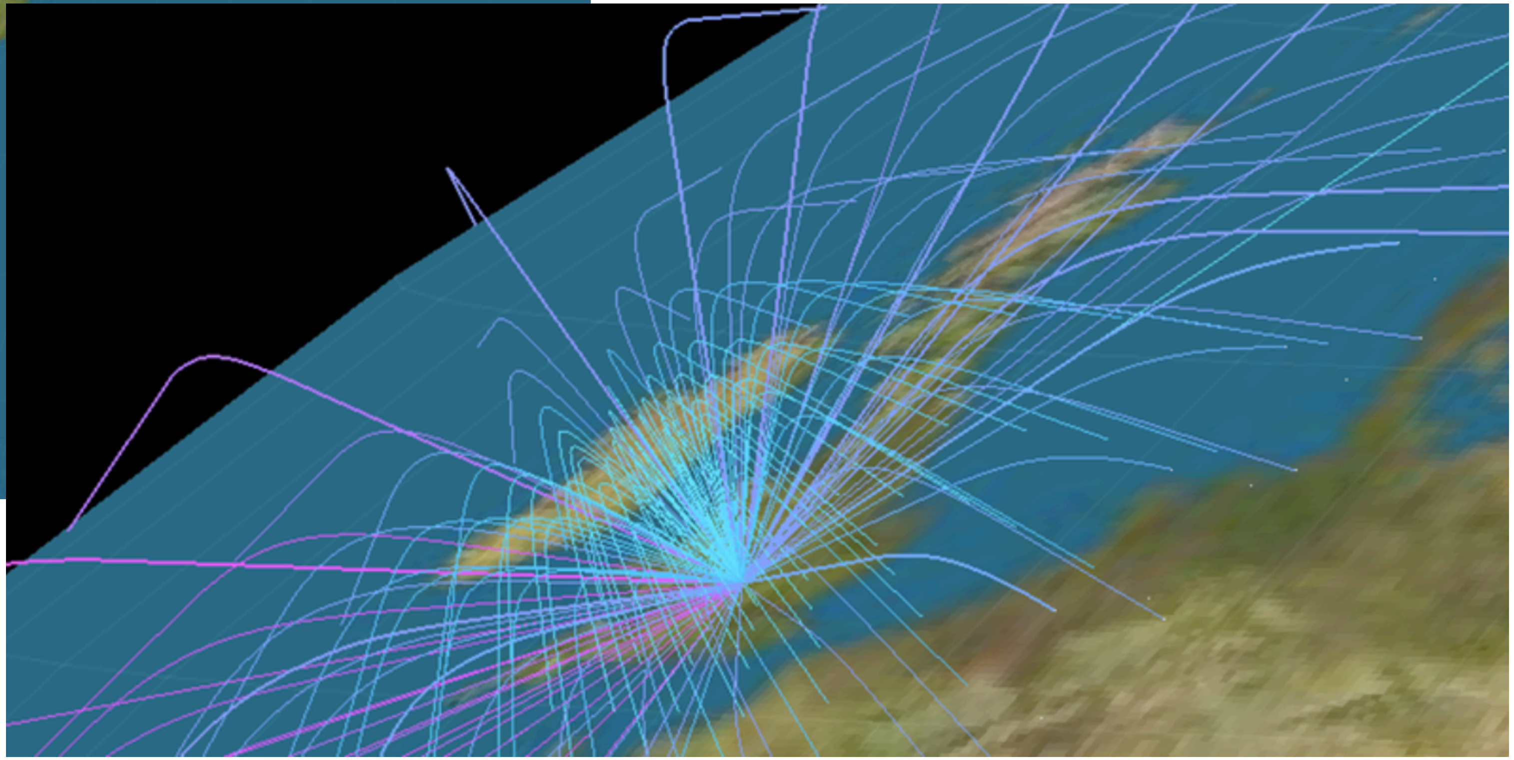
System Impacts

System Impacts

System Impacts

System Impacts





System Impacts

- X rays (e-m radiation)
- Particle radiation
- Magnetic fields

- Radio interference
- Electronics malfunction
- Exposure to radiation
- Induced electric currents
- Ionospheric disturbance
- Navigation errors
- Timing errors
- Communication disturbances

- Spacecraft
- Power grid
- Rail
- Aircraft
- Road
- Shipping
- Finance
- Telecoms
- HF comms

Recommended reference book:

The Solar-Terrestrial Environment: An Introduction to Geospace - the Science of the Terrestrial Upper Atmosphere, Ionosphere, and Magnetosphere (Cambridge Atmospheric and Space Science Series) by John Keith Hargreaves

Summary

- Introduction & discovery
- Physics
- Instruments
- Storm conditions
- System impacts

C.N.Mitchell@bath.ac.uk